



# **BGRS/SB-2022**

**The 13th International Multiconference**

## **Bioinformatics of Genome Regulation and Structure/ Systems Biology**

**BGRS/SB-2022  
NOVOSIBIRSK, RUSSIA  
04–08 July, 2022**

<https://bgrssb.icgbio.ru/2022/>

## Organizers



Institute of Cytology and Genetics, SB RAS (ICG SB RAS)



Novosibirsk State University (NSU)



Boreskov Institute of Catalysis, SB RAS (IC SB RAS)



Novosibirsk branch of The Vavilov Society of Geneticists and Breeders (VOGIS)



Department of Information Biology,  
Novosibirsk State University (KIB NSU)

## Sponsors

---



The conference is held with the support of the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (Novosibirsk), project No. 075-15-2019-1662

### GOLD SPONSORS

---



DIA-M  
Russia, 129346, Moscow, PO box 100. Russia, 129345,  
Moscow, Magadanskaya 7, building 3  
Tel.: +7 (495) 745-05-08, fax: +7 (495)745-05-09  
E-mail: info@dia-m.ru, URL: www.dia-m.ru

### SILVER SPONSORS

---



MAXMEDIKAL  
Moscow, st. Narodnogo Opolcheniya, 34, building 1, office 202  
Tel.: +7 (495) 374-62-80, e-mail: max@maxmedikal.com,  
URL: www.maxmedikal.com



HELICON COMPANY  
121374, RUSSIA, Moscow, 88 Kutuzovskiy prospect  
Tel.: +7 (499) 705-50-50, e-mail: mail@helicon.ru,  
URL: www.helicon.ru/en/



SESANA  
Russia, 129345, Moscow, Podolskoe shosse, 8/5 (Tulskaya metro station)  
Tel.: +7 (495) 128-82-74, e-mail: sales@sesana.ru,  
URL: www.sesana.ru

### BASIC SPONSORS

---



BIOLABMIX  
Ready-to-use solutions for molecular biology.  
Legal and actual address: Russia, 630090, Novosibirsk, st. Ingenernaya, 28  
Sales department: svt@biolabmix.ru, tel.: +7 (383) 363-22-40  
URL: https://biolabmix.ru/en/

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

**BIOINFORMATICS OF GENOME REGULATION  
AND STRUCTURE/SYSTEMS BIOLOGY  
(BGRS/SB-2022)**

The Thirteenth International Multiconference

Abstracts

04–08 July, 2022  
Novosibirsk, Russia

Novosibirsk  
ICG SB RAS  
2022

**Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022) :**

The Thirteenth International Multiconference (04–08 July 2022, Novosibirsk, Russia); Abstracts /  
Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences. –  
Novosibirsk: ICG SB RAS, 2022. – 1148 p. – ISBN 978-5-91291-059-3. –  
DOI 10.18699/BGRS/SB-2022-000.

**Organizing committee**

Institute of Cytology and Genetics, SB RAS:

Svetlana Zubova (Chairperson)  
Evgeniya Antropova  
Georgy Batukhtin  
Tatiana Chalkova  
Larisa Eliseeva  
Roman Ivanov  
Tatyana Karamysheva  
Andrey Kharkevich  
Vasilij Koval  
Pavel Linkevich  
Zakhar Mustafin

Olga Petrovskaya  
Anna Smirnova  
Erlan Tokpanov  
Vladimir Zamyatin

Novosibirsk State University (NSU):

Natalia Aksenova  
Dmitry Churkin  
Olga Kalinina  
Olga Makarova

**Contacts**

All questions, please contact the official address  
of the Organizing committee of the conference: [bgrs2022@bionet.nsc.ru](mailto:bgrs2022@bionet.nsc.ru)  
Organizing committee Chairman: Zubova Svetlana  
Address: Lavrentieva Avenue, 10, Novosibirsk, 630090  
Phone: +7 (383) 363 4977, email: [svetazubova@gmail.com](mailto:svetazubova@gmail.com)

## Presidium, co-chairs

**Nikolay Kolchanov**, Academic Advisor of Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Alexey Kochetov**, Director of Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Mikhail Fedoruk**, Rector of Novosibirsk State University, Novosibirsk, Russia

**Ralf Hofestädt**, Professor of University of Bielefeld, Germany

## Program committee

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

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

**Ilya Akberdin**, Biosoft.Ru; Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia; Sirius University, Sochi, Russia

**Tamara Amstislavskaya**, Scientific-Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

**Vladimir Anisimov**, Department for Carcinogenesis and Oncogerontology at the Petrov Institute of Oncology, Ministry of Health of the Russian Federation; Institute of Cytology of RAS, St. Petersburg, Russia

**Konstantin Anokhin**, Lomonosov Moscow State University, Moscow, Russia

**Sergey Arkhipov**, Novosibirsk State University, Novosibirsk; Synchrotron Radiation Facility SKIF, Koltsovo, Russia

**Yurii Aulchenko**, Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia

**Anastasiya Bakulina**, Novosibirsk State University, Novosibirsk, Russia

**Matteo Barberis**, University of Surrey, Guildford, United Kingdom

**Sergey Bartsev**, Institute of Biophysics of SB RAS, Krasnoyarsk, Russia

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

**Gennady Bocharov**, Marchuk Institute of Numerical Mathematics of RAS; I.M. Sechenov First Moscow State Medical University, Moscow, Russia

**Elena Boldyreva**, Boreskov Institute of Catalysis of SB RAS; Novosibirsk State University, Novosibirsk, Russia

**Tatyana Chernigovskaya**, St. Petersburg State University, St. Petersburg, Russia

**Valery Danilenko**, Vavilov Institute of General Genetics of RAS, Moscow, Russia

**Anatoly Derevianko**, Institute of Archaeology and Ethnography of SB RAS, Novosibirsk, Russia

**Yulia Dyakova**, National Research Centre “Kurchatov Institute”, Moscow, Russia

**Roman Efremov**, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry of RAS, Moscow, Russia

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

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

**Efim Frisman**, Institute for Complex Analysis of Regional Problems of FEB RAS, Birobidzhan, Russia

**Vladimir Golubyatnikov**, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia

**Sergey Goncharov**, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia

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

**Vladimir Ivanisenko**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Sergey Kabanikhin**, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia

**Elza Khusnutdinova**, Institute of Biochemistry and Genetics, Ufa Federal Research Centre of RAS, Ufa, Russia

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

**Nataliya Kolosova**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Fedor Kolpakov**, Sirius University, Sochi; Federal Research Center for Information and Computational Technologies, Novosibirsk; Biosoft.Ru, Novosibirsk, Russia

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

**Sergey Kostrov**, Institute of Molecular Genetics of National Research Centre “Kurchatov Institute”, Moscow, Russia

**Olga Krivorotko**, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia

**Ivan Kulakovskiy**, Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology of RAS; Institute of Protein Research of RAS, Pushchino; Vavilov Institute of General Genetics of RAS, Moscow, Russia

**Denis Larkin**, Royal Veterinary College, RVC · Department of Comparative Biomedical Sciences, London, United Kingdom; Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Sergey Lashin**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Inna Lavrik**, Otto von Guericke University, Magdeburg, Germany

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

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

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

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

**Mikhail Marchenko**, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia

**Yury Matushkin**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Mikhail Moshkin**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Alexey Moskalev**, Institute of Biology of Komi Science Centre of UB RAS, Syktyvkar, Russia

**Sergey Medvedev**, Federal Research Center for Information and Computational Technologies of SB RAS, Novosibirsk, Russia

**Vyacheslav Molodin**, Institute of Archaeology and Ethnography of SB RAS, Novosibirsk, Russia

**Vladimir Naumenko**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Alexey Okunev**, Novosibirsk State University, Novosibirsk, Russia

**Yuriy Orlov**, I.M. Sechenov First Moscow State Medical University, Moscow, Russia

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

**Andrey Palyanov**, A.P. Ershov Institute of Informatics Systems of SB RAS, Novosibirsk, Russia

**Elena Pasyukova**, Institute of Molecular Genetics of National Research Centre “Kurchatov Institute”, Moscow, Russia

**Maxim Patrushev**, National Research Centre “Kurchatov Institute”, Moscow, Russia

**Sergey Peltek**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Aleksandr Pilipenko**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Nikolay Pimenov**, Federal Research Centre “Fundamentals of Biotechnology” of RAS, Moscow, Russia

**Andrey Pokrovskiy**, Novosibirsk State University, Novosibirsk, Russia

**Vladimir Poroikov**, Institute of Biomedical Chemistry, Moscow, Russia

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

**Nikolai Ravin**, Federal Research Centre “Fundamentals of Biotechnology” of RAS, Moscow, Russia

**Alexey Romanyukha**, Marchuk Institute of Numerical Mathematics of RAS, Moscow, Russia

**Elena Salina**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Maria Samsonova**, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

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

**Dmitry Sherbakov**, Limnological Institute of SB RAS, Irkutsk, Russia

**Vadim Stepanov**, Tomsk National Research Medical Center of RAS, Tomsk, Russia

**Aleksey Troitsky**, A.N. Belozersky Research Institute of Physico-Chemical Biology MSU, Moscow, Russia

**Yakov Tsepilov**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

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

**Pedro Antonio Valdes-Sosa**, Cuban Neurosciences Center, Havana City, Cuba

**Evgeniy Vityaev**, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia

**Alexander Yanenko**, National Research Centre “Kurchatov Institute”, Moscow, Russia

**Nick Yankovsky**, Vavilov Institute of General Genetics of RAS, Moscow, Russia

**Nikolay Yudin**, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

**Maksim Zakhartsev**, Gazprom Neft – Industrial Innovations, St. Petersburg, Russia

**Elena Zemlyanskaya**, Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia

**Yan Zubavichus**, Borekov Institute of Catalysis of SB RAS, Novosibirsk; Synchrotron Radiation Facility SKIF, Koltsovo, Russia



## **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](mailto: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 June 10, 2022, ICG included 120 scientific units, which employed 1475 members; of them 474 researchers, 2 RAS Advisors, 7 Full Members of the RAS, 4 Corresponding Members of the RAS, 92 Doctors of Science, and 292 Candidates of Science. ICG trains 72 postgraduates and 27 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 2020 the overall number of publications in peer-reviewed periodicals is 725. In 2016–2020 the WoS system published 2068 articles from ICG researchers, and 37160 references to articles of the researchers were made within these five years.

*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/](http://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.bionet.nsc.ru](http://www.bionet.nsc.ru); email: [icg-adm@bionet.nsc.ru](mailto:icg-adm@bionet.nsc.ru)



# \* NSU IN FACTS & FIGURES

**8600**  
STUDENTS

**1400**

FOREIGN  
STUDENTS  
from 57 countries



**6**

departments



**3**

Institutes



**2500**

UNIVERSITY STAFF MEMBERS

**880** ASSOCIATE  
PROFESSORS

**570** FULL PROFESSORS  
with doctoral degrees

**79** MEMBERS  
OF THE RAS

MORE THAN **80**

COUNTRIES  
REPRESENTED BY MOOCs STUDENTS

**141**

PARTNER  
UNIVERSITIES  
in 26 countries

**JOINT PROGRAMS  
WITH COMPANIES:**

Baker Hughes, OSCiAI, Intel,  
Yandex, 2GIS, Academpark,  
Biotechnopark, Sberbank  
Technologies

**35**

PARTNER  
INSTITUTES  
in Academgorodok

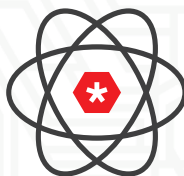
**139**

LEADING  
RESEARCHERS

RESEARCH GROUPS ARE  
INVOLVED IN NSU

**24**

INTERNATIONAL  
COLLABORATIONS  
in the field of elementary  
particle physics and life  
sciences



**115**

LABORATORIES

**6**

MIRROR  
LABORATORIES

**H>20**

**TOP-100** IN PHYSICS

QS University Rankings

**TOP-100** IN PETROLEUM  
ENGINEERING

QS University Rankings

**TOP-50** THE BEST  
STUDENT-TO-  
STAFF RATIO

THE World University  
Rankings

**8<sup>th</sup>** POSITION

QS UNIVERSITY RANKINGS EMERGING  
EUROPE AND CENTRAL ASIA

**74<sup>th</sup>** POSITION

QS UNIVERSITY RANKINGS  
IN NATURAL SCIENCES

**TOP-260**

QS UNIVERSITY RANKINGS  
among 20 000 world universities

**N\***

# Капиллярный гель-электрофорез – системы Qsep

**ДИА•М**  
современная лаборатория

**Bi optic**  
Inc.



**Qsep 1 Plus**  
(1 капилляр,  
15 образцов)



**Qsep 100**  
(1 капилляр,  
96 образцов)



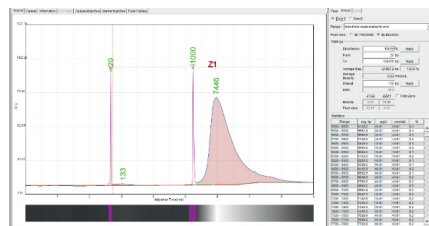
**Qsep 400**  
(4 капилляра,  
96 образцов)

- скрининг ПЦР-продуктов
- генотипирование (STR (SSR), SNP, AFLP)
- количественный и качественный анализ библиотек для секвенирования
- анализ фрагментированности геномной ДНК
- анализ внеклеточной ДНК
- анализ олигонуклеотидов
- анализ продуктов рестриктазного расщепления (RFLP)
- количественный и качественный анализ РНК (RIN)

**Недорогие расходные материалы  
(от 0,5\$ за образец)!**



**Простое и понятное ПО**



[www.dia-m.ru](http://www.dia-m.ru)

# Исследования *in vivo*: визуализация, документирование, анализ

ДИАМ

Центральный институт  
диалогов и коммуникаций

VILBER  
Smart Imaging

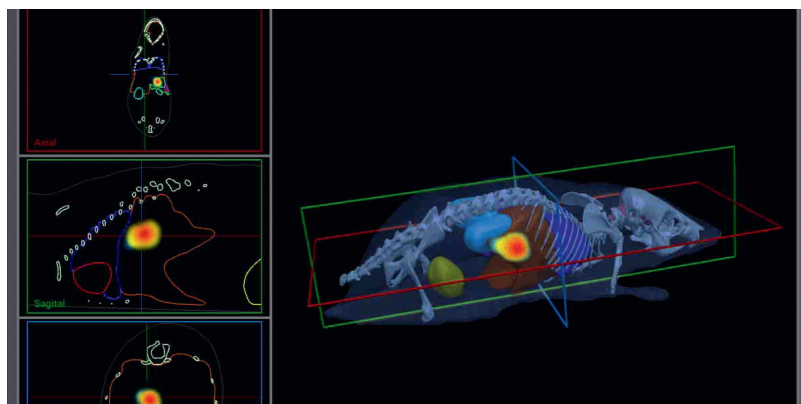
Newton

Онкология • Иммунология • Наномедицина • Вирусология • Нейробиология

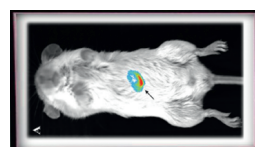
Никакой радиации • Длительные неинвазивные исследования • До 5 мышей

**Newton** – система видеонаблюдения, документирования биолюминесценции и флуоресценции, новый уровень визуализации сигнала в живых объектах неинвазивным методом

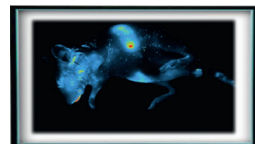
- Объекты исследований:**
- Грызуны (мыши, крысы, морские свинки, маленькие кролики)
  - Данио-рерио и др.
  - Растения
  - Срезы органов и тканей *ex vivo*
  - Образцы *in vitro* (в микропланшетах, чашках Петри и т. д.)



Bioluminescence



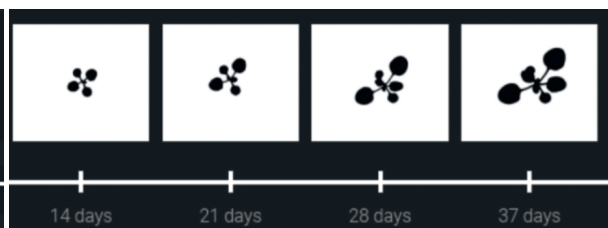
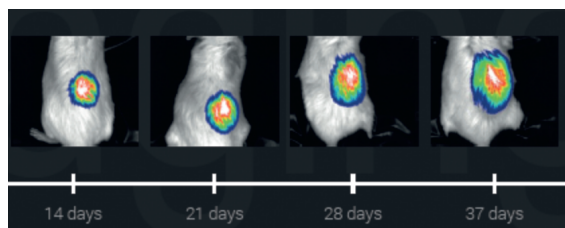
Fluorescence



Cell cultures

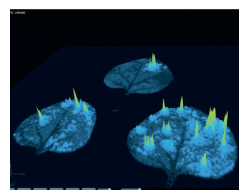
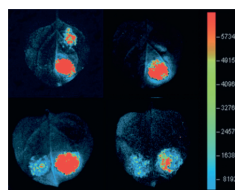


- Количественная оценка интенсивности биолюминесцентного или флуоресцентного сигнала
- Локализация биолюминесцентного сигнала на 3D-модели (биолюминесцентная томография) и количественный анализ интенсивности сигнала в тканях и органах
- Изучение развития опухолей и миграции клеток в динамике
- Оценка распределения терапевтического препарата в организме (всасывание, накопление, распределение)
- Визуализация развития воспалительных процессов и реакций организма на введение инфекционного агента



Эксперименты с растительными объектами (целые растения, листья или проростки):

- Оценка экспрессии генов по сигналу GFP
- Изучение воздействия УФ-облучения на ткани растений
- Изучение распространения фитопатогенов в растительных тканях;
- Возможность симуляции циклов день/ночь при постановке длительных экспериментов.



**Newton** – непревзойденная чувствительность и количественный анализ.

- Цифровой атлас органов и тканей
- Мультиплексные исследования в УФ, ИК, БИК и видимом свете (до 8 каналов возбуждения).
- Большое количество красителей и зондов: GFP, YFP, Pro-Q Emerald 300, Sypro-Ruby, FITC, DAPI, Alexa Fluor 680, 700, 750, Cy3, 5, 5.5, DyeLight, IRDye 800CW, VivoTrack 680, VivoTag 750 и т.д.
- Разрешение изображения – 10 Мп
- Светосильный объектив; охлаждение камеры -90 °С
- Моторизованная темная комната; подогреваемый столик
- Автофокус, автоэкспозиция, автоматическое управление с ПК



**Newton 7.0-BT500**  
**Newton 7.0-FT500**

биолюминесценция  
биолюминесценция, флуоресценция (8 источников IR, NIR, RGB и UV-света, 9 фильтров 535, 565, 595, 655, 710, 695, 750, 820, 840 нм);

**Newton 7.0-BIO FT500**

биолюминесценция, флуоресценция растений (источники света 440 и 480 нм), фильтры 535 и 565 нм, моторизованная по осям x, y, z платформа с возможностью вращения на 30°

## Анестезия лабораторных животных

- Единовременная и продолжительная ингаляция изофлураном или севофлураном
- Мыши, крысы, хомяки, кролики, морские свинки, кошки, собаки, обезьяны и т.д.
- Высокая точность подачи газа и безопасность в работе (цифрового расходомера и системой сбора отходов на основании угольного фильтра)
- Разработка моделей болезней животных (ишемия головного мозга, ишемия миокарда, остеопороз)
- Инъекции хвостовой вены
- Отбор проб крови из брюшной полости, аорты и сердца
- Инъекции лекарственного средства в желудочек
- Удаление ткани или органа и т.д.

Для удобства работы с разными животными используются специально разработанные маски, отдельные индукционные боксы разного размера, терморегулируемые столики, платформы для стереотаксиса и т.д.

Компактные и надежные устройства для работы автономно на столе:

**R540** – для животных до 7 кг, ингаляция изофлураном или севофлураном, RWD;

**R550** – для животных до 7 кг, ингаляция изофлураном или севофлураном, с возможностью одновременно подключить до 5 животных, RWD;

**Bioesthesia** – для животных до 7 кг, ингаляция изофлураном, адаптирована для использования и системами визуализации Newton, Vilber.

Мобильные системы на штативе с колесиками:

**R520** – для животных до 7 кг, ингаляция изофлураном или севофлураном;

**R620** – для животных до 50 кг, ингаляция изофлураном или севофлураном.



000 «Диаэм»

Москва

ул. Магаданская, д. 7, к. 3 ■ тел./факс: (495) 745-0508 ■ sales@dia-m.ru

www.dia-m.ru

**Новосибирск**

пр. Академика  
Лаврентьева, д. 6/1  
тел.  
(383) 328-0048  
nsk@dia-m.ru

**Казань**

ул. Парижской  
Коммуны, д. 6  
тел.  
(843) 210-2080  
kazan@dia-m.ru

**С.-Петербург**

ул. Профессора  
Попова, д. 23  
тел.  
(812) 372-6040  
spb@dia-m.ru

**Ростов-  
на-Дону**

пер. Семашко, д. 114  
тел.  
(863) 303-5500  
rnd@dia-m.ru

**Пермь**

Представитель  
тел.  
(342) 202-2239  
perm@dia-m.ru

**Воронеж**

Представитель  
тел.  
(473) 232-4412  
voronezh@dia-m.ru

**Армения**

Представитель  
тел.  
(094) 01-0173  
armenia@dia-m.ru

**Узбекистан**

Представитель  
тел.  
(90) 354-8569  
uz@dia-m.ru

# ООО «МАКСИМ МЕДИКАЛ»

Компания осуществляет поставку, обслуживание научного и медицинского оборудования и разрабатывает системы анализа и инспекции на базе машинного зрения и микросенсоров.

## Оборудование и реагенты для генетических исследований

**KASP™** – экономичная и точная аллель-специфическая ПЦР для поиска и валидации селекционно-значимых маркеров до SNP. Демо-наборы в наличии.

**Cleaver Scientific** – оборудование для электрофореза и сопутствующих продуктов, используемых в сфере медико-биологических исследований. Горизонтальные установки для гель-электрофореза, широкий спектр систем для вертикального электрофореза, компактные лабораторные источники питания и другое.



**TwistAmp и LavaLAMP** – реагенты для изотермической амплификации методом рекомбиназной полимеразной амплификации (РПА) и LAMP. Являются альтернативой ПЦР и позволяют диагностировать патогены и ГМО в течение 10-15 мин.

Демо-наборы в наличии.

**Axxin** – мобильные анализаторы для изотермической амплификации в полевых условиях.

Демо оборудование в наличии.



**Lucigen&Epicenter** – специальные ферменты, наборы реагентов для подготовки библиотек NGS (наборы для целевого обогащения, секвенирования одиночных клеток, циркуликазы, РНКаза-R, РНКаза-N и др.)

## Оборудование общего назначения

**HiPoint** – модульные камеры роста растений и растительных культур, теплицы.

- Широкий выбор источников LED-освещения, включая искусственный солнечный свет.
- Модульная и масштабируемая конструкция
- Автоматический контроль параметров: освещение, температура, влажность, CO<sub>2</sub>.



## Оборудование для спектрального анализа

**Spectricon** – мульти- и гиперспектральные камеры и микроскопы для спектрального анализа в диапазоне 370 - 1700 нм.

**QCELL** – оборудование на основе мультиспектральной визуализации Spectral Vision.

**NanoMabna** – микрообъемные спектрофотометры (аналог Nanodrop), анализаторы поверхностного плазмонного резонанса.

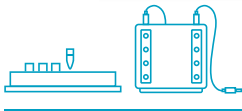
**Phenospex** – роботизированные системы фенотипирования растений от 10 см до 1.1 м на основе лазерных 3D сканеров Planteye F600. Автоматический анализ параметров: NDVI, EVI, цифровая биомасса, высота растения, наклон листа, площадь листовой поверхности, степень озеленённости и др.





Helicon offers a comprehensive range of laboratory equipment for scientific research in the fields of cell and molecular biology.

## Delivering world-class solutions for your laboratory!



*Helicon offers advanced solutions from top-level manufacturers and produces equipment for molecular biology.*

*For more details visit our website at [helicon.ru](http://helicon.ru)*



**DELIVERY**



**EDUCATION**



**TECHNICAL ADVICE**



**METHODOLOGICAL  
SUPPORT**

**Helicon is one of the leading Russian distributors of laboratory equipment, reagents, and plastic consumables in the areas of cell and molecular biology with 25 years of experience.**

**Our mission is to keep Russia at the forefront of scientific innovation!**

**We are committed to providing superior quality solutions for:**

- Cell and Molecular Biology
- Veterinary Medicine
- Food Safety
- Bioindustry
- Forensic Science
- Clinical Diagnostics

Our portfolio includes numerous well-known brands.

Helicon has its own R&D center and manufacturing divisions.

The company produces laboratory equipment and supplies under its own brand, including plastic consumables, vertical and horizontal electrophoresis chambers, gel documentation systems, laboratory furniture, tube racks, dispensers, etc.

One of the many advantages of working with Helicon is the opportunity to test products before making a purchase decision.

Helicon provides delivery and installation services free of charge.

Helicon has its headquarters in Moscow and maintains regional offices in Saint Petersburg, Novosibirsk, Kazan, Rostov-on-Don, Voronezh, and Yekaterinburg. Its established logistics network and warehouse facilities ensure timely delivery, even during peak periods.



**000 "Helicon Company"**

121374, Kutuzovsky, 88, Moscow, Russia

+7 499 705 50 50

mail@helicon.ru

[helicon.ru](http://helicon.ru)





# SE ANA

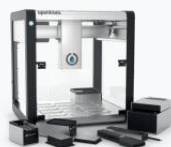
Лидер в области генетических исследований

Мы предлагаем для вас комплексные решения в области молекулярно-биологических исследований

## Обучение и услуги

- Секвенирование по Сэнгеру
- Фрагментный анализ
- Высокопроизводительное секвенирование
- Анализ данных
- Разработка тест-систем
- Консультирование и выполнение проектов

## Оборудование



### OPENTRON

Робот для выделения нуклеиновых кислот и подготовки NGS библиотек



### GENOLAB M

Высокопроизводительная система для секвенирования нуклеиновых кислот



### OPTOLANE

Анализатор [LOAA] (On-Point)  
Новое решение для цифровой ПЦР

И многое другое оборудование

## Реактивы



### RAISSOL

Наборы для выделения ДНК и её количественного определения.



### TECAN

Наборы для подготовки библиотек для NGS

**Разработка и производство компонентов и наборов для ПЦР и молекулярной биологии.**



- ◆ **Наборы для выделения ДНК и РНК**
- ◆ **Мастермиксы для проведения ПЦР**  
классическая ПЦР, ПЦР в режиме реального времени, ПЦР длинных фрагментов, ОТ-ПЦР одношаговым методом
- ◆ **Наборы для проведения обратной транскрипции**
- ◆ **Системы детекции РНК вируса SARS-CoV-2**
- ◆ **Маркеры молекулярных весов для ДНК**  
Ready-to-use, диапазон длин от 100 до 10000 п.н.

**Стандартные нуклеозидтрифосфаты** ◆  
dATP, dCTP, dGTP, dTTP и dUTP

**Модифицированные нуклеозидтрифосфаты (22 типа)** ◆

**Ферменты** Taq ДНК-полимераза, Hot-Start Taq ДНК-полимераза, ревертаза MoMLV, Bst ДНК-полимераза ◆

**Иммунохимические реактивы:** моноклональные антитела, ◆  
аффинноочищенные антитела, конъюгаты с пероксидазой хрена и биотином

**Олигонуклеотидный синтез:** праймеры и зонды для ПЦР, ◆  
модифицированные олигонуклеотиды

**Реактивы для мРНК:** m1ΨTP, ΨTP, m6ATP, m5CTP, ARCA ◆



**BIOSAN**

biosan-nsk.ru



**Biolabmix®**

biolabmix.ru

Тел.№: +7 (383) 363 22 40

e-mail: sales@biolabmix.ru

ГОТОВЫЕ РЕШЕНИЯ ДЛЯ МОЛЕКУЛЯРНОЙ БИОЛОГИИ

# НОВЫЙ НАБОР ДЛЯ ВЫДЕЛЕНИЯ ДНК ИЗ ОБРАЗЦОВ КРОВИ

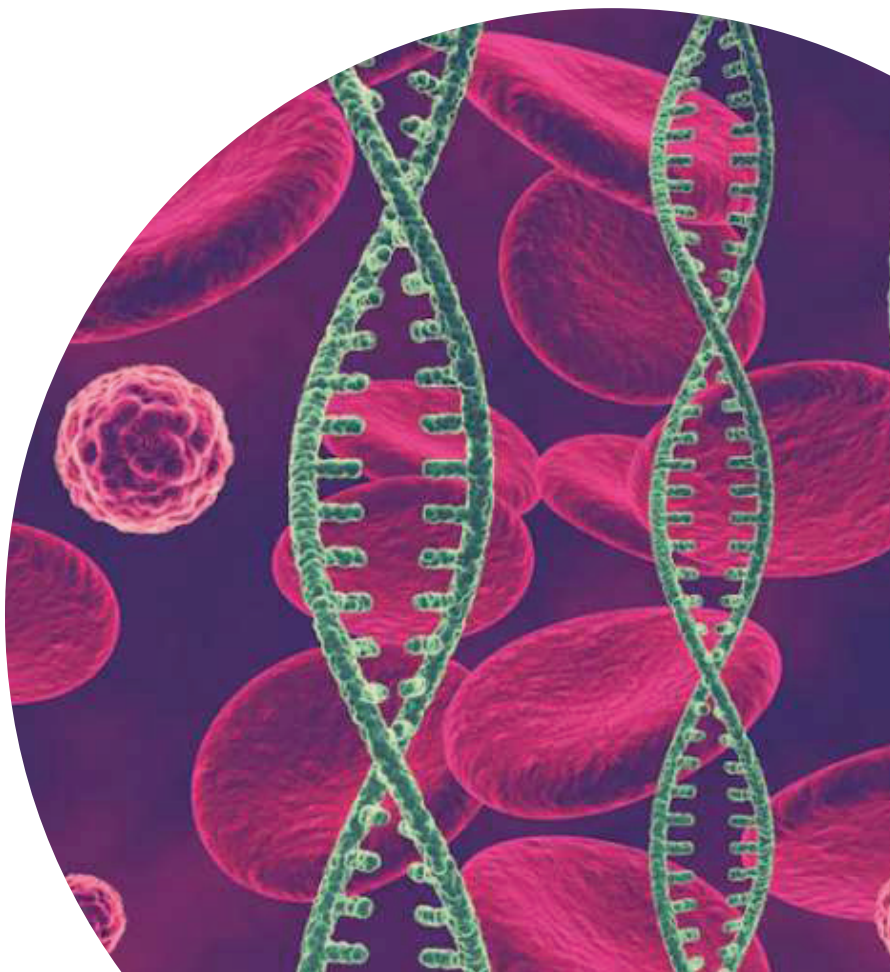
- ◆ Цельная кровь из пробирки с антикоагулянтами;
- ◆ Плазма крови;
- ◆ Сыворотка крови;
- ◆ Криопреципитат;
- ◆ Лейкоцитарная масса;
- ◆ Ликвор.

Электронный заказ  
на сайте производителя:



**Biolabmix®**

biolabmix.ru



# Contents



1	Symposium “Genomics, transcriptomics and bioinformatics”	<a href="#">21</a>
2	Symposium “Systems computational biology”	<a href="#">190</a>
3	Symposium “Structural biology and pharmacology: computational and experimental approaches”	<a href="#">261</a>
4	Symposium “Evolutionary, population and medical genomics/genetics of human”	<a href="#">377</a>
5	Symposium “Biotechnologies: computational and experimental approaches”	<a href="#">489</a>
6	Symposium “Genetics, bioinformatics and systems biology of plants”	<a href="#">592</a>
7	Symposium “Animal genetics, bioinformatics and systems computational biology”	<a href="#">667</a>
8	Symposium “Biomedicine, bioinformatics and systems biology”	<a href="#">754</a>
9	Symposium “Mathematics, bioinformatics and systems computational biology of COVID-19”	<a href="#">865</a>
10	Symposium “Cognitive sciences, neurogenetics, neuroinformatics and systems computational biology”	<a href="#">925</a>
11	Symposium “Systems biology and bioinformatics of DNA repair processes and programmed cell death”	<a href="#">968</a>
12	Section “Systems biology of aging: experimental and computational approaches”	<a href="#">1028</a>
13	Symposium “Big genetic Data Analysis, deep learning, mathematical modeling and supercomputing”	<a href="#">1082</a>
	Author index	<a href="#">1130</a>

# 1 Symposium “Genomics, transcriptomics and bioinformatics”



- 1.1 Section “Genomics and transcriptomics of plants and animals” [21](#)
- 1.2 Section “Regulatory genomics” [67](#)
- 1.3 Section “Functional and applied 3D genomics” [110](#)
- 1.4 Section “Evolutionary genomics, bioinformatics, and molecular phylogeny” [136](#)

## Genome-wide identification and expression analysis of lectins during cell development in elongation in maize root

Aglyamova A. \*, Gorshkova T.

*Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia*

\* [aliaglyamova@yandex.ru](mailto:aliaglyamova@yandex.ru)

**Key words:** plant lectins, maize, root

*Motivation and Aim:* Lectin domains are a group of different structural units of proteins capable to specifically recognize carbohydrate structures. They are found in combination with a wide variety of protein domains and are involved in various physiological processes. Although some lectin motifs are very similar in plants, animals and bacteria, most often they are associated with different domains and perform completely different functions. Plant lectins are quite divergent and can be localized in various cell compartments suggesting the differences in their function. In plants, lectins are quite numerous. The large proportion of lectins are transmembrane proteins; their lectin domains are oriented outside of plasma membrane and together with lectins fully located in apoplast have direct access to the cell wall polysaccharides. Several proteins with lectin domains from specific family of proteins with lectin domain were reported to be the possible elements to monitor the cell wall state and to evolve signals that regulate organ development [1, 3]. However, an involvement of proteins with lectin domains in the process of cell elongation that involves extensive cell wall rearrangements has not received yet the adequate and systemic attention. A renowned model to study plant growth is maize (*Zea mays* L.) primary root that is characterized by the presence of consecutive zones that differ by cell sizes and relative rates of their increase [2].

*Methods and Algorithms:* We performed a genome-wide screening for maize genes that encode proteins with lectin domains. The domain organization of proteins encoded by identified genes was characterized and amino acid sequences were checked both for the presence of a signal peptide and transmembrane domain. The putative subcellular localization of proteins with lectin domains was predicted. We also performed the multiple alignment of protein sequences and build phylogenetic trees for eight families of maize proteins with lectin domains. Using transcriptomic data obtained earlier [2] for 5 zones of the apical part of the primary maize root (before the root hairs initiation) we tried to relate the expression of these genes to certain stages of root cell development and peculiarities of cell wall metabolism.

*Results:* Screening of the maize genome revealed 307 genes that encode members of 15 different families of proteins with lectin domains. The half of these genes encode lectin receptor-like kinases (RLK). The largest maize families of lectin domain-containing proteins are GNA and legume, which is typical for plants. All members of maize RLKs were predicted to be membrane-bound with a lectin domain faced to the extracellular space. In addition to transmembrane proteins there are members of jacalin-related lectin, galactose-binding, hevein, RicinB and EUL families that are predicted as cell wall proteins.

In total, 70 % of all identified genes for proteins with lectin domains were expressed in the analyzed zones of maize root. For most part, genes within the same family have

diverse expression patterns. Each zone has its own set of specifically upregulated genes for proteins with lectin domains, both cell-wall related and non-related. The expression of most genes raised at the active elongation and the start of cell maturation and differentiation. Homologues of some characterized proteins with lectin domains of other species were identified in maize genome.

*Conclusion:* For the first time we revealed the genes for proteins with lectin domains that are specifically upregulated at initiation of cell elongation, are important in the course of active elongation or late elongation, opening the door for further studies on the molecular mechanisms of lectin involvement in plant morphogenesis.

*Acknowledgements:* The study is supported by Russian Science Foundation (project No. 20-64-47036).

### *References*

1. Guo H. et al. Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *PNAS*. 2009;106:7648-7653.
2. Kozlova L.V. et al. Elongating maize root: zone-specific combinations of polysaccharides from type I and type II primary cell walls. *Sci Rep*. 2020;10:1-20.
3. Li C. et al. FERONIA and her pals: functions and mechanisms. *Plant Physiol*. 2016;171:2379-2392.

## Design primers for LAMP-amplification

Akhmetzianova L.U.<sup>1,2\*</sup>, Davletkulov T.M.<sup>2</sup>, Garafutdinov R.R.<sup>3</sup>, Gubaydullin I.M.<sup>1,2</sup>, Chemeris A.V.<sup>3</sup>

<sup>1</sup> Institute of Petrochemistry and Catalysis, Ufa Federal Research Center RAS, Ufa, Russia

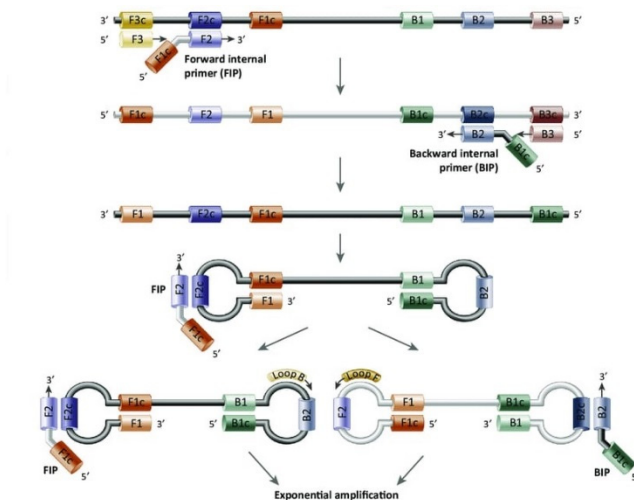
<sup>2</sup> Ufa State Petroleum Technological University, Ufa, Russia

<sup>3</sup> Institute of Biochemistry and Genetics, Ufa Federal Research Center RAS, Ufa, Russia

\* [www.lianab@mail.ru](mailto:www.lianab@mail.ru)

**Key words:** primer design, algorithm, Python, LAMP, software

For successful loop-mediated isothermal amplification (LAMP) [1] it is necessary to correctly select sets of at least four primers, and find their occurrences in the complete genome or target site. The selection of primers for LAMP is a non-trivial task due to the large number of primers and the length of the genome, and therefore special computer programs are required. This paper presents an overview of the existing most popular computer programs for the design of primers for LAMP. The main parameters that must be taken into account when selecting primers are described. The analysis of mathematical algorithms for searching for a sample in a string is carried out. To implement the search and selection of primers in the genome, a new algorithm for selecting primers based on direct search was developed. A new computer program “LAMPrimers iQ” has been developed that takes into account a number of requirements for the selection of primers, searches for them, analyzes compliance with recommended conditions and groups them into sets, taking into account the absence of homo- and heterodimers, as well as the minimum difference in the melting temperature of primers in one set. One set contains four primers: FIP (forward inner primer), F3 (forward outer primer), BIP (backward inner primer) and B3 (backward outer primer). Primers location scheme used to initiate the LAMP process is shown in the Fig. 1 [2].



**Fig. 1.** Primer location scheme for loop-mediated isothermal amplification (LAMP)



This application implements work with files (plain text format, FASTA, GenBank) of sequences, it is also possible to insert a sequence through the clipboard. The results can be saved in Excel format or continue working with another file or sequence.

The program is implemented in the Python 3.10 programming language using the Biopython library, which contains a number of different submodules for general bioinformatics tasks and allows you to use different genetic databases and the Qt framework to create applications with a graphical interface.

Further, it is planned to expand the functionality of the program to increase the specificity of the reaction, namely, to add loop primers (LoopF, LoopB) [3] and a hybridization probe with an independent annealing site. Test the obtained sets of primers in the laboratory.

*Acknowledgements:* The reported study was funded by RFBR according to the research project No. 20-37-90091.

### *References*

1. Notomi T., Okayam H., Masubuchi H., Yonekawa T., Watanabe K., Amino N., Hase T. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 2000;28(12):E63.
2. Alhassan A., Li Z., Poole C., Carlow C. Expanding the MDx toolbox for filarial diagnosis and surveillance. *Trends Parasitol.* 2015;31(8):391-400. doi: 10.1016/j.pt.2015.04.006.
3. Nagamine K., Hase T., Notomi T. Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Mol Cell Probes.* 2002;16(3):223-229.

## Relationship patterns of the genetic regulation of carotenoid and flavonoid accumulation in tomato fruits (*S. lycopersicum*)

Babak O.<sup>1\*</sup>, Anisimova N.<sup>1</sup>, Nekrashevich N.<sup>1</sup>, Yatsevich K.<sup>1</sup>, Drozd E.<sup>1</sup>, Fateev D.<sup>2</sup>, Soloviova A.<sup>2</sup>, Kurina A.<sup>2</sup>, Artemyeva A.<sup>2</sup>, Kilchevsky A.<sup>1</sup>

<sup>1</sup> Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus

<sup>2</sup> Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St.

Petersburg, Russia

\* O.babak@igc.by; babak\_olga@mail.ru

**Key words:** tomato, genetic regulation of the carotenoid and anthocyanin accumulation, MYB transcription factors

Carotenoid and flavonoid compounds of plants are secondary metabolites of terpenoid and phenolic nature respectively that perform important signaling and protective functions developed in the course of evolution, providing a wide degree of adaptability to the effects of abiotic and biotic stresses (resistance to excessive solar radiation, low temperatures, damage caused by insects, etc.). From the point of view of food production for humans, the composition of carotenoids and flavonoids determines the nutritional, taste and commercial qualities of products, as well as their purpose during consumption and processing. One of the most important properties of products determined by the level of accumulation of carotenoids and flavonoids is antioxidant activity, which determines a degree of the body's protection against oxidative stress, resistance to infections, and as a result, an increased life expectancy. In order to develop food products with maximum antioxidant activity, it is important to understand the relationship between substance accumulation mechanisms of the above classes.

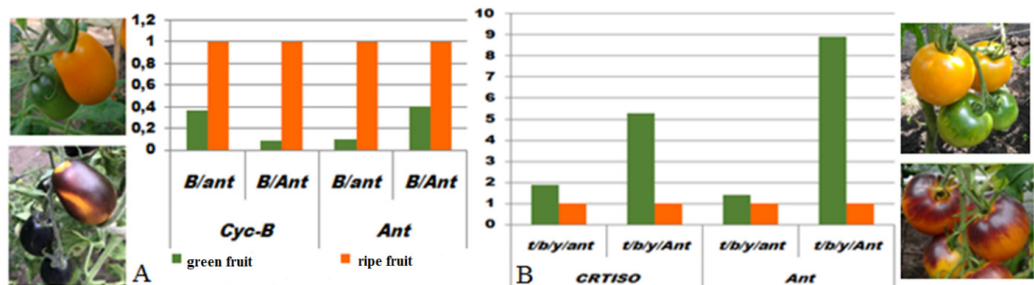
In order to study the relationship between the genetic regulation of carotenoid and flavonoid accumulation in tomato fruits, the following experiments were carried out: assessment of changes in the expression pattern of the structural *tangerine* (*t*) genes of carotenoid isomerase (CRTISO) and *Beta* (*B*) lycopene- $\beta$ -cyclase (*Cyc-B*), as well as determine carotenoid and anthocyanin concentration in tomato fruits depending on the allelic composition of carotenoid biosynthesis genes, the genes of MYB transcription factors (TF) *Yellow* (*Y/y*) SlMYB12 TF, *Anthocyanin 1* (*Ant1/ant1*) R2R3MYB TF, *Golden 2-like* (*U/u*) MYB TF regulating an expression level of the structural genes of flavonoid biosynthesis.

The study material was a collection of tomato accessions from the Institute of Genetics and Cytology of the National Academy of Sciences of Belarus with a different allelic combination determining flavonoid and carotenoid accumulation, as well as the coloration of fruits (r, at, t, og, ogc, B, gf-3, gf -5, hp-1, hp-2dg, u, gs, SlMyb12, and Ant1); a collection of tomato accessions with different fruit coloration from the All-Russian Institute of Plant Genetic Resources.

The study results related to the expression of alleles that determine the accumulation of carotenoids *tangerine* (*t*), and *Beta* (*B*) and anthocyanins *Anthocyanin1* (*Ant*) in tomato forms with their various combinations demonstrated the following:

- a different expression level at different stages of development depending on the combination of alleles: the maximum expression level of the *Ant* gene when combined

with the allele *t* at the stage of technical ripeness and in combination with the allele *B* at the stage of the complete ripeness of fruits (Fig. 1);  
 - a decreased expression level of the allele *t* of carotenoid isomerase at the stage of fruit ripening in forms with the alleles in their genotype that determine the high content of chalcone (*Y*) and anthocyanin (*Ant*) flavonoids.



**Fig. 1.** Relative expression level of alleles that determine the accumulation of carotenoids and anthocyanins in fruits at the stages of technical (green fruit) and biological (ripe fruit) ripeness in tomato forms with their various combinations: A – *Beta* (*B*) alleles of lycopene- $\beta$ -cyclase (*Cyc-B*) and *Anthocyanin 1* (*Ant*) R2R3MYB transcription factor (TF); B – *tangerine* (*t*) alleles of carotenoid isomerase (CRTISO) and *Anthocyanin 1* (*Ant*) R2R3MYB TF

Evaluation of carotenoid and anthocyanin accumulation in tomato fruits depending on the allelic composition demonstrated an increased concentration of pigments (chlorophyll, carotenoids, and anthocyanins) in the fruits with the *U* alleles of the *Golden 2-like*, *hp-2<sup>dg</sup>* of the DET1 nuclear protein gene, *gf-3* alleles, and *gf-5* of the Stay-Green gene in the genotype. The patterns of accumulation of carotene forms in tomato fruits depending on the alleles of the lycopene- $\beta$ -cyclase gene were confirmed: the maximum lycopene accumulation in forms with the *og<sup>c</sup>* allele, carotene – with the *B* allele, and the predominant accumulation of lycopene – in forms with the *b* allele. Nature of the relationship between the genetic regulation of flavonoid and carotenoid accumulation was shown – the maximum accumulation of anthocyanins in tomato fruits with *Ant1*, *Y*, and *U* alleles in the genotype is associated with the decreased concentration of lycopene in fruits; the allele *y* of the chalcone synthase gene increases the accumulation of lycopene in tomato fruits.

Based on the study results, tomato accessions with various combinations of the gene alleles of carotenoid and flavonoid accumulation were selected as models for the further study of the genetic regulation of pigment accumulation in fruits and use in the breeding process aimed to develop forms with a high level of antioxidant activity.

*Acknowledgements:* The work was supported by the Grant No. 20-516-00017Bel\_a, as well as the Grant B20R-285 of the Belarusian Republican Foundation for Fundamental Research “Study of genetic mechanisms to regulate the accumulation of anthocyanins and carotenoids in vegetable Solanaceae and Brassicaceae crops”.

## Evolution of gene exon-intron structure by alternative splicing

Babenko V.\*, Babenko R., Kudryavtseva N.N.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\*bob@bionet.nsc.ru

**Key words:** exon skipping, gene structure evolution, intron phase

*Motivation and Aim:* At least half of the AS events, manifested by exon skipping/insertion/change of splicing sites 5'-3', are known to result in premature stop codon emergence and are subject to nonsense-mediated decay (NMD) acting as additional means of rapid abolition of expression depending on the dynamics of the local compartment of protein metabolism. It is widely used to maintain homeostasis of multicomponent complex subunits such as ribosomes, spliceosomes, and chromatin remodeling mechanisms, and is especially pronounced in brain cells. However, of greatest interest is the result of splicing encoding a functional protein to assess the increase in diversity of a specific protein in a sample. One approach to filter out NMD-prone splice products is to target ribosome-engaged transcripts, which guarantees the coding ability of the mRNAs used (Furlanis et al., 2020). *Methods and Algorithms:* We arrange a pipeline for examining the exon-skipping events derived from RNA-Seq data.

*Results:* We observed specific distribution of coding vs non-coding skipping events in the brain tissues. We propose an efficient approach for the selection of coding transcripts beyond non-coding ones by translating the appropriate exon skipping events and validating them against an established protein database.

*Conclusion:* We managed to reduce the scan time to a reasonable range, thus proving the possibility of a wide application of the algorithm. We report the frame/phase distribution of alternative exon encoding underlining exon shuffling evolution mode featured the most in animal kingdom [1, 2]. No apparent long range phase compensation observed even in 'old' alternative exons.

*Acknowledgements:* The work was supported by the Russian Science Foundation project 19-15-00026-II to N.K., V.B.

### References

1. Rogozin I.B., Carmel L., Csuros M., Koonin E.V. Origin and evolution of spliceosomal introns. *Biol Direct.* 2012;7:11. doi: 10.1186/1745-6150-7-11.
2. Pathy L. Exon shuffling played a decisive role in the evolution of the genetic toolkit for the multicellular body plan of Metazoa. *Genes (Basel).* 2021;12(3):382. doi: 10.3390/genes12030382.

# Light perception in Beroidae ctenophores: evidence from laboratory experiments and genomics data

Baiandina Iu.\*, Kuleshova O., Krivenko O.

A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Sevastopol, Russia

\* [sepulturka@mail.ru](mailto:sepulturka@mail.ru)

**Key words:** photoreception, ctenophores, *Beroe ovata*, opsin genes

**Motivation and Aim:** Ctenophores are the basal Metazoa branch in which the evolution of the mechanisms of animals' light perception could be realized for the first time [2, 3]. However, there is a little bit of evidence about any aspects of light perception in ctenophores. Ctenophore behavioral responses to light were not known until recently, although photoreceptor-like structures have been described previously in several species. We have shown early that adult ctenophores *Mnemiopsis leidyi* respond to a sharp illumination change with a short-term increase in speed of movement [1]. Molecular aspects of light reception were studied on the same lobate ctenophore. In the *M. leidyi* genome, three opsin genes have been identified [2], two of them can be assigned to the group of ciliary opsins with a high confidence [3]. Here, we present the first experimental results on *Beroe ovata* response to high-intensity light exposure, as well as an analysis of the available transcriptomic data aimed to identify genes responsible for light perception in Beroidae.

**Methods and Algorithms:** Adult *Beroe ovata* were collected from the Black Sea nearshore waters (Martynova Bay, Sevastopol) in August 2021. Animals were placed in laboratory aquariums and the video recording of their behavioral responses to point illumination of various parts of their body (aboral organ, cten, and lobes) with lasers of violet, green, and red spectra was carried out. Video processing and analysis was conducted using Davinci Resolve video editor, ImageJ image analysis program with wrMTrack\_Batch plug-in, and CtenophoraTrack data analysis program.

The search for protein sequences that may relate to photoreception was performed based on the analysis of transcriptomes of four Beroidae species available at NCBI: *Beroe forskalii* /USA: South Carolina/ (SRA: SRR6074515); *Beroe ovata* /USA: South Carolina/ (SRR6074516), *Beroe sp.* UF-2017 /Australia: Moreton Bay, Queensland/ (SRR5892577), *Beroe sp.* UF-2017 /Antartica: Weddell Sea/ (SRR5892576), as well as own transcriptome data on *Beroe ovata* /Russia: Sevastopol/. Data pre-processing, 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) Databases are used for annotation, as well as the Opsin database formed on the basis of SwisProt and TrEMBL

**Results:** After illumination of the adult *B. ovata* aboral organ with a violet spectrum laser pronounced behavioral responses were observed. Within 1-5 sec after the starting of exposure, the ctenophores shrink the body and entrails near the statocyst and begin to move from the laser beam (at a speed of about 3 mm/s) rotating the body so that the laser does not hit the aboral organ. No pronounced reactions of *B. ovata* to high-intensity light

exposure on their ctenes and lobes were found. The reactions of ctenophores to red laser were less pronounced than the response to violet light exposure.

We have identified protein sequences that are highly homologous to proteins of the opsin family in the transcriptomes of all considered species. The best homology rates corresponded to the Opsin 1 and Opsin 3 of *M. leidy* and *Pleurobrachia bachei* ctenophores (more than 65 % identity with > 90 % coverage), as well as vertebrate Opsin 5 sequences (more than 40 % identity with > 70 % coverage). Based on it, we searched for loci in the genomes of *Beroe forskalii* (GCA\_011033025.1) and *Beroe ovata* (GCA\_900239995.1) containing the found protein-coding sequences.

*Conclusion:* Ctenophores have the ability to photoreception. The response of *Beroe ovata* to high-intensity violet light radiation is observed with a point impact on the aboral organ. Other parts of the ctenophores body are not sensitive to exposure to high-intensity light. Genes sequences that can respond to the photoreception in Beroidae have been obtained.

*Acknowledgements:* The study is supported by IBSS GA No. 121030100028-0.

### References

1. Baiandina Iu.S. et al. Black Sea *Mnemiopsis leidy* (Ctenophora) adult locomotion and light-induced behavior in laboratory experiments. *J Sea Res.* 2022;180:102152. doi: 10.1016/j.seares.2021.102152.
2. Schnitzler C.E. et al. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidy*: a new view of ctenophore photocytes. *BMC Biol.* 2012;10:107. doi: 10.1186/1741-7007-10-107.
3. Fleming J.F. et al. 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.

## QTLs associated with the duration of the developmental phases in wheat

Berezhnaya A.<sup>1\*</sup>, Kiseleva A.<sup>1,2</sup>, Salina E.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* BerezhnayaAA@bionet.nsc.ru

**Key words:** common wheat, heading time, maturity time, QTLs

**Motivation and Aim:** The adaptability and yield of wheat is strongly dependent on the flowering and maturity times. Flowering time has been shown to be affected by three major genetic systems: vernalization response (*Vrn*) loci, photoperiod response (*Ppd*) genes, and some other loci [1]. Meanwhile, knowledge about specific genetic factors controlling maturity time in wheat is limited. The aim of the study was to identify QTLs associated with the duration of the developmental phases in wheat using SNP markers.

**Methods and Algorithms:** The study was conducted on a F<sub>2</sub> mapping population developed from a cross between two spring wheat varieties – Novosibirskaya 31 and 124-1 line. A mapping population of 80 F<sub>2</sub> progenies and the two parents were genotyped at the TraitGenetics GmbH, Germany, with the the Illumina Infinium 25k Wheat array. The F<sub>2</sub> progenies along with the parental lines were phenotyped for the time of heading and maturity in the glasshouse. The F<sub>3</sub> progenies were divided into 2 groups and evaluated separately for flowering and maturity times under field conditions during 2020 and 2021 growing seasons. MultiPoint UltraDense was employed to construct the linkage map. Kosambi mapping function was used to convert recombination fractions to centimorgans [2]. QTL analyses were performed using the software package MultiQTL.

**Results:** In both environments parent 124-1 headed and matured later compared to parent Novosibirskaya 31: there was 8 days difference in the first experimental year and 13 days in the second. The F<sub>3</sub> progenies required 42-49 days to flowering and 81–95 days to maturity. Shapiro–Wilk tests confirmed normal distributions of all investigated traits in both environments. Heritability estimates were low for days to maturity (0.32) and moderate for days to heading (0.6) and vegetation period (0.57).

Traits	Parents		F <sub>2</sub> 2019		F <sub>3</sub> 2020		F <sub>3</sub> 2021		H <sup>2</sup>
	N31	L124-1	Average	Range	Average	Range	Average	Range	
Days to heading	34.3 ± 0.2	35.1 ± 0.3	31.8 ± 0.2	27.0–37.0	47.4 ± 0.2	43.1–52.9	46.3 ± 0.21	42.0–51.2	0.60
Days from heading to maturity	34.8 ± 0.2	41.5 ± 0.2	37.7 ± 0.3	32.0–47.0	37.8 ± 0.23	33.8–44.0	42.8 ± 0.24	38.7–49.0	0.32
Vegetation period	69.1 ± 0.3	76.6 ± 0.3	69.5 ± 0.4	63.0–78.0	85.2 ± 0.35	48.8–94.0	88.8 ± 0.35	82.8–95.9	0.57

No correlation was detected between days to heading and days to maturity.

A genetic map was constructed based on 4774 marker loci, including 814 skeletal markers, and spanned 2368 cM. The average distance between markers was 3.1 cM. Chromosome size ranged from 8.5 cM (chromosome 5D) to 212.9 cM (chromosome 5A). During the linkage map construction 25 linkage groups were developed. This linkage map did not represent 3D, 4D, 6D chromosomes, which is probably due to the lack of SNP polymorphism on the D genome. QTL analysis was performed using both composite interval mapping and multiple interval mapping.

We identified a total of 15 QTLs. Six QTLs were identified to control time to heading. The most significant QTLs were QHD.icg-7D and QHD.icg-2A, which had a LOD score > 3, and explained 20 and 19 % of the phenotypic variation respectively. Three QTLs on 5B, 6A, and 6B were associated with maturity time. The strongest QTL was QMT.icg-6A, which explained 21 % of the variance. The LOD score peak was in the *AX-158527841 – AX-109855005* interval (LOD = 3.5). Another six QTLs were associated with vegetation period. The most significant QTLs were located on chromosomes 1D, 5A, and 6B. QTL on 5B associated with maturity, and another two QTLs on 1D and 7B associated with vegetation period. These were moderate effect QTLs, which individually explained 14–18 % of the phenotypic variance.

*Conclusion:* The present study revealed several QTLs associated with the duration of developmental phases of wheat. The identified QTLs can be used in further studies on the genetic control of heading and maturity time.

*Acknowledgements:* This study was supported by the RFBR 20-016-00059.

### *References*

1. Snape J.W., Butterworth K., Whitechurch E. et al. Waiting for fine times: genetics of flowering time in wheat. *Euphytica*. 2001;119(1-2):185-90.
2. Kosambi D.D. The estimation of map distances from recombination values. *Ann Eugen*. 1944;12:172-175.



## Sequencing *Linum usitatissimum* L. genomic DNA extracted from nuclei

Dvorianinova E.M.<sup>1,2\*</sup>, Bolsheva N.L.<sup>1</sup>, Rozhmina T.A.<sup>1,3</sup>, Sigova E.A.<sup>1,2</sup>,  
Pushkova E.N.<sup>1</sup>, Melnikova N.V.<sup>1</sup>, Dmitriev A.A.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup> Federal Research Center for Bast Fiber Crops, Torzhok, Russia

\* [dvorianinova.em@phystech.edu](mailto:dvorianinova.em@phystech.edu)

**Key words:** flax, high-molecular-weight DNA, whole genome sequencing, Nanopore

*Motivation and Aim:* The genetics of agricultural plants has shown the importance of assembling accurate and contiguous genomes [1]. These sequences form the basis for studying crucial agricultural traits at the molecular level. However, poor assembly quality undermines the validity of further genetic analysis [2]. Thus, obtaining high-quality plant genomes with available techniques is a relevant challenge. The current work aimed at producing a high-quality genome sequence of a flax (*Linum usitatissimum* L.) variety using the Oxford Nanopore Technologies (ONT) platform, which requires extremely pure DNA.

*Methods and Algorithms:* We adapted a plant nuclei isolation protocol for flax. The organelles were obtained from the *L. usitatissimum* line #3896 (provided by the Institute for Flax) leaves. Pure high-molecular-weight DNA was extracted from the nuclei with the Nanobind Plant Nuclei Big DNA Kit (Circulomics, USA). The SQK-LSK109 kit was used to prepare 1D genomic DNA libraries for sequencing on the ONT platform (MinION, R9.4.1 flow-cells). The received fast5 reads were basecalled with Guppy 5.0.11 (`dna_r9.4.1_450bps_sup.cfg`); the adapters were removed with Porechop. *L. usitatissimum* line #3896 genome was assembled with Canu 2.2 (`genomeSize = 400m`) and polished with Racon (two iterations) and Medaka [3–5]. Genome statistics were evaluated using QUAST and BUSCO (`eudicots_odb10` dataset) [6, 7]. Cytosine methylation levels and methylation frequencies were quantified with Megalodon and the `megalodon.sh` script from METEORE respectively [8, 9].

*Results:* The developed nuclei isolation protocol allowed us to extract pure high-molecular-weight DNA and perform two successful sequencing runs on the ONT platform. We received 14.5 and 16.5 Gb of raw fast5-reads with an N50 of 12.4 and 15.7 kb respectively. Having merged the two datasets, we assembled and polished the genome sequence of *L. usitatissimum* line #3896. The assembly demonstrated 93.8 % BUSCO completeness and consisted of 1697 contigs with a total length of 447.0 Mb and an N50 of 6.2 Mb. In the assembled genome, cytosine nucleotides were modified in the CG, CHG, and CHH contexts (Table 1). Nearly a half of the CG cytosines were highly methylated ( $\geq 50\%$ ). In contrast, only 0.02 % of the CHH cytosines had high methylation frequency.

**Table 1.** Cytosine methylation in the genome assembly of *L. usitatissimum* line #3896

Methylation context	CG	CHG	CHH
Context abundance, % of the called CN sites	17.0	11.9	71.1
Percentage of sites with high methylation levels ( $\geq 50$ %)	53.8	2.8	0.02

**Conclusion:** The employed method of flax nuclei isolation enabled us to extract pure high-molecular-weight DNA of *L. usitatissimum*. The received amount of ONT data was adequate to produce an assembly with acceptable QUASt and BUSCO statistics. Our dataset can be used in molecular genetic research on flax, including such genome regulation studies as methylation analysis.

**Acknowledgements:** The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. 075-15-2021-1064.

### References

1. Li G. et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat Genet.* 2021;53(4):574-584.
2. Van Bel M., Bucchini F., Vandepoele K. Gene space completeness in complex plant genomes. *Curr Opin Plant Biol.* 2019;48:9-17.
3. Koren S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722-736.
4. Vaser R. et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017;27(5):737-746.
5. <https://github.com/nanoporetech/medaka>
6. Gurevich A. et al. QUASt: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072-1075.
7. Simão F.A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210-3212.
8. Yuen Z.W.-S. et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat Commun.* 2021;12(1):1-12.
9. <https://github.com/nanoporetech/megalodon>

## Longread-only approach to the organellar genome assembly of a rare endemic non-model species *Crepis callicephala* Juz. (Asteraceae)

Emirsaliiev A.<sup>1,2\*</sup>, Tsyupka V.<sup>1,2</sup>, Grebennikova O.<sup>1,2</sup>, Bulavin I.<sup>1,2</sup>, Krivenko O.<sup>1,2</sup>, Ivanova N.<sup>2</sup>, Nikiforov A.<sup>2</sup>, Mitrofanova I.<sup>1,2,3</sup>

<sup>1</sup> Kurchatov Genomic Centre – NBG-NSC, Yalta, Russia

<sup>2</sup> FSFIS "Nikita Botanical Gardens – National Scientific Center of RAS", Yalta, Russia

<sup>3</sup> FSBIS Main Botanical Gardens named after N.V. Tsitsin, Moscow, Russia

\* [asantaraqtasli@gmail.com](mailto:asantaraqtasli@gmail.com)

**Key words:** cpDNA, plastid genome, chloroplast, endangered species, adaptation, Crimea

*Motivation and Aim:* *Crepis callicephala* Juz. is a rare endemic species of Crimean flora. It belongs to the Asteraceae family – one of the most important for human and widespread eudicots family in the world. The *Crepis* genus had a great impact on the plant embryology and cytogenetics as a useful object for chromosome observation [1]. But its genome has not been in detail investigated yet via molecular methods. The Crepidinae subtribe members of Asteraceae, including the genus *Crepis*, are the perspective source of different biologically active compounds of a high interest, as well some species of their parental tribe are included into local medicinal plant lists and pharmacopoeia. *C. callicephala* and relative species demonstrate notable niche specialisation [2] that also accords to the high level of endemism in this genus. Plastid genome codes a set of crucial photosynthetic genes [3] that contribute to the adaptive evolution. Being the wild relative of such important cultivated plants as *Lactuca* and *Cichorium*, the *Crepis* sp.sp. can be useful for plant geneticists to better understand adaptive abilities of that plants. As the modern omics approaches (including genomic researches) still has been unclear for the *Crepis* genus, we aimed to sequence and assemble the chloroplast genome of rare endemic Crimean species *C. callicephala*. The modern single molecule sequencing methods would be a good choice for that purposes due to relatively low cost and ability to resolve complex structure of plant organellar genomes in an assembly stage. Also we aim to compare different assembly methods in this application.

*Methods and Algorithms:* Seeds of *C. callicephala* was harvested previously in natural populations. *In vitro* plants were obtained by induction of multiple adventive proliferation from upper part of seedlings. DNA extraction, genomic library preparation and sequencing have been previously described [4]. Raw genomic reads base called with ONT Guppy v4.0.15 and v4.0.19, reads' statistics was scored using MinIONQC [5] NanoPack and LongQC packages. Adapters were trimmed with Porechop, reads were filtered using NanoFilt and bioawk.

For assembly, we initially selected the corresponding to plastids fraction of long reads. Reads have been aligned with minimap2. Unique mapped reads have been extracted using samtools.

Genome assembly was based on different reference-guided and *de novo* approaches. We used *de novo* tools such as Canu, flye, pomoxis (*de novo*) and NECAT [5], and also we

used pomoxis in a reference-guided mode. Assemblies were polished with long reads only using medaka. The final annotation was performed via GeSeq service [6] with enabling the CDS, tRNA and rRNA annotations on BLAT v36x7 (with search against Crepidinae and several other Cichorieae plastomes), HMMER v3.3.1, Chloë v0.1.0 and tRNAscan-SE v2.0.7.

**Results:** To examine the quality of selection previously extracted reads was one more realigned resulting in a good alignment with 94.34 % long reads mapped and total coverage depth x1,354.7. The longest mapped read was 124,313 bp in length with average 3,059 bp. Different assembly pipelines were performed. Some of them produced a single contig of expected length (NECAT and pomoxis in reference-guided mode), but the others (Canu, flye, pomoxis (*de novo*)) resulted either with very short or oversize and fragmented assemblies. Produced by NECAT and pomoxis assemblies had different total lengths, inverted repeat sizes and *ycf1* gene sizes. To avoid such bias, we selected reads contained homologous to plastome sequences one more again. New reference was constructed from three plastomes, each of them was linearly concatenated – the one belonged to *L. sativa*, and the others were produced by pomoxis-medaka pipeline both in reference-guided and *de novo* modes. Mapped reads were used for assembly with pomoxis-medaka on *L. sativa* reference. Resulting plastome of *C. callicephala* had length 149,463 bp and contained 35 tRNA, 120 protein-coding genes and 13 putative pseudogenes.

**Conclusion:** Draft plastome sequence of rare endemic Crimean species *C. callicephala* Juz. was obtained from total genomic nanopore reads based on longread-only assembly approach. None of widely used *de novo* assembly tools worked well for this purpose.

**Acknowledgements:** The research was done within the framework of the grant assignment No. 075-15-2019-1670 and conducted on the base of the Unique Scientific Installation "Scientific Center of Plant Biotechnology, Genomics and Conservation" of FSFIS "NBG-NSC" and facilities of Shared Access Center Bioinformatics of FRC Institute of Cytology and Genetics of Siberian Branch of the RAS.

## References

1. Navashin M.S. Karyology and Cytogenetics problems in Crepis genus research. Moscow: Nauka, 1985.
2. Nikiforov A.R. Aging population repertoire features of *Lagoseris callicephala* Juz. (Asteraceae). *Bull. State Nikita Botan Gard.* 2017;125:83-87.
3. Skobeyeva V., Omelchenko D., Logacheva M. et al. Plastid genome evolution in the genus *Allium*. In: Bioinformatics of genome regulation and structure/systems biology (BGRS/SB-2020): The Twelfth International Multiconference. Abstracts. Novosibirsk, 2020;243. doi: 10.18699/BGRS/SB-2020-154.
4. Emirsaliiev A.O., Brailko V.A., Grebennikova O.A. et al. Genomic research of the rare relict endemic plant *Crepis callicephala* Juz. (Asteraceae). In: Proceedings of the First All-Russian scientific and practical conference with international participation «Genomics and modern biotechnologies in plant propagation, breeding and preservation» October, 27–31, 2020. Simferopol: ARIAL, 2020;73-74. doi: 10.47882/GENBIO.2020.52.77.030.
5. Chen Y., Nie F., Xie Sh.-Q. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun.* 2021;12(1):1-10. doi: 10.1038/s41467-020-20236-7.
6. Tillich M., Lehwark P., Pellizzer T. et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucl Acids Res.* 2017;45:W6-W11. doi: 10.1093/nar/gkx391.

# Transcriptomics as an effective tool to study plant cell wall formation and function

Gorshkova T.

*Kazan Institute of Biochemistry and Biophysics, FRC KazRC RAS, Kazan, Russia*

\* [gorshkova@kibb.knc.ru](mailto:gorshkova@kibb.knc.ru)

**Key words:** plant cell wall, polysaccharides, transcriptomics, transcription factors, lectins, tissue-specific expression

*Motivation and Aim:* Plant cell wall is the most abundant organic matter on our planet and the major renewable recourse. Together with that, this compartment largely determines the peculiarities of plant biology. All that warms up the interest to this complex supramolecular structure and demands for the approaches for its purposeful design. However, plant cell walls are mainly composed of polysaccharides – complex polysaccharides, most of which have branched irregular structure with numerous types of linkages and are extremely difficult to analyze. Moreover, polysaccharides are synthesized by non-template mechanism with participation of large membrane-bound enzymatic complexes, many participants of which have not been fully identified yet. Besides, the *in muro* modifications of cell wall polysaccharides are performed by the enzymes encoded by huge multigene families. The effective tool to figure out the general picture of cell wall dynamics and the important molecular players is transcriptomics.

*Methods and Algorithms:* We have developed several plant models that permit to study tissue- and stage-specific processes occurring in cell walls of various tissues and performed corresponding transcriptomic analysis followed by the complementary bioinformatics [1, 2].

*Results:* The results will be presented on the example of flax phloem fibers – the crucial elements of plant biomechanical system – that in the course of cell development consequently deposit three different cell wall types performing corresponding changes in cell wall machinery. The analysis of transcriptomic data was considerably facilitated by the developed online platform FIBexDB [3] that also helped to characterize individual genes and reveal their coexpression. Though the transcript amount of a certain gene is not always proportional to the amount of corresponding protein and especially to its activity, the changes in transcriptome patterns are very informative to realize the peculiarities of general cell metabolism and to detect the key molecular players, with various cell wall related enzymes and the involved transcription factors among them. Of special interest are proteins with lectin domains that interact with specific cell wall polysaccharides and help to trace the cell wall state [4].

*Conclusion:* Transcriptome analysis, especially if combined with immunocytochemistry and atomic force microscopy, helps to considerably advance the understanding of the cell wall polysaccharide fine-tuning in everyday life of a plant cell, its importance and possibilities to regulate.

*Acknowledgements:* The study is supported by the RSF grant No. 19-04-00361.

## References

1. Gorshkova T. et al. Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci Rep.* 2018;8:14570.
2. Kozlova L.V. et al. Elongating maize root: zone-specific combinations of polysaccharides from type I and type II primary cell walls. *Sci Rep.* 2020;10:1-20.
3. Mokshina N. et al. FIBexDB: a new online transcriptome platform to analyze development of plant cellulosic fibers. *New Phytologist.* 2021;231:512-515.
4. Petrova N. et al. Gene expression patterns for proteins with lectin domains in flax stem tissues are related to deposition of distinct cell wall types. *Frontiers Plant Sci.* 2021;12:634594.

## Isolation of high quality RNA from *Phytophthora infestans* for transcriptome analysis

Ivanov A.<sup>1,2\*</sup>, Golubeva T.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* a.ivanov2@g.nsu.ru

**Key words:** late blight, plant pathogen, transcriptome analysis, RNA isolation

**Motivation and Aim:** *Phytophthora infestans* is one of the main pests of agricultural plants of the *Solanaceae* family, especially potatoes: the cost of combating against pathogen can be up to 30 % of the product's final price. Studies of its transcriptome, especially transcriptome changes during the colonization of the host plant, can be very useful for the development of new highly selective and eco-friendly drugs for pest control. For such studies, it is necessary to develop methods for isolating high-quality RNA from *P. infestans* samples, which is significantly complicated by the presence of a rigid cell wall and aggressive RNAses. The aim of our work was to study the effect of various parameters on the quality of total RNA isolated from samples of *P. infestans*.

**Methods and Algorithms:** To achieve this goal, we identified several key parameters that affect the integrity of total RNA during isolation: this is the process of sample preparation and grinding, as well as the use of RNA stabilizing solutions. The Plant MiniKit and Qiazol (Qiagen) were used to purify the RNA fraction. To optimize the protocol, we varied the above parameters and examined the obtained RNA using an RNA analyzer, the quality of the RNA was assessed based on the RNA integrity number (RIN).

**Results:** RNA isolation should be carried out immediately after collection of *P. infestans* samples. Grinding *P. infestans* is best done with Tissue Lyser. However, the original Qiagen lysis buffers are insufficient to prevent RNA degradation during the homogenization. Best results were achieved with Qiazol added to the samples prior to the homogenization. The samples were then proceeded according to Qiazol protocol. Qiagen spin columns were used for RNA isolation after phase separation step. Tissue Lyser grinding time is critical: RIN was decreased with longer grinding.

**Conclusion:** We have developed a protocol that allows isolate the total RNA from *P. infestans* with a sufficiently high quality, which is suitable for RNA-seq analysis (RIN  $\geq$  7).

**Acknowledgements:** The work was supported by a grant from the President of the Russian Federation for state support of young Russian scientists (MK-4311.2022.5).

## Insights of the genome evolution in the *Nicotiana* genus

Ivanov N.V.

PMI R&D, Philip Morris Products S.A., Neuchatel, Switzerland

nikolai.ivanov@pmi.com

**Key words:** *Nicotiana*, tobacco, genomics, sequencing

*Motivation and Aim:* *Nicotiana* genus includes ~90 naturally occurring species divided in 13 sections. The diversification between the *Nicotiana* genus, containing common tobacco (*Nicotiana tabacum*), and the *Solanum* genus, which includes tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*), is estimated to occur 24 million years ago. Although majority of *Nicotiana* species originate in South America, the Suaveolentes section constitutes the largest subsection in the *Nicotiana* genus with 39 species distributed in Africa (1 species), Pacific Islands (3 species) and Australia (35 species).

*Methods and Algorithms:* We sequenced and analyzed the genomes of numerous *Nicotiana* species using the Illumina and Oxford Nanopore sequencing technology with at least a sequencing coverage of 20X representing different chromosome number and different biogeographic regions.

*Results:* In this work the latest state of the *Nicotiana* genomics is presented and the insights to their genome evolution are discussed.

# Viral metagenomic analysis of publicly available genomic and transcriptomic samples from Simuliidae and Ceratopogonidae insects

Kamanova E.P.<sup>1,2\*</sup>, Starchevskaya M.E.<sup>1,2</sup>, Antonets D.V.<sup>1,2</sup>

<sup>1</sup> State Research Center of Virology and Biotechnology “Vector”, Novosibirsk, Russia

<sup>2</sup> Novel Software Systems LLC, Novosibirsk, Russia

\* kamanovae@gmail.com

**Key words:** metagenomics, *Nematocera*, viruses of insects

*Motivation and Aim:* Insects make the largest group of multicellular organisms. They also often act as carriers of numerous infection diseases of plants, animals, and humans. Blood-sucking insects from the suborder Nematocera deserve special attention, as they can transmit dangerous human viral infections such as Zika and Dengue fever, yellow fever and many more. Viral infections of livestock animals, transmitted by blood-sucking insects, can cause serious economic losses. To date, the Culicidae family is the most well-studied family of *Nematocera* insects. And there are numerous studies throughout the world dedicated to mosquito-borne viral infections. However, other important groups of *Nematocera* insects such as Simuliidae and Ceratopogonidae families as well as viruses associated with them remain largely unknown. Nevertheless, it has been shown that the members of Simuliidae family can transmit vesicular stomatitis virus to domestic pigs [1]. They were also found to play a role in transmission of Venezuelan equine encephalomyelitis virus and myxomatosis virus [2].

*Methods and Algorithms:* We decided to investigate the available datasets to identify the sequences belonging to known and new viruses. From the NCBI SRA database we collected 8 genomic and 3 transcriptomic samples of Simuliidae family and 8 genomic and 144 transcriptomic samples of Ceratopogonidae family. The analysis of Simuliidae samples was started. The sequencing reads were classified against the NCBI Nucleotide database (nr) with Kraken2 software [3]. More than 20 % of reads were remained unclassified. The unclassified reads and those attributed to viruses were extracted and assembled with metaSPAdes [4]. The resulting contigs with a coverage of no less than 20 and with a length of 500 bp or more were used for functional analysis. The open reading frames were predicted with Prodigal [5].

*Results:* The obtained amino acid sequences were analyzed for conserved domains and several genes were proposed to encode proteins specific for nucleopolyhedroviruses belonging to Baculoviridae family known to infect different groups of insects. Many sequences were attributed to bacteriophages of Myoviridae family and unclassified dsDNA viruses and a large number of contigs was still not attributed to any taxon.

*Conclusion:* Metagenomic studies are very important as they can provide us with better understanding microorganisms diversity, ecology and evolution and also help us to identify new pathogens and their transmission routes. We believe that metagenomic studies of Simuliidae and Ceratopogonidae species are of critical importance. However, the available genomic and transcriptomic data for these insects is scarce, and especially, for Simuliidae family.



*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-15-2019-1665).

### *References*

1. Mead D.G. et al. Biological transmission of vesicular stomatitis virus (New Jersey serotype) by *Simulium vittatum* (Diptera: Simuliidae) to domestic swine (*Sus scrofa*). *J Med Entomol.* 2004;41:78-82.
2. Manguin S. et al. Main Topics in Entomology: Insects as Disease Vectors. In: Lopez O., Fernandez-Bolanos J. (Ed.). *Green Trends in Insect Control*. RSC Green Chemistry, 2011:1-52.
3. Wood D.E. et al. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):257.
4. Nurk S. et al. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27(5):824-834.
5. Hyatt D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119.

## Meta-analysis of transcriptomic changes in prefrontal cortex after chronic social defeat stress

Kisaretova P.<sup>1,2\*</sup>, Reshetnikov V.<sup>1</sup>, Bondar N.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* [kisaretova@bionet.nsc.ru](mailto:kisaretova@bionet.nsc.ru)

**Key words:** chronic social defeat stress, prefrontal cortex, transcriptome

**Motivation and Aim:** Chronic social defeat stress (CSDS) is a valid classic model of depression in rodents. The use of animal models of depression opens new horizons in research of molecular brain changes.

**Methods and Algorithms:** Here we report a meta-analysis of transcriptomic changes in prefrontal cortex (PFC) of C57BL/6 male mice. We searched for all case-control studies on CSDS with 10 or 30 days duration that reported RNA sequencing measurements on PFC in C57BL/6 male mice and integrated it with our own transcriptomic data.

We analyzed stress effects after 10 days of CSDS (5 datasets), 30 days of CSDS (2 datasets) and compared resilient and susceptible to stress animals that were identified in 2 studies after 10 days of CSDS. RNA-seq data were subjected to standard pipeline: preprocessed with fastp v0.20.1, mapped to GRCm38 using HISAT2 aligner v2.2.1 and quantified with featureCounts v2.0. Differential gene expression analysis was performed via the DESeq2 R-package. Genes with  $p\text{-adj} < 0.05$  were designated as differentially expressed genes (DEGs). GSEA was conducted using gseGO function from ClusterProfiler (v4.0.5) R package.

**Results:** We found 17 DEGs (11 upregulated and 6 downregulated) in CSDS10 group compared to control. 10 days of CSDS lead to upregulation of genes involved in neuroinflammation (*Vwf*, *Il1r1*, *Il6ra*), oxidoreductase activity (*Hbb-bt*, *Hbb-bs*) and melanocortin receptor 4 gene (*Mc4r*) involved in energy homeostasis. Change in expression of glucocorticoid responsive genes wasn't unidirectional (*Fkbp5* upregulated, *Hsd11b1* downregulated). GSEA showed that 10 days of CSDS resulted in activation of gene expression related to social behavior, regulation of T cell cytokine production and neural tube patterning.

When comparing resilient and susceptible to stress mice we identified 92 DEGs (23 upregulated and 69 downregulated). Most downregulated genes in susceptible compared to resilient mice were oligodendrocyte-specific. Susceptible mice were shown to have lowered expression of myelin-related genes – structural (*Mbp*, *Mal*, *Mobp*, *Plp1*), enzymes (*Ugt8a*), transcription factor – *Olig2* etc.

The biggest changes we saw were in CSDS30 group with 580 DEGs (303 upregulated and 277 downregulated). Upregulated genes are involved in the following processes – transcription (*Polr2k*, *Polr2g*, *Rbbp4*), translation (*Eif1*, *Eif2s2*), alternative splicing (*Snrpa1*, *Snrpd1*, *Snrpd2*, *Snrpe*, *Snrpg*) and ATP synthesis (*Atp5g3*, *Atp5e*, *Atp5j*, *Atp5k*, *Atp5l*, *Atp5o*). While downregulated genes are involved in NMDA receptor signaling (*Grin1*, *Grin2c*, *Camk2g*), neuronal trafficking (*Htt*, *Disc1*) and steroid receptor modulators (*Ncor2*, *Ncoa3*, *Bcor*).

*Conclusion:* Collectively, our results on PFC gene expression changes suggest that (1) the most persistent expression changes induced by 10 days of CSDS are associated with glucocorticoid signaling, neuroinflammation, and oxidoreductase activity; (2) susceptibility to 10 days of CSDS is strongly associated with a decrease in the expression of key genes associated with myelination processes; 3) 30 days of CSDS lead to changes in energy-consuming processes such as transcription and translation, the activation of these processes is probably required for adaptation at the cellular level under conditions of prolonged chronic stress.

*Acknowledgements:* The study is supported by Russian Science Foundation (grant 21-15-00142).

## Analysis of multiple transcription factors binding in the model plant genomes

Kononov V.A.<sup>1</sup>, Dergilev A.I.<sup>2</sup>, Orlova N.G.<sup>3,4</sup>, Orlov Y.L.<sup>1,5\*</sup>

<sup>1</sup> Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Financial University under the Government of the Russian Federation, Moscow, Russia

<sup>4</sup> Moscow State Technical University of Civil Aviation, Moscow, Russia

<sup>5</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

\* orlov-yul@rudn.ru

**Key words:** plant biology, genomics, transcription factor binding, ChIP-seq, databases, *Arabidopsis thaliana*

**Motivation and Aim:** The development of high-throughput genomic sequencing coupled with chromatin immunoprecipitation technologies allows studying the binding sites of the protein transcription factors (TF) in the genome scale. The growth of data volume on the experimentally determined binding sites raises qualitatively new problems for the analysis of gene expression regulation, prediction of transcription factors target genes, and regulatory gene networks reconstruction. Genome regulation remains an insufficiently studied though plants have complex molecular regulatory mechanisms of gene expression and response to environmental stresses. It is important to develop new software tools for the analysis of the TF binding sites location and their clustering in the plant genomes, visualization, and the following statistical estimates.

**Methods and Algorithms:** We used experimental ChIP-seq data to model TF binding as the alternative to sequence-based prediction [2, 4]. This study presents application of the analysis of multiple TF binding profiles in evolutionarily distant model plant organisms. The construction and analysis of non-random ChIP-seq binding clusters of the different TFs in mammalian embryonic stem cells were discussed earlier using similar bioinformatics approaches and computer scripts [3]. We used ChIP-seq data in several plant plants, including *Arabidopsis thaliana*, *Physcomitrella patens*, and *Chlamydomonas reinhardtii* to model TF binding clusters in their genomes and estimate the distribution of the TF in the clusters found [2]. The data on TF binding were collected from GEO NCBI, PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>, <http://ucsc.gao-lab.org/>), PlantRegMap (<http://plantregmap.cbi.pku.edu.cn>).

**Results:** The clusters of TF binding sites may indicate the gene regulatory regions, enhancers and gene transcription regulatory hubs. Further, it can be used for analysis of the gene promoters as well as a background for transcription networks reconstruction. The heatmaps of TF binding in *Arabidopsis* allow delineating two major groups of factors: (GATA, ARF, BBR-BPC, HSF, LBD, FAR1, C2H2, GRAS, MYB, TCP, E2F/DP, HD-ZIP, CH3, NAC, AP2, bZIP, WRKY) and (B3, G2-like, ZF-HD, MADS, DOF, YABBY), or, in other words, relating to the GATA and related G2-like transcription factors. We discuss the statistical estimates of the TF binding sites clusters in the model plant genomes. The distributions of the number of different TFs per binding cluster follow same power law distribution for all the genomes studied. the evolutionarily

more ancient factors GATA and MYB are presented in the clusters of sites in all studied plant species. The binding clusters in *Arabidopsis thaliana* genome will be discussed in detail.

*Conclusion:* We show that regulatory gene networks structure in plants corresponds to similar transcription factor binding in plant genomes. The statistical estimates of the transcription factor binding clusters in genomes could be extended to other evolutionary not related genomes. Integration of experimental genomic information, Big Data, in general, represents an important problem in bioinformatics, requiring the integration of existing software tools and approaches [1, 5]. The presented statistical analysis of clusters of binding sites in plant genomes helps in solving the problem, developing approaches to the study of the evolutionary origin of enhancers and gene networks

*Acknowledgements:* This publication has been supported by the RUDN University Scientific Projects Grant System, project No. R.3-2022-ins).

### References

1. Chen M., Harrison A., Shanahan H., Orlov Y. Biological big bytes: Integrative analysis of large biological datasets. *J Integr Bioinform.* 2017;14(3):20170052. doi: 10.1515/jib-2017-0052.
2. Dergilev A.I., Orlova N.G., Dobrovolskay O.B., Orlov Y.L. Statistical estimates of multiple transcription factors binding in the model plant genomes based on ChIP-seq data. *J Integr Bioinform.* 2021;19(1):20200036. doi: 10.1515/jib-2020-0036.
3. Dergilev A.I., Spitsina A.M., Chadaeva I.V., Svichkarev A.V., Naumenko F.M., Kulakova E.V., Vityaev E.E., Chen M., Orlov Y.L. Computer analysis of colocalization of the TFs' binding sites in the genome according to the ChIP-seq data. *Russ J Genet Appl Res.* 2017;7(5):513-522. doi: 10.1134/S2079059717050057.
4. Kononov V.A., Orlova N.G., Orlov Y.L. Computer analysis of plant stress response: gene networks and transcriptional regulation. In: Proceedings of the International Conference "Scientific research of the SCO countries: synergy and integration". Part 3. Reports in English. March 31, 2022. Beijing: PRC. Scientific publishing house Infinity, 2022;143-152.
5. Orlov Y.L., Galieva A.G., Orlova N.G., Ivanova E.N., Mozyleva Y.A., Anashkina A.A. Reconstruction of gene network associated with Parkinson disease for gene targets search. *Biomeditsinskaya Khimiya.* 2021;67(3):222-230. doi: 10.18097/PBMC20216703222.

## Influence of tropomodulin 1 on calcium regulation of actin-myosin interaction in the ventricles

Kopylova G.<sup>1\*</sup>, Kochurova A.<sup>1,2</sup>, Beldiia E.<sup>1,2</sup>, Shchepkin D.<sup>1</sup>

<sup>1</sup> Institute of Immunology and Physiology, UB RAS, Yekaterinburg, Russia

<sup>2</sup> Ural Federal University, Yekaterinburg, Russia

\* g\_rodionova@mail.ru

**Key words:** myocardium, tropomodulin, actin-myosin interaction, calcium regulation

**Introduction and Aim:** For the contractile function of the myocardium, a strictly ordered and stable structure of the contractile apparatus of the cardiomyocyte is important determined by a number of proteins, which include tropomodulin (Tmod) [1]. Tmod binds to the thin filament and prevents it from disassembling and attaching new actin globules [1, 2]. Recently, it was found that in skeletal muscles, Tmod is involved in the activation of the thin filament, the regulation of actin-myosin interaction and affects the force developed by muscles [3]. The aim of our work was to study the molecular mechanism of the effect of the cardiac isoform of tropomodulin, Tmod1, on the calcium regulation of the actin-myosin interaction underlying myocardial contraction.

**Methods:** Myosin and native thin filaments (NTF) consisting of actin, troponin, and tropomyosin were isolated from the left ventricle of sheep. Human Tmod1 was expressed in *E. coli* [4]. To study the effect of Tmod1 on actin-myosin interaction we used an *in vitro* motility assay [4]. The calcium dependence of the sliding velocity of NTF over myosin in the *in vitro* motility assay was approximated by the Hill equation:  $V = V_{max} \times (1 + 10^{h(pCa - pCa_{50})})^{-1}$ , where  $V$  and  $V_{max}$  are the NTF velocity and the maximum velocity at a saturating calcium concentration;  $pCa_{50}$  – calcium sensitivity,  $pCa$  value at which half of  $V_{max}$  is reached;  $h$  is the Hill coefficient. To study the effect of Tmod1 on the calcium regulation of actin-myosin interaction, a concentration of 500 nM of Tmod1 was used in accordance with previous studies [5]. Data are presented as mean  $\pm$  standard deviation. Differences were assessed using the Mann–Whitney U-test,  $p < 0.05$ .

**Results:** It was found that Tmod1 affects the cooperative mechanisms of calcium regulation of actin-myosin interaction in the myocardium. Tmod1 significantly reduced the calcium sensitivity of sliding velocity of NTF from  $pCa_{50}$   $5.99 \pm 0.01$  to  $pCa_{50}$   $5.65 \pm 0.01$ . Tmod1 statistically insignificantly increased the Hill cooperativity coefficient from  $1.66 \pm 0.11$  to  $1.78 \pm 0.21$ . Tmod1 did not affect the NTF sliding velocity at saturating calcium concentration. The maximum sliding velocity of NTF without Tmod1 was  $2.3 \pm 0.1$   $\mu\text{m/s}$ , while with the addition of Tmod1 it was  $2.2 \pm 0.1$   $\mu\text{m/s}$ .

**Conclusion:** Thus, Tmod1 can act as a new regulator of actin-myosin interaction in the myocardium. The results obtained will contribute to understanding the molecular mechanisms of the influence of Tmod on the activation of a thin filament and actin-myosin interaction in the myocardium and the search for new therapeutic targets in heart pathologies.

**Acknowledgements:** The study is supported by Russian Science Foundation (22-24-00729).

## References

1. Gokhin D.S., Fowler V.M. Tropomodulin capping of actin filaments in striated muscle development and physiology. *J Biomed Biotechnol.* 2011;2011:103069.
2. Gokhin D.S. et al. Tropomodulin 1 directly controls thin filament length in both wild-type and tropomodulin 4-deficient skeletal muscle. *Development (Cambridge, England).* 2015;142(24):4351-4362.
3. Ochala J. et al. Pointed-end capping by tropomodulin modulates actomyosin crossbridge formation in skeletal muscle fibers. *FASEB J.* 2014;28:408-415.
4. Bershtsky S.Y. et al. Myopathic mutations in the  $\beta$ -chain of tropomyosin differently affect the structural and functional properties of  $\beta\beta$ - and  $\alpha\beta$ -dimers. *FASEB J.* 2019;33(2):1963-1971.
5. Colpan M. et al. The cardiomyopathy-associated K15N mutation in tropomyosin alters actin filament pointed end dynamics. *Arch Biochem Biophys.* 2017;630:18-26.

## First insight into regeneration potential of the ears of *Acomys cahirinus* in the single-cell resolution

Kozlova O.<sup>1\*</sup>, Bilyalov A.<sup>1,3</sup>, Nesmelov A.<sup>1</sup>, Gazizova G.<sup>1</sup>, Shagimardanova E.<sup>1</sup>, Gusev O.<sup>1,2</sup>

<sup>1</sup>Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia

<sup>2</sup>Graduate School of Medicine, Juntendo University, Tokyo, Japan

<sup>3</sup>Moscow Clinical Research Center Named After A.S. Loginov MHD, Moscow, Russia

\* [olga-sphinx@yandex.ru](mailto:olga-sphinx@yandex.ru)

**Key words:** regeneration, single-cell RNA sequencing

*Motivation and Aim:* It is well-known that spiny mouse, *Acomys*, shows exceptional regenerative ability after damaging its internal organs, full-thickness excision of dorsal skin and biopsy punch of ears. However, investigating such outstanding model of regeneration using modern molecular and bioinformatics methods is limited not only due to challenges associated with breeding and colony maintenance, but also due to inaccessibility of high quality *Acomys* genome information. This research presents the first attempt of looking into regeneration process of the *A. cahirinus* ear by RNA sequencing of individual cells.

*Methods and Algorithms:* We utilized 10X Genomics Single Cell 5' R2-only protocol for RNA-sequencing of the control (right after punch biopsy) and regenerating external ear cells (2 days of regeneration after operation). Due to inaccessibility of high quality genome assembly and annotation for *A. cahirinus*, we chose mouse genome for mapping with cellranger count, using some relaxed settings for STAR (that gave us approximately 97 % mapping rate and 71 % of reads uniquely mapped). Further processing was carried out with Seurat and Conos R packages.

*Results:* Approximately 4600 and 6800 cells (control and regenerating dataset respectively) were recovered during reads processing and quality control. Using certain gene markers for mouse cell types, we were able to identify all major types of external ear cells, including endothelial cells, cartilage, keratinocytes, smooth muscle cells, fat cells and dermis, glia, neurons and immune cells. Integration of two datasets showed us some differences in clusters structure, and the most remarkable notion was formation of specific cartilage subcluster in the regenerating dataset.

*Conclusion:* The first results obtained in the given research suggest that, even if the genome assembly for the non-model organism is currently unavailable, usage of closely-related genome may be a good alternative. This is especially applicable to single-cell studies, where good annotation is crucial for successive analysis – for example, for mitochondrial contamination deletion and finding cell-cycle dependent variation.

*Acknowledgements:* The study is supported by Ministry of Science and Higher Education of the Russian Federation grant 075-15-2021-1344.



## Chicken oocyte transcriptome: RNAseq of nuclear and cytoplasmic long and short RNAs

Krasikova A.<sup>1\*</sup>, Kulikova T.<sup>1</sup>, Maslova A.<sup>1</sup>, Bergardt V.<sup>1</sup>, Makarova N.<sup>2</sup>, Popov A.<sup>2</sup>, Schelkunov M.<sup>2,3</sup>, Fedotova A.<sup>2,4</sup>

<sup>1</sup> Saint-Petersburg State University, St. Petersburg, Russia

<sup>2</sup> Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>3</sup> A.A. Kharkevich Institute for Information Transmission Problems, Moscow, Russia

<sup>4</sup> Lomonosov Moscow State University, Moscow, Russia

\* [alla.krasikova@gmail.com](mailto:alla.krasikova@gmail.com)

**Key words:** chicken oocyte, hypertranscription, lampbrush chromosomes, transcriptome, RNAseq, RNA-FISH, nuclear RNA

*Motivation and Aim:* Hypertranscription during oogenesis is responsible for transformation of the meiotic chromosomes into the lampbrush form. Ongoing transcription on the lateral loops of avian lampbrush chromosomes was demonstrated by intense and rapid BrUTP incorporation. However, spectrum of sequences transcribed on lampbrush chromosomes of birds remained unknown. Here, we systematically characterized chicken oocyte transcriptome at the lampbrush chromosome stage of oogenesis.

*Methods:* We performed RNA sequencing of RNA isolated from the cytoplasm and nuclei manually dissected from chicken oocytes at the lampbrush chromosome stage. We examined nuclear and cytoplasmic mRNAs, as well as long and short non-coding RNAs. To verify RNAseq data, we used RNA-FISH on lampbrush chromosome preparations with BAC-clones overlapping with genes.

*Results:* Similar to *X. tropicalis* oocyte cytoplasm (Garder et al., 2012), in chicken ooplasm we detected mature mRNAs and long non-coding RNAs consisting of spliced exons. Visual examination of expression profiles for hundreds of expressed genes revealed that nuclear RNA contains both full-length gene transcripts and intronic sequences. By sequencing small RNA libraries we characterized nuclear and cytoplasmic small house-keeping non-coding RNAs and short regulatory RNAs including more than 150 miRNAs. Since we detected full-length nuclear transcripts for many genes, with reads covering the entire transcription unit, we aimed to detect nascent transcripts *in situ*. RNA-FISH with BAC-clones that overlap with genes whose transcripts appear in the nuclear and cytoplasmic RNAseq data on isolated chicken lampbrush chromosomes revealed nascent transcripts along lateral loops. RNase A pretreatment eliminated all transcripts from the RNP matrix of these loops. In total, by FISH-mapping we verified transcription of 42 protein-coding genes and 8 long non-coding RNA genes on the lateral loops of chicken lampbrush chromosomes. Among them are genes essential for oocyte maturation, ovarian development and early stages of embryogenesis. Transcripts of these genes could be used for oocyte maturation and/or early embryo development.

*Acknowledgements:* RNAseq was supported by Russian Science Foundation grant No. 19-74-20075. The work was performed using the equipment of Resource Centers ‘Molecular and Cell Technologies’ (Saint-Petersburg State University) and ‘Genomics Core Facility’ (Skolkovo Institute of Science and Technology).

# Computer annotation of plant protein sequences based on sequence similarity and orthology

Malyugin E.<sup>1,2\*</sup>, Mustafin Z.<sup>2,3</sup>, Pronozin A.<sup>2,3</sup>, Genaev M.<sup>2,3</sup>, Afonnikov D.<sup>2,3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* [evgeny.malyugin98@gmail.com](mailto:evgeny.malyugin98@gmail.com)

**Key words:** annotation, orthologs, gene age

**Motivation and Aim:** With the advent of new generation sequencing methods, the number of sequenced genome sequences is increasing tremendously. All this makes it possible to recognize the structure of the genome and predict the localization of protein-coding genes with high accuracy. However, for recognized genes, the function is usually unknown. Methods existing for this problem mainly use the search for homologous sequences with known functions described in terms of Gene Ontology to predict the function of genes. In this case, an important factor in sequence annotation is the age of the gene [1, 2], which is associated with the presence of its orthologs in a large number of taxa. For genes that have occurred recently during evolution (young or taxon-specific genes), orthologs are absent or poorly represented in other organisms. For them, function prediction is difficult and its accuracy is lower than for ancient genes, for which homologs are more widely represented. Thus, the development of a method for annotation of gene functions that would work with high accuracy regardless of the age of the genes is relevant.

**Methods and Algorithms:** We proposed an algorithm for gene function annotation based on similar sequence search and orthologous group analysis in the OrthoDB v. 10.0 [3]. The search for homologous sequences was performed in the database using the Usearch v 11 algorithm [4]. The method was tested on the sequences of protein-coding genes of a number of plant genomes, including *Arabidopsis thaliana*, rice, moss, algae, etc. Orthoscape software was used to estimate gene ages [5].

**Results:** The accuracy of the proposed algorithm when analyzing the *A. thaliana* genome was comparable with the eggNOG-mapper method [6]. Thus, the proportion of correct predictions (TP) among all positive predictions for our method was 86.6 %, for eggNOG-mapper method 85 %. It was shown that with certain parameters of the algorithm, the accuracy of function prediction in genes of different ages (including young genes) is close. Thus, the proposed method can reduce the differences in the accuracy of function prediction for genes of different ages, for better performance.

**Acknowledgements:** Development of algorithms 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.

## References

1. Liebeskind B.J., McWhite C.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.
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;47(D1):D807-D811.
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(1):1-9.
6. Cantalapiedra C.P., Hernández-Plaza A., Letunic I., Bork P., Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825-5829.

## Insights into *Opisthorchis felineus* genome organization

Maslov D.E.<sup>1,2\*</sup>, Ershov N.I.<sup>1</sup>, Pakharukova M.Y.<sup>1,2</sup>, Mordvinov V.A.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *bochlit2@gmail.com*

**Key words:** genome assembly, Hi-C, Opisthorchiidae, Rabl configuration

**Motivation and Aim:** *Opisthorchis felineus* (cat liver fluke) is a member of epidemiologically important liver trematodes, is one of the causative agents of human opisthorchiasis. An estimated 1.2 million people are infected with *O. felineus* and 12.5 million are at risk of infection [1, 2]. Liver fluke infection is known to adversely affect bile ducts and gall bladder and is recognized as one of the major factors of risk of cholangiocarcinoma [3, 4]. The vast majority of genome assemblies of flatworms, as well as non-model organisms in general, are built based on short-read sequencing and often fragmented. The only available draft assembly of *O. felineus* is no exception. Genome assembly fragmentation hinders the study of the genome functional organization and excludes any possibilities of analyzing spatial organization. The latter, although being thoroughly studied in vertebrate species for a decade now, is still poorly investigated in flatworms.

**Results:** We obtained a highly-continuous genome assembly of *O. felineus* utilizing an approach integrating long-read sequencing data (PacBio Sequel) and information on the physical proximity of genomic loci generated by Hi-C method. The assembly consists of 7 separate scaffolds reproducing haploid karyotype of the parasite represented by two larger and 5 lesser submetacentric chromosomes. The assembly completeness is further supported by a comprehensive gene annotation. We were able to identify Rabl-like chromosome configuration which contrasts *O. felineus* to vertebrates, which are predominantly characterized by more prominent chromosome territories, at the same time this feature is not unique to *O. felineus* and similar pattern had been previously described for another Opisthorchiidae species *Clonorchis sinensis*. Additionally, low instructiveness of eigenvector-based compartmentalization for chromatin transcriptional activity and inconsistency with observable spatial features indicates a need for more informative compartmentalization metric for species exhibiting similar patterns of chromatin organization.

**Conclusion:** A highly continuous and complete *O. felineus* genome provides a basis for delineating general peculiarities of its spatial-functional organization using Hi-C. Rabl-like chromosome configuration is a predominant characteristic of *O. felineus* chromatin organization, similar to several other non-vertebrate species.

**Acknowledgements:** The study was supported by the State Budget Project FWNR-2022-0016.

### References

1. Keiser J., Utzinger J. Emerging foodborne trematodiasis. *Emerg Infect Dis.* 2005;11(10):1507-1514.
2. Fedorova O.S., Fedotova M.M., Zvonareva O.I., Mazeina S.V., Kovshirina Y.V., Sokolova T.S. et al. *Opisthorchis felineus* infection, risks, and morbidity in rural Western Siberia, Russian Federation. *PLoS Negl Trop Dis.* 2020;14:e0008421.
3. Pakharukova M.Y., Mordvinov V.A. The liver fluke *Opisthorchis felineus*: Biology, epidemiology and carcinogenic potential. *Trans R Soc Trop Med Hyg.* 2015;110:28-36.
4. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. IARC, 1994;61:1-241.

## Peroxidase gene family in *Prunus persica* (L.)

Meger Y.<sup>1,2\*</sup>, Vodiasova E.<sup>1</sup>

<sup>1</sup> Kurchatov Genomic Centre of the Nikita Botanical Garden – National Scientific Center of RAS, Yalta, Russia

<sup>2</sup> Sevastopol State University, Sevastopol, Russia

\* meger\_yakov@mail.ru

**Key words:** peroxidase, transcriptomics, *Prunus persica*

**Motivation and Aim:** Peroxidase (POD) is the key antioxidant enzyme, which involved in cell response to different negative factors (cold, drought, infection, heavy metal, atmospheric and soil pollutants, etc.). This gene family includes a lot of genes. At the same time, in studies that aim to assess the impact of various negative factors on peaches usually one or two genes are analyzed only [1]. It should be noted that another key enzyme, catalase, is encoded by 2 genes in peach with low nucleotide differences, but they have tissue-specific expression [2]. The aim of this study, therefore, was to evaluate the expression of all genes encoding peroxidase in peach using transcriptomics techniques and determine the possible correlation between the individual gene expression and specific organ.

**Methods and Algorithms:** The peach genome assembly (accession number GCA\_000346465.2) contains 60 annotated genes encoding POD. The nucleotide diversity for all genes were analyzed. Phylogenetic tree analysis of all PODs was constructed with Maximum Parsimony method. The expression of genes encoding peroxidase in different tissues was analyzed using transcriptomic data available from NCBI. All genes expression was calculated using tpm method (transcripts per million) by Kallisto. The transcripts with tpm < 1 were excluded from analyses. Normalized counts by the relative log expression method implemented in DESeq2.

**Results:** Expression profiles of peroxidase encoding genes in different tissues were obtained. Phylogenetic analysis performed did not reveal a clear clustering of peroxidases. The exon-intron structure also varies greatly: some genes are intronless, while others have several long introns. The high nucleotide variability detected made it impossible to develop a single primer to estimate the total expression of the genes encoding a given protein.

**Conclusion:** Incorrect primer design and evaluation of the expression of not all genes encoding the target protein can lead to incorrect results. This needs to be taken into account during primer design as gene selection can play a significant role in studying the molecular mechanisms of adaptation of peach varieties to stress conditions.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Centre of the Nikita Botanical Garden – National Scientific Center of RAS (075-15-2019-1670).

### References

1. Haider M.S. et al. Drought stress revealed physiological, biochemical and gene-expressional variations in ‘Yoshihime’ peach (*Prunus Persica* L) cultivar. *J Plant Interact.* 2018;13(1):83-90.
2. Bagnoli F. et al. Molecular cloning, characterisation and expression of two catalase genes from peach. *Funct Plant Biol.* 2004;31(4):349-357.

## Efficient detection of structural variations using capture-Hi-C data

Mozheiko E.A., Babenko V.N.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [mozheiko@bionet.nsc.ru](mailto:mozheiko@bionet.nsc.ru)

The diagnosis and treatment of patients with hereditary diseases require the creation of efficient methods for the study of individual genomes. The existing approaches either are aimed at searching for a narrow set of genomic variants or are too expensive to use in routine practice. In this way, it is an actual issue to investigate efficient methods, which could couple detection point mutation in exons and searching for novel translocations and copy number alterations (CNA). We proposed a novel method based on sequencing of enriched Hi-C libraries to detect such variants. This method is based on the assumption that Hi-C technology is more informative for the search for chromosomal rearrangements than whole genome data with the same sequencing depth. This is achieved due to the fact that translocations and CNA significantly change the profile of the chromatin spatial contacts.

In this study, we prepare 10 enriched Hi-C-libraries from 8 patients' peripheral blood samples, and 2 cancer cell lines: k562 and A549 [1]. We used these cell lines as a reference of highly rearranged genomes to test our bioinformatics algorithm for detecting chromosomal rearrangements [2].

We have developed a bioinformatics algorithm for detecting interchromosomal translocations, which takes into account key features of altered profile of the chromatin spatial contacts. In this algorithm, we also took into account that chromosomal rearrangement can influence each other.

For CNA detection, we develop a coverage normalization algorithm, based on intrinsic properties of the chromatin. This type of normalization has never been used in other studies, and highly improves accuracy of CNA detection.

*Acknowledgements:* DNase I Hi-C library preparation and sequencing were supported by RFBR Grant 18-29-13021. Data analysis performed by E.M. was supported by RFBR Grant 20-34-90110. High-throughput computations were performed on the nodes of the computational cluster of the Institute of Cytology and Genetics, SB RAS (budget project No. 0259-2021-0016).

### References

1. Gridina M., Mozheiko E., Valeev E., Nazarenko L.P., Lopatkina M.E., Markova Z.G., Yablonskaya M.I., Voinova V.Yu., Shilova N.V., Lebedev I.N., Fishman V. A cookbook of DNase Hi-C. *Epigenetics Chromatin*. 2021;14(1):15. doi: 10.1186/s13072-021-00389-5.
2. Mozheiko E.A., Fishman V.S. Detection of point mutations and chromosomal translocations based on massive parallel sequencing of enriched 3C libraries. *Russ J Genet*. 2019;55:1273-1281. doi: 10.1134/S1022795419100089.

## Less is more: filtering and trimming ONT sequencing data

Nenasheva N.V.<sup>1,2,3\*</sup>, Ilyinsky V.V.<sup>3</sup>, Makeev V.J.<sup>1,2</sup>

<sup>1</sup> *Moscow Institute of Physics and Technology, Moscow, Russia*

<sup>2</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

<sup>3</sup> *Medical Genetic Center Genotek, Moscow, Russia*

\* *nenasheva.nv@phystech.edu*

**Key words:** Oxford Nanopore Technology (ONT), quality of ONT data, genome reference

Sequencing with Oxford Nanopore technology is rapidly gaining in its popularity as it has a number of advantages over sequencing by synthesis or other 1st and 2nd generation methods. Yet, this technology has its own limitations, particularly the unstable read quality. Since many tasks in bioinformatics, such as identification of genomic variants, structural changes, or epigenetic modifications, require high quality reads with unique genome mapping, we tested if the overall ONT data quality can be improved by filtering out low-quality reads or read fragments. To this end, we developed a set of criteria applied to the raw (fastq) base calling data, including the phred score, the level of read fragmentation and the sequence alignment coefficient normalized to the fragment length, which allowed us to remove poor quality fragments of long reads or even entire reads. To further improve the potential of ONT data we used the combined Illumina – ONT sequencing and mapped ONT reads to the “personal reference sequence” with corrected individual single nucleotide variants rather than to the standard database genome reference sequence. We assembled this reference sequence using PICARD (AddOrReplaceReadGroups, NormalizeFasta, CreateSequenceDictionary) and GATK (MarkDuplicates, BaseRecalibrator, ApplyBQSR, HaplotypeCaller, FastaAlternateReferenceMarker) tools. By so doing we visibly improved the quality of read mapping and alignment. We performed a series of comparative tests to assess contributions of the read errors and genome SNVs to mapping/alignment errors. Testing was performed on the data from six genomes belonging to the members of the same family obtained by Oxford Nanopore Technology sequencing, with the genome of one family member sequenced with Illumina. Both of these areas, improving the quality of sequencing data and improving the quality of mapping, can be applied together and used for analysis of individual DNA methylation with the help of Oxford Nanopore Technology data.

## Diversity of CRISPR arrays identified into plant mitochondrial genomes

Petrushin I.S.<sup>1,2\*</sup>, Gorbenko I.V.<sup>1</sup>, Konstantinov Yu.M.<sup>1,2</sup>

<sup>1</sup> Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia

<sup>2</sup> Irkutsk State University, Irkutsk, Russia

\* [ivan.kiel@gmail.com](mailto:ivan.kiel@gmail.com)

**Key words:** mitochondria, CRISPR, horizontal gene transfer

Mitochondrial genome of plants is much bigger than that of animals or fungi, and stores many noncoding sequences of non-mitochondrial origin. Moreover, mitochondrial genome of higher plants is actively involved in horizontal gene transfer processes where it can act both as a donor and a gene acceptor. In this work we identified many CRISPR-like arrays in number of mitochondrial genomes of higher plants. These CRISPR-like arrays are widespread among species and include agricultural (*Vitis vinifera*, *Beta vulgaris*, *Daucus carota*, *Vicia faba*), wild types (*Zea luxurians*) model plants (*Arabidopsis thaliana*), and even liverworts (*Treubia lacunose*, *Marchantia polymorpha*). Existence of such elements may be associated with mobile genetic elements (MGEs) presence or defense against DNA of foreign origin (including MGEs). Some of mitochondrial genomes contain dozens of CRISPR-like arrays: *Cucumis sativus* – 57, *Ophioglossum californicum* – 42, *Cucurbita pepo* – 36, *Cycas taitungensis* – 20. We suggest that CRISPR arrays were transferred to mitochondrial genomes during evolution probably from ancestor of *Alphaproteobacteria* host, which is probably a sister group to *Asgardarchaeota* (which possess CRISPR-Cas system). Due to the existence of mitoviruses, that infect mitochondria directly, evolution of plants led to the need for the CRISPR-Cas system to maintain the immunity of mitochondria. The mechanism of discovered CRISPR-like system might be different of those in Bacteria and Archaea, so our next step is to identify CRISPR associated proteins in mitochondrial and nuclear genomes of higher plants.



## **ICAnnoLncRNA: automatic pipeline of identification, classification and annotation of long noncoding RNAs**

Pronozin A.\*, Afonnikov D.

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

\* *pronozinartem95@gmail.com*

Long non-coding RNAs (lncRNAs) typically defined as transcripts of more than 200 nucleotides in length and without any protein coding potential. These RNAs are involved in important plant development processes such as phosphate homeostasis, flowering, photomorphogenesis and stress response. Information about lncRNA sequences and their expression typically obtained from transcriptome sequencing. To process these data relevant tools for automated lncRNA are required.

Here, we propose a ICAnnoLncRNA, pipeline for automatic identification, classification and annotation of plant lncRNAs. The pipeline includes: (1) identification of lncRNAs by LncFinder and CPC2; (2) transmembrane potential prediction by TMHMM; (3) Mapping lncRNA to reference genome by GMAP; lncRNAs classification by its localization in the genome using gffcompare; (5) analysis of the lncRNAs structural features; (6) search for conservative lncRNA with homology to sequences from other organisms. The pipeline implemented using the Snakemake [1] workflow management system language.

We analyzed 45 million *Z. mays* transcripts. Identify ~31 million lncRNAs by LncFinder and CPC. For ~10 million (26 %) lncRNAs the significant transmembrane potential was predicted. 21 million (47 %) lncRNAs were aligned to the reference genome. We identified antisense – 3.565.080, intron – 2.785.105, intergenic – 2.856.326 lncRNAs. The 3 million lncRNAs in maize have homologs and 297.299 of them have homologs among known lncRNAs from other plants. Tissue-specificity analysis showed that the vast majority of lncRNAs are expressed in ovary tissues.

The proposed automatic pipeline made it possible to identify 9.206.511 new lncRNAs in maize genome, classified into classes depending on their localization in the genome, annotate them and evaluate their structural features.

*Acknowledgements:* The work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

### *References*

1. Köster J., Rahmann S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;19:2520-2522.

## Post-stress changes in the gut microbiota in rat strains with different excitability of the nervous system

Shevchenko A.U.<sup>1\*</sup>, Matskova L.V.<sup>1</sup>, Shalaginova I.G.<sup>1</sup>, Dyuzhikova N.A.<sup>2</sup>

<sup>1</sup> Immanuel Kant Baltic Federal University, Kaliningrad, Russia

<sup>2</sup> Pavlov Institute of Physiology, RAS, St. Petersburg, Russia

\* shevch2009@yahoo.com

**Key words:** stress, nervous system excitability, post-stress disorders, gut microbiota, microbiota-brain axis

*Motivation and aim:* The great diversity of bacteria that inhabit the mammalian gut constitutes a complex and dynamic ecosystem [1] that plays an important role in the homeostasis of the host organism. Changes in the microbial community are the influence of various genetic, environmental, and dietary factors. The aim of our study was to show the effect of stress on the gut microbial community of two different strains of rats, with low and high nervous system excitability.

*Methods and Algorithms:* The experiment was performed using 24 male rats (5-month-old) from two strains which differed in their nervous system excitability; the high excitability HT strain (12 rats) and the low excitability LT strain (12 rats). Animals were divided randomly into two groups. The control and experimental groups consisted of 6 rats each. The same standard environmental conditions were used for all animals. The Hecht protocol for prolonged emotional and painful stress was used on experimental groups of rats [2] as follows: every day for 15 days, the animals were exposed to 6 unsupported (10 seconds each) and 6 current-reinforced (2.5 mA, 2 ms) light signals. Samples of feces were collected before the stress and on the 7th and 24th days after.

Total DNA from fecal samples was isolated using the QIAamp DNA Stool Mini kit (51504, QIAGEN). The V4 region of the bacterial 16S rRNA gene was amplified with dual-indexing bacterial 16S RNA primers F515 (5'-GTGBCAGCMGCCGCGGTAA-3') and R806 (5'-GGACTACHVGGGTWTCTAAT-3'). Amplification was performed with the HS SYBR qPCR mix reaction mixture (Eurogen, Moscow, Russia) on a CFX96 real-time PCR system (BioRad, Hercules, CA, USA).

The quality of the PCR products was checked by agarose gel electrophoresis. Purification of amplicons was performed with the QIAquick Gel Extraction Kit (250).

The purified libraries were pooled in an equimolar concentration and further paired-end sequencing was performed using an Illumina Miseq instrument (Illumina Inc., San Diego, CA, USA).

The raw reads were checked with FastQC, then filtered with fastp and demultiplexed with the deML package. Further analysis was performed using the Quantitative Insights Into Microbial Ecology 2 (Qiime2-2020.14) software package. The alpha diversity of the operational taxonomic units (OTU) was calculated using the q2-diversity Qiime2 plugin. The taxonomy of the reads was annotated to the Greengenes database using a naive Bayes classifier via the q2-feature-classifier plugin.

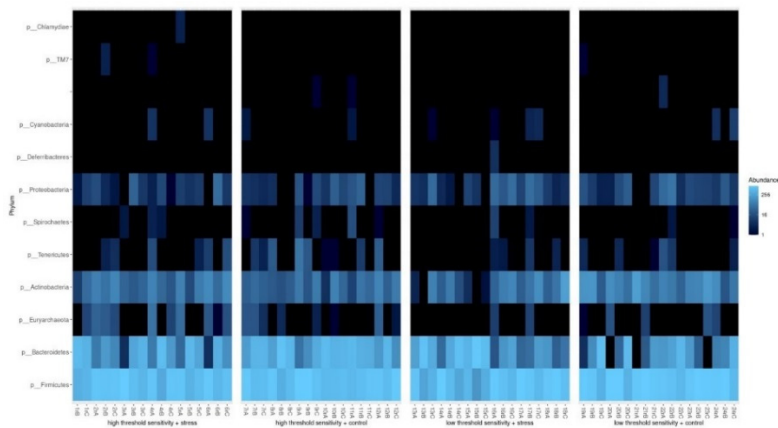
Distribution of data was evaluated by Shapiro-Wilk test using the GraphPad Prism software (v.6, GraphPad Software Inc., La Jolla, CA). Non-parametric tests (Mann-Whitney U tests, Fisher's exact test, Wilcoxon) were also performed with the GraphPad Prism software.

*Results:* In the intact animals of high excitable strain (LT), we detected a significant reduction of the Chao1 and Shannon indexes for the  $\alpha$ -diversity of the bacterial

composition compared with the low excitable strain (HT). Our analysis revealed 11 Phyla among all samples. The most diverse taxonomic levels were represented by *Firmicutes* and *Bacteroidetes*, which were found in every sample. Less represented Phyla were *Actinobacteria*, *Proteobacteria*, *Euryarchaeota*, *Tenericutes*, *Spirochaetes*, *Cyanobacteria*, *Deferribacteres*, *Chlamydiae*, *Saccharibacteria* (TM7). The most distributed Orders in our samples were *Lactobacillales*, *Bacteroidales*, *Clostridiales*. We identified 59 genera, which the most represented were *Lactobacillus*, *Prevotella*, *Allobaculum*, *Enterococcus*, *Blautia*, *Collinsella*, *Faecalibacterium*, *Dorea*, *Bifidobacterium*.

We have found the following post-stress changes in the composition of the microbial community. A progressive decrease of *Lactobacillus* Genus in both strains from the 7th until the 24th day after the stress. *Enterococcus* was not detected in the LT strain but appeared in the HT strain after 24 days in response to stress. Genus of *Blautia*, *Collinsella* and *Methanobrevibacter* were overrepresented in the LT strain but were reduced in the HT strain. In the HT group, the Genera *Faecalibacterium* and *Prevotella* had low relative frequencies before the stress stimulus but increased significantly on days 7 and 24 after stress.

**Conclusion:** Our present data on dynamic changes in microbial diversity after stress provide support for further studies and may be correlated to the effects of probiotic therapy. Also, knowledge of the dynamic changes in bacterial species under stressful conditions will add to our understanding of how interspecies relationships contribute to shifts in the gut microbial community.



**Fig. 1.** Heatmap of the Phylum level of the gut microbiota of rats with high and low excitability of the nervous system. 1rA, 1rB, 1rC are the same animals at different time points

## References

1. Rogers G.B., Kozłowska J., Keeble J., Metcalfe K., Fao M., Dowd S.E., Mason A.J., McGuckin M.A., Bruce K.D. Functional divergence in gastrointestinal microbiota in physically-separated genetically identical mice. *Sci Rep.* 2014;4:5437. doi: 10.1038/srep05437.
2. Vaido A., Shiryaeva N., Pavlova M., Levina A., Khlebaeva D., Lyubashina O., Dyuzhikova N. Selected rat strains Ht, Lt as a model for the study of dysadaptation states dependent on the level of excitability of the nervous system. *Laboratory Animals Sci.* 2018;3. doi: 10.29296/2618723X-2018-03-02. (in Russian)

## Genome assembly of *Colletotrichum lini* from long Nanopore reads

Sigova E.A.<sup>1,2\*</sup>, Dvorianinova E.M.<sup>1,2</sup>, Rozhmina T.A.<sup>1,3</sup>, Kudryavtseva L.P.<sup>3</sup>,  
Melnikova N.V.<sup>1</sup>, Dmitriev A.A.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup> Federal Research Center for Bast Fiber Crops, Torzhok, Russia

\* sigova.ea@phystech.edu

**Key words:** flax pathogens, *Colletotrichum lini*, genome assembly, Oxford Nanopore

**Motivation and Aim:** *Colletotrichum lini* is the malicious flax anthracnose causative agent. Studying the fungus at the genetic level is vital for the successful disease control. However, the lack of the *C. lini* whole genome sequence hinders extensive molecular research on the pathogen. Therefore, our aim was to obtain the first genome assembly of *C. lini* using the Oxford Nanopore Technologies (ONT) sequencing platform.

**Methods and Algorithms:** *C. lini* highly pathogenic strain #811 was provided by the Institute for Flax (Torzhok, Russia). Pure high-molecular DNA was obtained according to our previously developed protocol. The spectrophotometry (Nanodrop) and fluorometry (Qubit 4) methods were used to evaluate the quality and quantity of the extracted DNA. DNA libraries were prepared and sequenced on the ONT (MinION instrument, FLO-MIN-106 R9.4.1 flow-cell) platform according to the manufacturer's protocol. The obtained reads were basecalled using Guppy 6.0.1 with different quality filtration thresholds (min\_qscore, i.e. the minimum average read quality of a basecalled read, was taken in the range from 7 to 10). Porechop 0.2.4 was used for adapter removing. Draft assemblies for each minimum quality score value were performed using Canu 2.2, Flye 2.8.1, Raven 1.5.1, Shasta 0.8.0, Wtdbg-cns 1.1 (Wtdbg2 0.0), NextDenovo 2.5.0, Miniasm 0.3-r179, Ra 0.2.1, and SmartDenovo tools. BUSCO 5.3.2 and QUASt 5.0.2 were used to analyze the quality of the obtained assemblies.

**Results:** We obtained 1.7 Gbases (Gb) of raw ONT reads with an N50 of 15.7 kb. Genomes assembled from this data with different tools and quality filtration thresholds have the statistics shown in Table 1.

The assemblers gave better results at lower min\_qscore values (7–8), since at high min\_qscore values (9–10) the coverage was insufficient to obtain a quality assembly. The highest assembly completeness for each min\_qscore was achieved by Flye (up to 93.7 %), the lowest – by Miniasm (less than 30 %). The average length of the assemblies with BUSCO > 80 % was 52.2 Mb (48.1–53.5 Mb). For the basecalled data with a min\_qscore of 7, Flye also produced the most contiguous assembly: N50 of 4.4 Mb for a total length of 53.4 Mb, 42 contigs. Thus, the genome assembly of *C. lini* strain #811 obtained with Flye from the ONT data basecalled with min\_qscore = 7 can be considered the most complete and contiguous.

**Table 1.** Statistics on the obtained genome assemblies

Q	Assembler	Assembly length, Mb	BUSCO, %	Number of contigs	N50, Mb	Q	Assembler	Assembly length, Mb	BUSCO, %	Number of contigs	N50, Mb
10	<i>Canu</i>	48.1	80.9	443	0.15	8	<i>Canu</i>	52.9	88.8	174	0.53
	<i>Flye</i>	52.9	89.7	111	1.11		<i>Flye</i>	53.4	93.7	37	3.41
	<i>Raven</i>	32.2	50.8	351	0.11		<i>Raven</i>	51.4	85.5	160	0.49
	<i>Wtdbg2</i>	50.3	67.4	178	0.59		<i>Wtdbg2</i>	52.1	74.8	81	2.10
	<i>Shasta</i>	27.1	41.3	691	0.06		<i>Shasta</i>	45.0	59.3	709	0.10
	<i>NextDenovo</i>	–	–	–	–		<i>NextDenovo</i>	40.1	69.8	135	0.37
	<i>Ra</i>	26.8	45.4	248	0.12		<i>Ra</i>	50.5	82.2	177	0.43
	<i>SmartDenovo</i>	401.6	75.8	24544	0.02		<i>SmartDenovo</i>	622.4	76.9	38180	0.02
	<i>Miniasm</i>	27.9	20.2	290	0.11		<i>Miniasm</i>	48.5	27.0	191	0.39
9	<i>Canu</i>	51.7	86.6	260	0.33	7	<i>Canu</i>	53.5	88.8	134	0.81
	<i>Flye</i>	53.1	92.7	48	3.31		<b><i>Flye</i></b>	<b>53.4</b>	<b>93.5</b>	<b>42</b>	<b>4.44</b>
	<i>Raven</i>	45.5	71.6	285	0.20		<i>Raven</i>	52.8	89.9	90	0.95
	<i>Wtdbg2</i>	51.6	73.6	98	1.30		<i>Wtdbg2</i>	51.8	71.0	40	3.20
	<i>Shasta</i>	42.1	55.5	762	0.08		<i>Shasta</i>	46.1	61.6	663	0.11
	<i>NextDenovo</i>	0.08	0.3	1	0.08		<i>NextDenovo</i>	52.3	91.6	69	1.24
	<i>Ra</i>	43.4	70.7	274	0.19		<i>Ra</i>	52.5	85.7	98	0.87
	<i>SmartDenovo</i>	507.7	76.6	31063	0.02		<i>SmartDenovo</i>	753.5	77.1	46314	0.02
	<i>Miniasm</i>	41.9	26.3	297	0.18		<i>Miniasm</i>	50.5	23.6	96	0.84

**Conclusion:** We obtained the first *C. lini* genome assembly from long ONT reads. This knowledge is a starting point for further detailed research on *C. lini* and the flax-pathogen interaction.

**Acknowledgements:** This work was financially supported by the Russian Science Foundation, grant No. 22-16-00169.

# The metagenomic analysis of Colorado potato beetles public NGS data and the nanopore sequencing of its genetic material

Starchevskaya M.E.<sup>1,2\*</sup>, Kamanova E.P.<sup>1,2</sup>, Vyatkin Y.V.<sup>1,2,3</sup>, Tregubchak T.V.<sup>2</sup>, Bauer T.V.<sup>2</sup>, Rotskaya U.N.<sup>4</sup>, Kosman E.S.<sup>4</sup>, Antonets D.V.<sup>1,2</sup>

<sup>1</sup> State Research Center of Virology and Biotechnology "Vector", Novosibirsk, Russia

<sup>2</sup> Novel Software Systems LLC, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia

\* [starchevskayamaria@mail.ru](mailto:starchevskayamaria@mail.ru)

**Key words:** viral metagenomics, Colorado potato beetle, viruses of insects, nanopore sequencing

**Motivation and Aim:** The Colorado potato beetle (*Leptinotarsa decemlineata* from Chrysomelidae family) is one of the most serious insect pests. High ecological plasticity and ability to adapt to a variety of Solanaceae plants has led to its global expansion worldwide [1]. *L. decemlineata* is also known for its ability to rapidly develop resistance to insecticides [2]. Another promising approach to control insect pests is the use of biological insecticides based on viruses [3], but the information on viruses, which infect leaf feeding beetles, is scarce. Baculoviridae and Iridoviridae viruses were proposed to infect Chrysomelidae beetles [4]. Besides, leaf feeding beetles were shown to transmit different plant viruses [5, 6]. However, despite the undoubted importance of Colorado potato beetle virome for agriculture, it has not been studied yet and, as far as we know, our work is the first attempt to discover the viral material in DNA- and RNA-Seq data of *L. decemlineata*. We also hope that identification of new viruses will help to extend the arsenal of biopesticides.

**Methods and Algorithms:** 119 genomic and transcriptomic samples of *Leptinotarsa decemlineata* were selected from the NCBI SRA database. Genomic data obtained from muscle tissue (PRJNA508767, PRJNA369863) and whole insect tissue (PRJNA171749). The transcriptome datasets were obtained from antennas (PRJNA280017), heads (PRJNA400685), intestines (PRJNA400685, PRJNA336167), larvae (PRJNA275431) and whole insect (PRJNA438159, PRJNA384383, PRJNA297027, PRJNAJ427580, PRJNA297027, PRJNAJ42764380, PRJNA297027, PRJNAJ42764380). The quality control of reads was carried out and low-quality sequences were filtered, and possible contamination was eliminated. The reads that were not aligned to *L. decemlineata* reference genome (assembly GCA\_000500325.2 Ldec\_2.0) were assembled using metaSPAdes [7]. More than 3.5 million contigs were obtained, 7726 contigs with a coating of more than 20 and a length of more than 500 bp were selected. The selected contigs were analyzed with BLAST against NCBI nt database. The sequences of unidentified contigs were translated into 6 reading frames and analyzed with BLAST using NCBI nr and NCBI CDD databases. Possible gene sequences were predicted using MetaGeneMark [8] and the obtained gene sequences were also analyzed with BLAST using NCBI nr/nt and CDD databases. Contigs and non-aligned reads were also analyzed with Kraken2 software [9].

**Results:** 4334 contigs were attributed to different taxonomic groups. Among the identified viral families, the most reads were attributed to Baculoviridae, Iridoviridae,

Polydnaviridae, Iflaviridae, Mimiviridae, Pandoraviridae, Phycodnaviridae, and Poxviridae families. Also, full-length genome sequences of several plant viruses were assembled: Potato virus S, and Potato virus Y, Grapevine virus, and Ligustrum virus. The largest number of viral fragments was found in intestinal tissues. We have also found that some genes, predicted with MetaGeneMark, probably belong to bracoviruses. Bracoviruses belong to Polydnaviridae virus family and are associated with parasitoid wasps. Further examination of Colorado potato beetle reference genome revealed numerous long fragments of *Bracovirus* genomic sequences (over 25,000 bp in total) attributed to *Cotesia sesamiae Mombasa bracovirus* and *Cotesia vestalis bracovirus*. We analyzed 18 more *Coleoptera* genomes and found that Colorado potato beetle genome was the fourth most enriched with bracoviral sequences. As far as we know, this is the first report of a polydnavirus found in Chrysomelidae species. Bracoviruses are known to integrate into genomes of their hosts that is why we decided to examine if these sequences could be found in the genetic material extracted from *L. decemlineata* eggs collected by our colleagues from ISEA SB RAS. 2 contigs 1.5–2 kbp long, attributed to *Cotesia sesamiae Mombasa bracovirus*. A were selected and a number of oligonucleotide flanking the contig sequences were designed, and a set of primers for Sanger sequencing. The presence of bracoviral fragments in *L. decemlineata* genetic material was confirmed with PCR and with Sanger sequencing. The genetic material extracted from Colorado potato beetle eggs was sequenced with MinION sequencer (Oxford Nanopore Technologies). Over 1 million reads with a total length of 4,74 Gb were obtained. The average read length was above 4.6 kb and the maximal length was over 1 Mb. The bracoviral genetic sequences were found to be integrated into the genomic DNA of Colorado potato beetle.

**Conclusion:** Metagenomic studies of insects are extremely important. The identified viruses allow us to expand our understanding of the ecology and evolution of both the viruses themselves and their hosts. Discovery of new viruses will potentially expand the arsenal of biocontrol agents.

**Acknowledgements:** This work was supported by the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-15-2019-1665).

## References

1. Cingel A. et al. Extraordinary adaptive plasticity of Colorado potato beetle: “Ten-striped Spearman” in the Era of Biotechnological Warfare. *Int J Mol Sci.* 2016;17(9):1538.
2. Alyokhin A. et al. Colorado potato beetle resistance to insecticides. *Am. J. Potato Res.* 2008;85:395-413.
3. Cuartas P.E. et al. Novel biopesticide based on *Erinnyis ello* betabaculovirus: characterization and preliminary field evaluation to control *Erinnyis ello* in rubber plantations. *Pest Manag Sci.* 2019;75:1391-1399.
4. Lashomb J. et al. *Edovum putterli* (Hymenoptera: Eulophidae), an egg parasitoid of Colorado potato beetle (Coleoptera: Chrysomelidae): development and parasitism on eggplant. *J. Econ. Entomol.* 1987;80(1):65-68.
5. Wickizer S.L., Gergerich R.C. First report of Japanese Beetle (*Popillia japonica*) as a vector of Southern bean mosaic virus and Bean pod mottle virus. *Plant Dis.* 2007;91(5):637-637.
6. Kersch-Becker M.F., Thaler J.S. Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores. *Oecologia.* 2014;174:883-892.
7. Bankevich A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477.
8. Zhu W. et al. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 2010;38:e132.
9. Wood D.E., Salzberg S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46.

## Characterization of the glutathione S-transferase gene family in Mollusca

Vodiasova E.<sup>1\*</sup>, Uppe V.<sup>1</sup>, Chelebieva E.<sup>1</sup>, Rasskazov D.<sup>2</sup>

<sup>1</sup> A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Sevastopol, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* eavodiasova@gmail.com

**Key words:** molecular evolution, orthologous genes, gene expression, GST

**Motivation and Aim:** Glutathione S-transferase (GST) is one of the critical proteins in xenobiotic metabolism and a cellular protector under various stress factors as intoxication, hypoxic condition, infection, etc. It is known cytosolic, mitochondrial and microsomal subfamilies of GST [1]. The cytosolic subfamily has a lot of gene classes, representing which differs in various organisms. Each of GST classes may lead to a different expression profile depending on the negative factor or the type of tissue [2]. This should be considered when analyzing gene expression during toxicological and ecological studies. However, there are very few studies on GST encoding genes in marine invertebrates. This study aimed to determine all genes encoding glutathione S-transferases in Mollusca and classify them based on genome data.

**Methods and Algorithms:** To date the genome assemblies are available for 81 Mollusca species in NCBI. All GST genes for each species were identify through mapping Mollusca GSTs from non-redundant protein sequence and SwissProt databases on genomes. The total RNA was extracted from different tissues of *Mytilus galloprovincialis* for investigation of tissue-specific GSTs expression profile. The beta-actin as a housekeeping gene was used as a reference. All reactions were performed with three replicates. The relative expression levels of each targeted gene were counted by using the  $\Delta\Delta C_t$  comparative method.

**Results:** All classes of GST for each mollusk were characterized and their exon-intron structure analyzed based on genome data. For many species, these classes have been identified for the first time. It is shown that some classes have duplications in the genome (alpha, sigma, pi, omega, and mu). In addition, the expression of each gene under hypoxic conditions was investigated for *M. galloprovincialis*. It is worth noting that increased expression was found only for omega and sigma GST classes, confirming the assumption that the expression profile depends on the stress factor. The possible tissue specificity of the GST gene family in *M. galloprovincialis* was also analyzed.

**Conclusion:** The obtained results suggest a greater representation of different GST classes in Mollusca than previously thought. Probably such diversity of GST in the genome is some mechanisms of adaptation evolution resulting in the occurrence of an effective system of cellular detoxification. This question requires further investigation.

**Acknowledgements:** This work is funded by State Assignment No. 121030100028-0.

### References

1. Vaish S. et al. Glutathione S-transferase: a versatile protein family. *3 Biotech.* 2020;10(7):321.
2. Boutet I., Tanguy A., Moraga D. Characterisation and expression of four mRNA sequences encoding glutathione S-transferases pi, mu, omega and sigma classes in the Pacific oyster *Crassostrea gigas* exposed to hydrocarbons and pesticides. *Marine Biol.* 2004;146:53-64.



# Fold-change Specific Enrichment Analysis (FSEA): an approach for functional enrichment analysis of transcriptomic data

Wiebe D.<sup>1\*</sup>, Omelyanchuk N.<sup>1</sup>, Zemlyanskaya E.<sup>1</sup>, Mironova V.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute for Water and Wetland Research, Radboud University, Nijmegen, Netherlands*

\* *wiebe@bionet.nsc.ru*

Currently, transcriptomic experiments are the most common approach to studying gene expression dynamics in response to experimental conditions. Previously, we found that gene expression profiling data contains more information than is routinely extracted with standard approaches [1]. In order to fill this gap, we developed a new method for functional annotation of differentially expressed genes from transcriptome data with respect to their fold changes – Fold-change Specific Enrichment Analysis (FSEA). FSEA identifies Gene Ontology (GO) terms, which are annotated to genes with similar changes in the pattern of expression. We named GO terms found by FSEA as fold-change-specific in response to stimulus under investigation. Approbation identified many responses to abiotic factors, treatments, and diseases as regulated in a fold-change-specific manner. One of the most interesting findings offered by FSEA showed that there are two prevailing responses of functionally-related gene groups, either weak or strong. It is noteworthy that some fold-change-specific GO terms are invisible by classical algorithms for functional gene enrichment. Such GO terms were not enriched compared to the genome background, but were regulated by a factor within specific fold-change intervals. Application of FSEA to cancer-related transcriptomes revealed gene groups with a tightly coordinated response that could be potential regulators, markers, and therapeutic targets in oncogenic processes. To make FSEA available to researchers, we implemented it as the FoldGO Bioconductor R package (<https://www.bioconductor.org/packages/release/bioc/html/FoldGO.html>) along with a web-server version (<https://webfsgor.sysbio.cytogen.ru/>).

*Acknowledgements:* The work was financially supported by the Russian Science Foundation, project No. 21-14-00240.

## References

1. Omelyanchuk N.A., Wiebe D.S., Novikova D.D., Levitsky V.G., Klimova N., Gorelova V., Weinholdt C., Vasiliev G.V., Zemlyanskaya E.V., Kolchanov N.A. et al. Auxin regulates functional gene groups in a fold-change-specific manner in *Arabidopsis thaliana* roots. *Sci Rep.* 2017;7:2489.

## **RNA-seq analysis of the Asiatic *Vigna* to examine differential gene expression during germination and early plant development**

Ziobron A., Samsonova M., Wettberg E.

*University of Vermont, Burlington, USA*

Seed germination strategies varies between plant that place cotyledons above ground in epigeal germination or reserve them belowground in hypogeal germination. The type of cotyledon emergence is usually fixed for many plant families; however, for Fabaceae species (also known as the legumes), such as those in the genus *Vigna*, both epigeal and hypogeal germination are observed. In this study, we examined phenotypic variation of these germination traits, as well as differentially expressed genes (DEGs) during early seedling development among *Vigna* species in the subgenus *Ceratotropis* (“Asiatic *Vigna*”) using RNA sequencing. Thousands of DEGs were detected among multi-species comparisons, reflecting numerous genetic factors influencing timing of gene expression during early development of *Vigna* species. Gene ontology enrichment analyses further revealed functional variation associated with the germination strategies. Possible key functions and select genes further differentiating the Asiatic *Vigna* species, were also identified. These functions included, but were not limited to, stress and defense responses, signal transduction, and photosynthesis, reflecting trade-offs associated with differing germination types. Such trade-offs during seedling development have important agroecological implications in terms of crop productivity, crop diversity, and a plant’s response to environmental conditions. While there remains a need for a more in depth genetic analysis of seedling morphological traits among the Asiatic *Vigna*, results of this study do provide preliminary research into functions influencing gene expression during early plant development in this genus. The results further highlight the need for increased plant genomic resources, particularly a genus-wide pan-genome.

## 1.2 Section “Regulatory genomics”



# ITAS: integrated transcript annotation for small RNA

Bezuglov V.<sup>1,2\*</sup>, Stupnikov A.<sup>3,4</sup>, Skakov I.<sup>2</sup>, Shtratnikova V.<sup>1</sup>, Pilsner J.R.<sup>5</sup>,  
Suvorov A.<sup>1,6</sup>, Sergeev O.<sup>1\*</sup>

<sup>1</sup> *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *Moscow Institute of Physics and Technology, Moscow, Russia*

<sup>4</sup> *National Medical Research Center for Endocrinology, Moscow, Russia*

<sup>5</sup> *Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, USA*

<sup>6</sup> *Department of Environmental Health Sciences, University of Massachusetts, Amherst, MA, USA*

\* *vitaly1530@gmail.com, olegsergeev1@yandex.ru*

**Key words:** Transcript annotation, Differential Gene Expression, small RNA, microRNA, piRNA, tRNA-derived small RNA, RNA-seq, small RNA fragments, transcriptomics

*Motivation and Aim:* Transcriptomics analysis of various small RNA (sRNA) biotypes is a new and rapidly developing field. Annotations for microRNAs, tRNAs, tRNA-derived (tsRNA), piRNAs and rRNAs, contain information on transcripts sequences and loci that is vital for downstream analyses. Several databases have been established to provide this type of data for specific RNA biotypes. However, these sources often contain data in different formats, which makes bulk analysis of several sRNA biotypes in a single pipeline challenging. Information on some transcripts may be incomplete or conflicting with other entries. To overcome these challenges, we aimed to introduce ITAS, or Integrated Transcript Annotation for Small RNA, a filtered, corrected and integrated transcript annotation containing information on several types of small RNAs including tsRNAs for several species (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*), which was tested in several case studies for human-derived data against existing alternative databases.

*Methods and Algorithms:* Files with transcripts sequences in fasta format (referred to as fasta sequences) and annotation of transcripts in bed-format (loci) were obtained from corresponding databases: GtRNAdb Data Release 19 (June 2021) for mature tRNA sequences, piRNAdb v1.7.6 for piRNAs, miRBase Release 22.1 for microRNAs, UCSC for rRNA, MINTbase v2.0 (Human) and tRFdb for tsRNA. For some loci additional UCSC liftOver procedure followed to transform loci to the most recent genome version. To overcome the incompleteness of databases or conflicting data for some transcripts sequences in fasta-format from the reference genome (getfasta sequences) which correspond to the annotation bed-file were obtained by bedtools getfasta version 2.27.1, and fasta sequences were mapped on the reference genome of current version by hisat2. Then consensus database for each sRNA type was built, and transcripts correction was proceeded with the help of R instruments. Transcripts with conflicting data due to inbase and interbase (mature microRNA, piRNA, rRNA, tRNA) loci conflicts were detected with the help of bedtools intersect. Human sperm RNA-seq data from three publicly available datasets [1–3] were used for case studies. The differential expression analysis was done by DESeq2 package in relation to factors analyzed in original studies.

*Results:* The most popular databases of sRNA were analyzed. The following steps were invoked to approach outlined problems while generating the ITAS: (1) missing data for the incomplete entries were retrieved and filled in; (2) the problem of multiple loci per transcript was addressed; (3) transcripts with conflicting fasta-delivered and locus-delivered data were

identified, corrected if possible, and filtered out otherwise; (4) in-database and inter-databases loci-wise intersecting entries were identified and filtered out; (5) inter-databases loci-wise intersecting entries were identified and filtered out.

The identified problems were corrected where possible and entries with severe conflicts were filtered out. The statistics for Human annotations correction for different biotypes of small RNA is provided in Table 1. The conducted case studies using human sperm RNA-seq data [1–3] demonstrated the advantages of ITAS. Mapping of reads to ITAS which unifies in a single gtf format database transcripts from all source databases allowed for the identification of more significant transcripts as compared with the “map and remove” approach. In particular for sRNAs expression analysis, all cases revealed that ITAS based genome alignment approach identified more significant transcripts than the “map and remove” pipeline with default databases (11 vs 66 for [1], 43 vs 212 for [2], 26 vs 242 for [3]).

**Table 1.** Statistics on correction events in Human RNA transcript entries: cases when both locus and sequence were present (Loci & Seq, no correction), only locus or only sequence (Loci only, Seq only, sequence or locus retrieved from genome); cases that required extending entry’s locus, sequence or both (Ext Loci & Seq, Loci & Ext Seq, Ext Loci & Ext Seq); cases with transcripts loci intersections within same database (Inbase conflicts) and intersections between different databases (Interbases conflicts)

RNA type	Database	Loci & Seq	Loci only	Seq only	Ext Loci & Seq	Loci & Ext Seq	Ext Loci& Ext Seq	Inbase conflict	Interbases
mature microRNA	miRBase	1291	1227	1	0	0	0	140	21
piRNA	piRNadb	42 2017	0	1	0	0	0	388 826	1268
tRNA	GtRNadb	225	176	2	0	0	0	0	214
rRNA	UCSC	1591	0	186	0	0	0	37	126
tRNA-derived	MINTbase	0	0	8120	115 040	33	8	*	*

\* No intersection events were considered for tRNAs, as fragments of the same tRNA naturally have intersecting loci.

**Conclusion:** We have identified several issues during the inference of existing databases containing information on sRNA transcripts sequences and annotation for five species. Some transcripts had missing information on their sequences or loci; for others, their genome locus-retrieved sequence and database provided-sequence were not matching. Transcripts had both in-database and inter-databases intersecting loci with other entries. To address these drawbacks we established ITAS, a filtered, corrected, and integrated database for 5 species in a widely used gtf format, which was tested in several case studies for human-derived data against existing alternative databases.

**Acknowledgements:** The study was supported by the Russian Science Foundation grant No. 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 Metabolism*. 2016;23(2):369-378.
2. Ingerslev L.R., Donkin I., Fabre O., Versteyhe S., Mechta M., Pattamaprapanont P., Mortensen B., Krarup N.T., Barrès R. Endurance training remodels sperm-borne small RNA expression and methylation at neurological gene hotspots. *Clin Epigenetics*. 2018;10:12.
3. Hua M., Liu W., Chen Y., Zhang F., Xu B., Liu S., Chen G., Shi H., Wu L. Identification of small non-coding RNAs as sperm quality biomarkers for *in vitro* fertilization. *Cell Discov*. 2019;5:20.

# Grass frog (*Rana temporaria*) tandem repeats in the genome and oocytes' transcriptome

Dikaya V.<sup>1\*</sup>, Travina A.<sup>1</sup>, Komissarov A.<sup>2</sup>

<sup>1</sup> *Institute of Cytology, RAS, St. Petersburg, Russia*

<sup>2</sup> *Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia*

\* *animeshizik@gmail.com*

**Key words:** tandem repeats, oogenesis, satellite DNA, amphibian

**Motivation and Aim:** Tandem repeats (TR) or satellite DNAs' transcripts are of overwhelming majority at lamp-brush stage of oogenesis [2]. The fate of transcripts in late oogenesis stages remains obscure. TR of most organisms are not known because of TR species specificity. **Methods and Algorithms:** We used grass frog genome published (NCBI GCF\_905171775.1) and oocyte nuclei transcriptome got *de novo*. The cDNA was synthesized using a Mint-2 kit (Eurogen) and sequenced using Illumina HiSeq 2500 in Saint-Petersburg State University Resource Center. Adapters and optical duplicates were purified using the v2trim and rmdup programs. Ribosomal RNA was filtered by SortMeRNA. The alignment of the obtained message RNA reads was performed by the STAR program. *De novo* assembly of the transcriptome was performed with Trinity, Metabat2 and CheckM were used to detect contamination. The frog transcriptome was aligned to the detected genomes of the found contaminants. Matched regions were excised using the Python3 script from the assembly. The quality of the assembly was assessed by QUAST and BUSCO. Transcriptome annotation was performed using EggNOG, BLAST, Seq2Fun and Trinotate. In the genome, gene prediction was implemented by AUGUSTUS. Genome content of TR and transposable elements (TE) was determined with RepeatModeler2 and RepeatMasker. The search for tandem repeats was done by using the developed De Bruijn graph-based algorithm in Python2.7 and C++ in the genome and transcriptome. These repeats were localized on chromosomes using fluorescence in situ hybridization (FISH).

**Results:** The transcriptome was collected in 41 121 contigs. However, there were few annotated sequences in it. Trinotate annotated ~30% of the contigs in transcriptome. The BLAST annotation was only ~10%. Transcripts responsible for signaling molecules were most prevalent in the Trinotate annotation. Gene Ontology analysis revealed many transcripts related to nucleic acid binding, nucleus components, ribosome biogenesis, gene expression, and ribonucleoprotein complex formation. Total repeats content in genome is ~70% (~50% unclassified) and ~50% (~41% unclassified) in transcriptome. Tandem repeats are underrepresented in the genome and transcriptome but we found six well comprised TR ranging in length from 5 Mb to 114 Mb. This six TR were mapped on the frog chromosomes and localized to the pericentromeric regions with relative chromosome-specificity.

**Conclusion:** It was obtained that unknown, apparently species-specific repeats have a large number of representations in the data. The mode of TR transcription is under the data found on birds in similar studies: TR with genus specificity (TR 494A (S1a)) transcribed three orders less than species-specific 47A and 35B [1, 2], which probably reflects the feature common for TR transcription in gametogenesis.

**Acknowledgements:** This work was supported by Russian Science Foundation (grant No. 19-74-20102).

## References

1. Komissarov A.S. et al. New high copy tandem repeat in the content of the chicken W chromosome. *Chromosoma*. 2018;127:73-83.
2. Trofimova I., Krasikova A. Transcription of highly repetitive tandemly organized DNA in amphibians and birds: A historical overview and modern concepts. *RNA Biology*. 2016;12(13):1246-1257.

## **Dormice functional genome annotation project: the first insights on regulation of hibernation**

Gazizova G.<sup>1\*</sup>, Deviatiiarov R.<sup>1</sup>, Kozlova O.<sup>1</sup>, Komissarov A.<sup>2</sup>, Cherkasov A.<sup>1</sup>, Tyapkina O.<sup>3</sup>, Nurullin L.<sup>3</sup>, Shagimardanova E.<sup>1</sup>, Gusev O.<sup>4</sup>

<sup>1</sup> *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>2</sup> *ITMO University, St. Petersburg, Russia*

<sup>3</sup> *Federal Research Center “Kazan Scientific Center of RAS”, Kazan, Russia*

<sup>4</sup> *Department of Regulatory Transcriptomics for Medical Genetic Diagnostics, Graduate School of Medicine, Juntendo University, Tokyo, Japan*

\* *grgazizova@gmail.com*

**Key words:** hibernation, muscle atrophy, gene expression, promoter activity, dormice

*Motivation and Aim:* Hibernation is a phenomenon demonstrating in some groups of mammals and representing physiological adaptation for surviving seasonal severe environmental conditions. It is expressed by a long periods of a deep torpor characterized by a reduced metabolism, low body temperature and heart rate and interrupted by a short periods of an arousal. Remarkably, hibernating animals stay immobilized for a long time without significant features of hind limb muscle atrophy [1]. Edible dormouse (*Glis glis*) is a small arboreal rodent, which is superbly suitable for investigation of this phenomenon. Dormice are characterized by a long hibernation period that is 8 month and may be up to 11 month under unfavorable conditions [2]. At present time a plenty of molecular-genetic data on involvement of several molecules, including transcription factors, and signal pathways in the control of hibernation is accumulated [1, 3, 4], however, there are still no systematic data on the genome-wide promoters and enhancers activity in the tissues of a hibernating animal. The aim of present research is to identify molecular mechanisms regulating hypo-metabolic shifts and musculoskeletal adaptation during hibernation via analysis of promoter and enhancer activity in muscle of hibernator edible dormice.

*Methods and Algorithms:* Two groups of animals ( $n = 6$ ) were examined: active animals ( $n = 3$ ) kept in standard vivarium conditions with ambient temperature 23 °C and hibernated animals ( $n = 3$ ) kept in climatic camera with ambient temperature 0 °C for a 14 days. Two types of muscles were examined: “slow” type soleus muscle and “fast” type extensor digitorum longum (EDL) muscle. For genome assembly, libraries of Illumina short pair end reads, PacBio long reads and Hi-C reads were used. For genome annotation mRNA and total RNA sequencing of muscles and lumbar spinal cord were used. For analysis of gene expression and promoter and enhancer activity CAGE-seq (Cap Analysis Gene Expression) was applied.

*Results:* To identify molecular mechanisms regulating hypometabolic shifts and protective musculoskeletal adaptation “Dormice functional genome annotation project” was launched. For meaningful analysis of gene expression on promoter and enhancer level a well-assembled and annotated genome of edible dormouse is required. So far there was the edible dormouse’s genome assembly submitted to GenBank by scientific group from Broad University (BU) and based on Illumina HiSeq sequencing technology.

However, we attempted to complement and improve the dormouse genome assembly by long-read sequencing on Pacific Bioscience platform and applying Hi-C chromosome conformation capture technique. Here we provide our achievements on this issue and the first systematic annotation of promoters' landscape and genes activated in different types of skeletal muscles during hibernation.

According to preliminary study 8385 CAGE clusters were determined. Differential promoter expression analysis revealed a distinct program in regulation of hibernation. m. EDL response to hibernation ~2 times higher than m. Soleus. Differential promoter expression analysis and motif analysis demonstrate that response to hibernation is different in EDL and Soleus muscles. Interestingly, differentially expressed (DE) promoters are less rich by CpG regions than others. To identify conservative promoters we performed cross-species mapping and blastn analysis, which reveals 572 mouse conservative promoters in edible dormouse.

*Conclusions:* Obtained data facilitate the reconstruction of the key regulatory pathways and drivers involved in the regulation of hibernation and protective musculoskeletal adaptation in edible dormice. The study provides the first systematic annotation of promoters' landscape and genes activated in "fast" and "slow" muscle types during hibernation in edible dormice.

*Acknowledgements:* The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

### *References*

1. Carey H.V., Matthew T.A., Sandra L.M. Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. *Physiol Rev.* 2003;83(4):1153-1181.
2. Hoelzl F. et al. How to spend the summer? Free-living dormice (*Glis glis*) can hibernate for 11 months in non-reproductive years. *J Comp Physiol B.* 2015;185(8):931-939.
3. Vermillion K.L., Anderson K.J., Hampton M., Andrews M.T. Gene expression changes controlling distinct adaptations in the heart and skeletal muscle of a hibernating mammal. *Physiol Genomics.* 2015;47(3):58-74.
4. Storey K.B. Regulation of hypometabolism: insights into epigenetic controls. *J Experimental Biol.* 2015;218:150-159.



# Promoter motif inference and annotation of promoter sequences in bacterial genomes based on the analysis of structures of alternative sigma factor-promoter complexes

Gromyko D.\*, Vychyk P., Nikolaichik Y.

*Department of Molecular Biology, Belarusian State University, Minsk, Belarus*

\* *grom.dima.grom@gmail.com*

**Key words:** sigma factor, promoter, DNA motif, DNA binding domain

*Motivation and Aim:* Bacterial genome annotation pipelines can handle protein-coding sequences, rRNA and tRNA genes, but no pipeline deals with regulatory sequences. To close this void, we developed Sigmoid [1, 2] software for regulatory sequence analysis and annotation in bacterial genomes. Sigmoid can annotate transcription factor (TF) binding sites and terminators. It also contains a pipeline for *de novo* inference of TF binding site motifs, which could be used for annotating binding sites for uncharacterised TFs. The idea behind this pipeline relies on the conservation of positions of DNA bases-contacting amino acid residues of the TF. These conserved residues constitute the CR tag – the unique identifier of TF pair with its binding site. Due to the autoregulatory nature of most TFs, their DNA recognition motif could be inferred within a set of regulatory sequences of genes coding TFs with the same CR tag.

The same approach faces difficulties if applied to promoter identification due to a different mode of promoter recognition. Rather than making few specific contacts within a single short region of double-stranded DNA, sigma factors make many more base contacts within at least two promoter regions. In addition, many specific contacts are made within the single-stranded region of open promoter complexes. Therefore, CR tags derived from 3D structures of sigma factors bound to their promoters are excessively long for our pipeline resulting in failure due to insufficient diversity of the regulatory regions it collects. This work aims at modifying our pipeline to reliably infer promoter sequence motifs. We approach this problem by carefully analysing known sigma factor-promoter contacts and calculating covariation between CR tag and promoter sequence changes. We show that restricting the CR tag to the most significant positions allows efficient inference of promoter motifs for four families of sigma factors.

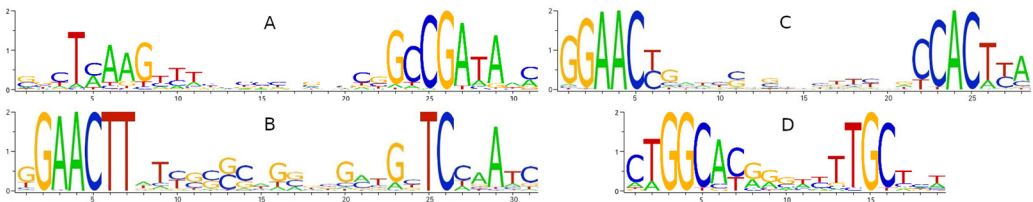
*Methods and Algorithms:* As changes in the promoter DNA sequences are followed by amino-acid replacements in the DNA-binding domain and *vice versa*, a correlation between residues in DNA-binding domains and positions in promoter sequences could reveal residues most important for specific promoter recognition. We looked for correlations with Prot-DNA-Korr [3], which counts Z-score for co-correlating amino-acid residue – DNA nucleotide pairs. Data sets for Prot-DNA-Korr were acquired via a custom Python script that collects sigma factor sequences and their promoter regions from the GenBank database and splits them into matching R2 domain/-10 region and R4 domain/-35 sets. PFAM and ECFhub were used as the source for the sigma factor family and promoter hidden Markov models. Collected sequences were aligned with Clustal Omega and further processed with Prot-DNA-Korr.

**Results:** To test our approach, four enterobacterial sigma factor families were selected. The main selection criterion was the availability of 3D structures of initiator complexes and characterized promoters. The four families had significant differences in DNA recognition mechanisms: ECF factors are characterized by the typical contacts described above between the promoter and the SR2 and SR4 domains, FliA has an insertion of the SR3 domain, also capable of contacting the promoter, while both DNA-binding domains of RpoN interact with double-stranded DNA.

With the help of Prot-DNA-Korr, we determined CR tag positions that are most important for specific promoter recognition (Fig. 1). This analysis allowed us to define CR tags for four sigma factor families: RpoE, HrpL, FliA and RpoN. The resulting CR tags have shortened from over 40 original contacts to 19 (or less) most important ones. The shortened CR-tags allowed the Sigmoid pipeline to successfully infer promoter motifs (Fig. 2) with high similarity to those described in the literature [4–8].



**Fig. 1.** Heatmap of correlations between SR2 domain amino acids and -10 promoter region nucleotides for RpoE sigma factor family



**Fig. 2.** Inferred promoter motifs for FliA (A), RpoE (B), HrpL (C) and RpoN (D)

**Conclusion:** We successfully expanded Sigmoid annotation capabilities, adding support for alternative sigma factor promoter inference. We are currently working to extend this approach to other families of bacterial sigma factors. Modified Sigmoid program code and calibrated profiles for identifying promoters of five alternative sigma factors are available in the repository [github.com/nikolaichik/sigmoid](https://github.com/nikolaichik/sigmoid). Calibrated profiles are suitable for automated annotation of promoters in bacterial genome sequences.

### References

1. Nikolaichik Y., Damienikan A.U. *PeerJ*. 2016;4:E2056.
2. Nikolaichik Y., Vychyk P. Abstr. BGRS/SB-2020:5-76. doi: 10.18699/BGRS/SB-2020-046.
3. Korostelev Y.D. et al. *PLoS One*. 2016;11(9) E0162681.
4. Lam H.N. et al. *PLoS One*. 2014;9(8):E106115.
5. Fitzgerald D.M. et al. *Mol Microbiol*. 2018;108(4):361-378.
6. Missiakas D., Raina S. *Mol Microbiol*. 1998;28(6):1059-1066.
7. Markel E. et al. *J Bacteriol*. 2016;198(17):2330-2344.
8. Wilson M.J., McMorran B.J., Lamont I.L. *J Bacteriol*. 2001;183(6):2151-2155.

# Tocilizumab partially reverses the altered transcriptional regulation of glycolysis in circulating CD8<sup>+</sup> T cells of rheumatoid arthritis patients

Harshan S.<sup>1,2\*</sup>, Dey P.<sup>3</sup>, Srivatsan R.<sup>1</sup>

<sup>1</sup> Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Bengaluru, India

<sup>2</sup> Manipal Academy of Higher Education, Manipal, Karnataka, India

<sup>3</sup> he Edu Bio LLC, Northwest Registered Agent Service Inc., Okemos, Michigan, USA

\* shilpa.harshan@gmail.com

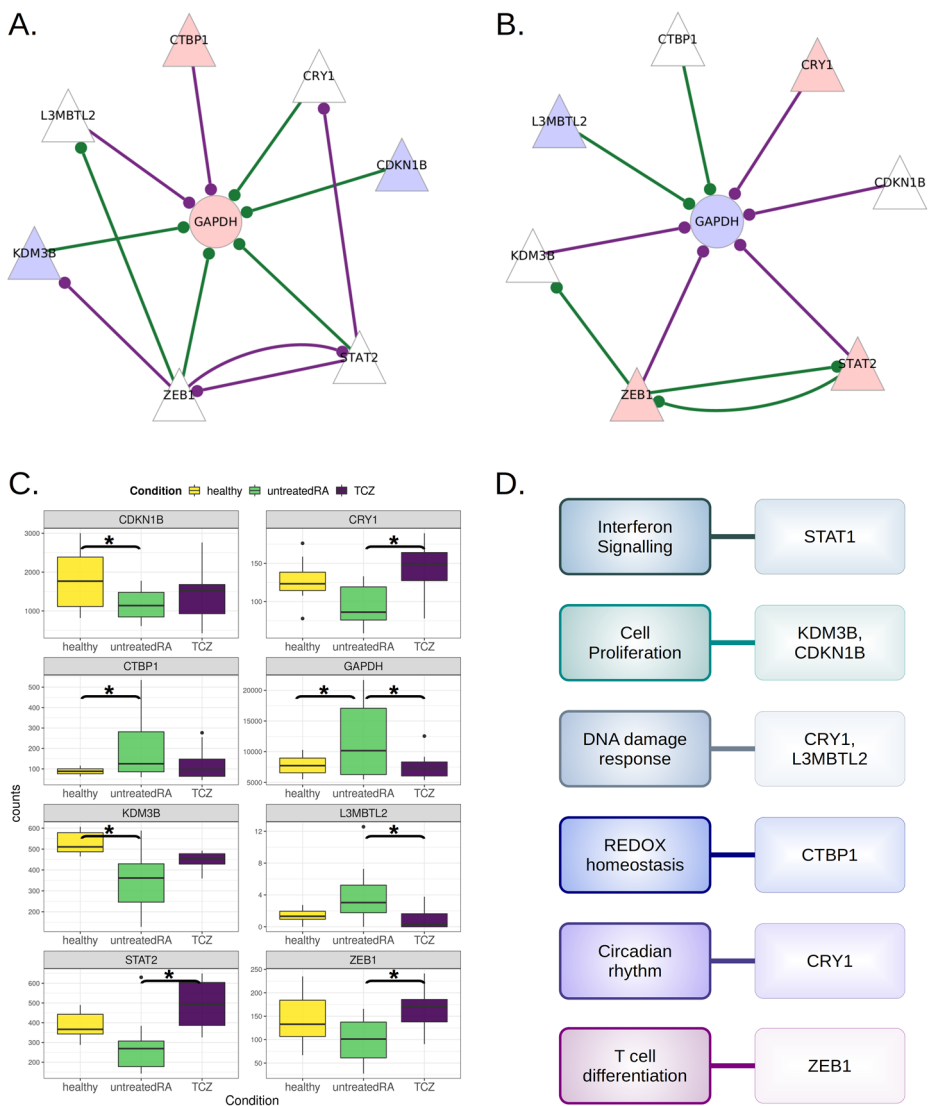
**Key words:** rheumatoid arthritis, tocilizumab, T lymphocyte, RNA seq, gene expression

*Background:* Peripheral T cells of rheumatoid arthritis (RA) patients show pathological changes. CD4<sup>+</sup> T cells use pentose phosphate pathway instead of glycolysis, resulting in a reductive phenotype. These changes drive increased proliferation and tissue invasiveness of these cells. CD8<sup>+</sup> T cells also show changes to the glycolysis pathway. However, the transcriptional regulation underlying the altered glycolysis in these cells is poorly understood. The effect of drugs used in the treatment of RA on glycolysis genes and their regulators also remains to be elucidated.

*Method:* We analysed publicly available RNA sequencing data from circulating CD4<sup>+</sup> and CD8<sup>+</sup> T lymphocyte subsets of healthy, untreated and treated RA individuals to identify differentially expressed genes (DEGs). Glycolysis related genes and transcription factor genes were selected to create differential co-expression networks for each cell type. Sample-wise edge weights were calculated using for all possible gene-pairs using “linear interpolation to obtain network estimates for single sample” (lionessR) analysis, and the gene-pairs with significantly different edge weights were identified using limma. The networks were annotated using the Gene Transcription Regulation Database (GTRD) and analysed to find high centrality genes.

*Results:* CD8<sup>+</sup> effector memory (Tem) cells and CD8<sup>+</sup>CD45RA<sup>+</sup> effector memory (Temra) cells showed large changes to the transcriptional regulation of glycolysis in untreated RA. GAPDH was upregulated and had high centrality scores in the untreated RA CD8<sup>+</sup> Tem cells. Its upregulation in the untreated RA CD8<sup>+</sup> Tem cells was reversed in the tocilizumab treated cells. The transcription factors that show opposing differential expression with GAPDH in the untreated and tocilizumab treated CD8<sup>+</sup> Tem cells were identified. Tocilizumab treatment partially reversed the RA associated differential expression of genes involved in fatty acid metabolism, citric acid cycle, oxidative phosphorylation complex 1 and ATP synthase. However, PFKFB3, which regulates glycolysis, is not affected in the tocilizumab treated CD8<sup>+</sup> cells.

*Conclusion:* Tocilizumab treatment is associated with a return to a healthy like expression state for several genes. However, tocilizumab does not restore PFKFB3 levels indicating that its effect on glycolytic activity in these cells is limited.



**Fig. 1.** A. GAPDH and its neighbors in the untreated RA network. Red nodes are upregulated and blue nodes are down regulated. Purple edges show higher mean edge weight in untreated RA samples. Green edges show higher mean edge weight in healthy samples. B. GAPDH and its neighbors in the toclizumab RA network. Purple edges show higher mean edge weight in the toclizumab samples. Green edges show higher mean edge weight in the untreated RA samples. C. Expression values of GAPDH and its differentially expressed neighbors in the healthy, untreated RA and toclizumab treated RA CD8<sup>+</sup> Tem cells. Statistically significant differential expression is indicated by an asterisk (\*). D. The functions of the differentially expressed transcription factor neighbours of GAPDH in the CD8<sup>+</sup> Tem cells

*Acknowledgements:* This work was fully supported by the Department of Electronics, IT, BT and S&T of the Government of Karnataka. Shilpa Harshan is a PhD student registered under Manipal Academy of Higher Education.

## Transcription factor binding sites: data integration, stable identifiers and incremental builds

Kolmykov S.<sup>1,2\*</sup>, Kondrakhin Y.<sup>2,3</sup>, Sharipov R.<sup>1,2,4</sup>, Yevshin I.<sup>1,2</sup>, Ryabova A.<sup>1,2</sup>, Kolpakov F.<sup>1,2,3</sup>

<sup>1</sup> *Sirius University of Science and Technology, Sochi, Russia*

<sup>2</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>3</sup> *Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia*

<sup>4</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *kolmykovsk@gmail.com*

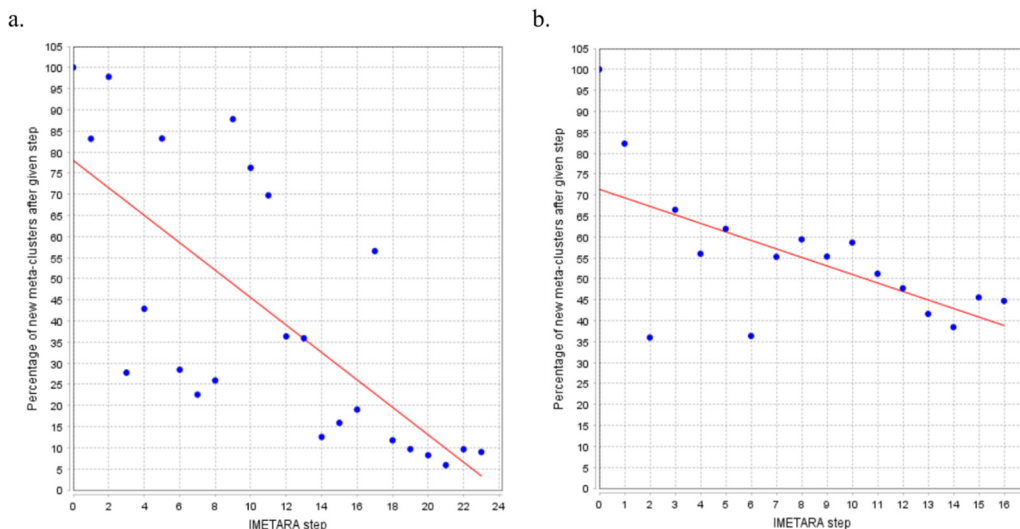
**Key words:** transcription factor binding sites, data integration, stable identifiers, incremental builds

*Motivation and Aim:* Transcriptional regulation is a complex mechanism at different levels, and transcription factors (TF) play a key role in this process. Most TFs bind to DNA by recognizing their short regulatory elements called TF binding sites (TFBSs). These TFBSs are the key components of transcriptional regulation, as well. One of the most widely used methods for experimental identification of TF binding sites (TFBS) is chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq). Currently, TFBS datasets from various ChIP-seq experiments are being accumulated in databases such as GTRD (<http://gtrd.biouml.org>), ChIP-Atlas (<https://chip-atlas.org/>), ReMap (<https://remap2022.univ-amu.fr/>) and ENCODE (<https://www.encodeproject.org/>).

*Methods and Algorithms:* For meta-processing of individual ChIP-seq datasets for a given TF, the METARA method (METa Analysis of ChIP-seq datasets using a Rank Aggregation approach) was previously developed. It produced a single integrated meta-dataset (known as a metacluster set in GTRD) by processing datasets obtained from individual ChIP-seq experiments for the given TF. It is important to note that the special Rank Aggregation score (RA-score) was assigned to each meta-cluster. It evaluates the quality (or reliability) of each meta-cluster.

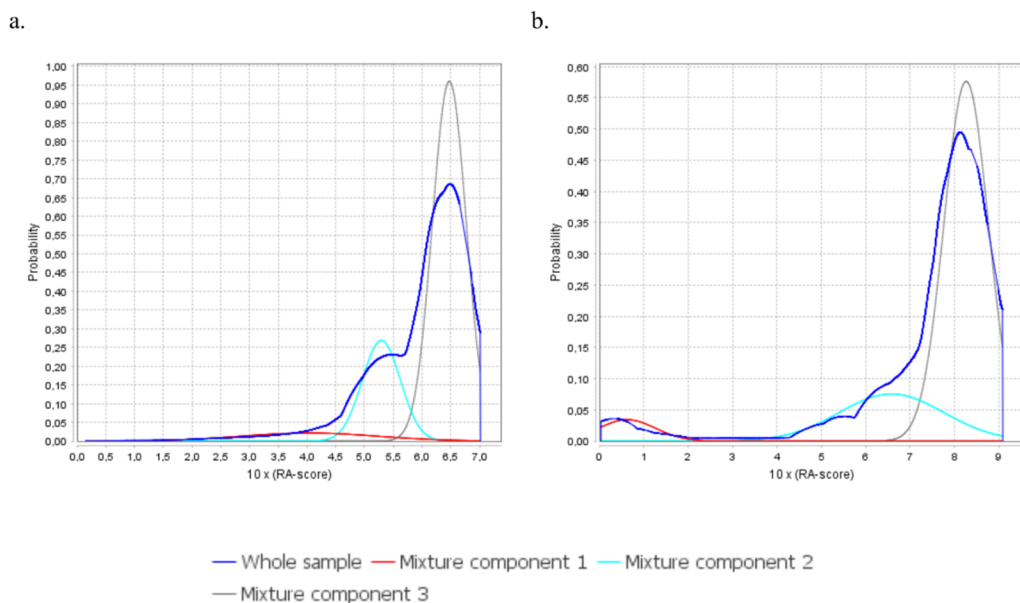
*Results:* We have developed a new method IMETARA (Incremental METARA). There are two main reasons for creating IMETARA: (a) to significantly reduce the run time of meta-cluster building; (b) introduce stable identifiers for each meta-cluster. Thus, IMETARA only recalculates the RA scores for existing meta-clusters and adds novel meta-clusters when several new ChIP-seq experiments are added to the ChIP-seq data already collected in GTRD.

It is important to note that there is a strong IMETARA continuity, namely that most (99.3–99.9 %) of the meta-clusters identified by METARA are also identified by IMETARA for many TFs. Transition from METARA to IMETARA makes it easier to control the completeness of the resulting set of meta-clusters. For this purpose, it is sufficient to consider the dependence of such a characteristic as "Percentage of new meta-clusters at this IMETARA step". Figure 1 demonstrates that the new steps result in only 5–10 % of new meta-clusters for ESR1 being included, while this percentage for FOXA1 is about 35–40 %.



**Fig. 1.** Percentage of new meta-clusters at the given IMETARA step for (a) ESR1 and (b) FOXA1

The RA scores can be interpreted as reliability scores. As a result, the algorithm of decomposition of a sample of RA-scores into several components of normal mixture (Fig. 2) demonstrates that meta-clusters can be divided into 2–3 reliability groups.



**Fig. 2.** Normal mixture of RA-scores for (a) JUND and (b) REST

*Acknowledgements:* The study was supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. 075-15-2021-1344.

## Lymphocyte migration factors in breast cancer: a transcriptome analysis

Korobeynikova A.V.<sup>1\*</sup>, Orlov Y.L.<sup>1,3</sup>, Kabieva S.S.<sup>1</sup>, Oparina N.Y.<sup>2</sup>

<sup>1</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

<sup>2</sup> University of Gothenburg: Department of Rheumatology a Inflammation Research, Gothenburg, Sweden

<sup>3</sup> Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia

\* aaaniich@yandex.ru

**Key words:** cancer, biomedicine, bioinformatics, e-Health, immunotherapy

*Motivation and Aim:* Breast cancer (the overgrowth and breast tissue cells proliferation) is currently documented as one of the most common types of cancer. Immunotherapy is high-potential treatment for many types of cancer, but it is a newly approach for fighting against breast cancer [1], so the prospect of studying this area is steadily growing, as it has been proven that breast cancer is immunogenic tumor. Immunotherapy has become a significant treatment strategy and enhanced the survival of breast cancer patients in the intervening years. Despite the success of this type of treatment, some patients do not respond to therapy or responding patients tend to recrudescence or worsen. The main causes of these adverse effects are complex internal or external mechanisms of resistance. It is also assumed that this is due to the low content (compared to other types of oncology) of tumor-infiltrating lymphocytes (TIL) in malignant tumors [2]. It is highly important to scrutinize the process of infiltration and migration of immune cells in breast cancer in order to improve modern immunotherapeutic interventions. For this reason we use data from the lymph nodes where the maturation of lymphocytes occurs.

*Methods and Algorithms:* We used open databases to collect transcriptomic data from lymph nodes of donors and patients with breast cancer. The objective was to compare gene expression and identify candidate genes for the analysis of lymphocyte migration in oncology. As a starting point normalization and rank standardization of data from heterogeneous platforms (microarray and RNA-seq) was carried out. In the sequel genes associated with CD4 cell lines were selected where expression differences between normal/control samples were found. With the help of GeneCards, an annotation of functional significance was obtained. The differential expression of the identified genes was tested for the same patients directly in tumor cells in order to obtain an immunological prognosis and obtain data on the migration of lymphocytes from the lymph nodes. Subsequently, these data can be used to develop personalized approaches to the prognosis of immunotherapy in patients with breast cancer, improve response and eliminate relapses.

*Results:* Immune response markers in oncology – NCAM1 and SIGLEC7 – revealed overexpression in lymph nodes in cancer samples. Proliferation processes, dependent on IL7, turned out to be active, while independent (SLAMF1), on the contrary, are inhibited. CD38, CD98, and SLAMF1 have low expression ranks and are associated with the presence of activated CD4 cells. Genes associated with the CD4+ cell line showed low differential expression in tumor samples from patients.

*Conclusion:* By the means of transcriptomic analysis, results indicating blocking of some pathways of proliferation of lymphocyte cell lines were obtained, as well as problems of migration into tumor cells. Incomplete picture is spotted: there is a great deal of evidence on reversible cells (eg, peripheral blood) and data on metastasized specimens, but a lack of clean sets from lymph nodes. We consider it promising to study the migration of T cells from lymph nodes for the progress in cancer immunotherapy. This may provide answers to the question of the difference in the patients sensitivity to this type of treatment and palindromia later on.

### *References*

1. Harbeck N., Gnant M. Breast cancer. *Lancet*. 2017;389(10074):1134-1150. doi: 10.1016/S0140-6736(16)31891-8.
2. Zacharakis N., Huq L.M., Seitter S.J., Kim S.P., Gartner J.J., Sindiri S., Hill V.K., Li Y.F., Paria B.C., Ray S., Gasmi B., Lee C.C., Prickett T.D., Parkhurst M.R., Robbins P.F., Langan M.M., Shelton T.E., Parikh A.Y., Levi S.T., Hernandez J.M., Hoang C.D., Sherry R.M., Yang J.C., Feldman S.A., Goff S.L., Rosenberg S.A. Breast cancers are immunogenic: Immunologic analyses and a phase II pilot clinical trial using mutation-reactive autologous lymphocytes. *J Clin Oncol*. 2022;40(16):1741-1754. doi: 10.1200/JCO.21.02170.
3. Huttenlocher A., Sandborg R.R., Horwitz A.F. Adhesion in cell migration. *Curr Opin Cell Biol*. 1995;7(5):697-706. doi: 10.1016/0955-0674(95)80112-x.
4. Goff S.L., Danforth D.N. The role of immune cells in breast tissue and immunotherapy for the treatment of breast cancer. *Clin Breast Cancer*. 2021;21(1):e63-e73. doi: 10.1016/j.clbc.2020.06.011.
5. Edechi C.A., Ikeogu N., Uzonna J.E., Myal Y. Regulation of immunity in breast cancer. *Cancers (Basel)*. 2019;11(8):1080. doi:10.3390/cancers11081080.
6. Gumbiner B.M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell*. 1996;84(3):345-357. doi: 10.1016/s0092-8674(00)81279-9.



# Interpreting non-coding genome variation with DNA sequence motifs

Kulakovskiy I.V.<sup>1, 2\*</sup>, Makeev V.J.<sup>2, 3</sup>

<sup>1</sup> *Institute of Protein Research, RAS, Pushchino, Russia*

<sup>2</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

<sup>3</sup> *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

\* *ivan.kulakovskiy@gmail.com*

**Key words:** non-coding genome variation, regulatory SNP, transcription factor binding, sequence motifs

Nowadays, it is impossible to imagine the future of human healthcare without personalized biomedicine and pharmacogenomics. While genome sequencing is becoming cheaper and more reliable, overall interpretability and thus the practical utility of the individual genome data remain limited. On the one hand, genome-wide association studies<sup>1</sup> reveal millions of statistical associations between genome variants and various traits, including predisposition to complex diseases, such as immune disorders and cancer. On the other hand, the detected associations do not immediately reveal the genuine causal variants, particularly, because of linkage disequilibrium. Prioritization of the causal variants requires statistical fine-mapping or functional annotation in a wet lab or *in silico*. Interpretation and prioritization are especially challenging for the non-coding variants in gene regulatory regions, which do not alter protein sequence and function directly. Instead, they disturb the cellular state by altering gene expression and protein abundance. For many pathologic phenotypes, such as immune disorders, the majority of causal variants are located in the transcription regulatory regions. Annotating individual regulatory elements of gene regulatory regions facilitates functional annotation of non-coding variants, and reveals underlying molecular mechanisms.

We present a versatile bioinformatics framework for the analysis of high-throughput sequencing data in the context of gene regulation and transcription factor binding, including *de novo* DNA motif discovery (ChIPMunk [1]), pattern matching (SARUS [2] and MoLoTool [3]), motif comparison (MACRO-APE [4]), and basic annotation of noncoding variants (PERFECTOS-APE [5]). With this toolkit, we assembled the well-known HOCOMOCO [3] collection of sequence motifs recognized by human TFs and used it to annotate single-nucleotide variants linked with autoimmune diseases. On top of that, we show how advanced statistics backed up with motif analysis allows for mapping causative allele-specific regulatory elements (ADASTRA [6] and ANANASTRA [7]), and discuss surprising pitfalls and peculiarities of machine learning applications in the analysis of regulatory variants [8].

*Acknowledgements:* Russian Science Foundation (20-74-10075 to I.V.K.).

## References

1. Kulakovskiy I.V. et al. Deep and wide digging for binding motifs in ChIP-Seq data. *Bioinformatics*. 2010;26(20):2622-3. doi: 10.1093/bioinformatics/btq488.
2. Kulakovskiy I.V. et al. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res*. 2013;41(D1):D195-D202. doi: 10.1093/nar/gks1089.
3. Kulakovskiy I.V. et al. 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(D1):D252-D259. doi: 10.1093/nar/gkx1106.

4. Vorontsov I.E. et al. Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms Mol Biol.* 2013;8(1):23. doi: 10.1186/1748-7188-8-23.
5. Vorontsov I.E. et al. PERFECTOS-APE: Predicting regulatory functional effect of SNPs by approximate P-value estimation. In: Proceedings of the 6th International Conference on Bioinformatics Models, Methods and Algorithms, BIOINFORMATICS 2015. Lisbon, 2015;102-108. doi: 10.5220/0005189301020108.
6. Abramov S. et al. Landscape of allele-specific transcription factor binding in the human genome. 2015. *Nat Commun.* 2021;12:2751. doi:10.1038/s41467-021-23007-0.
7. Boytsov A. et al. ANANASTRA: annotation and enrichment analysis of allele-specific transcription factor binding at SNPs. *Nucleic Acids Res.* 2022;gkac262. doi:10.1093/nar/gkac262.
8. Penzar D.D. et al. What do neighbors tell about you: the local context of cis-regulatory modules complicates prediction of regulatory variants. *Front Genet.* 2019;10:1078. doi: 10.3389/fgene.2019.01078.

## Web-MCOT server for motifs co-occurrence search in ChIP-seq data

Levitsky V.<sup>1,2\*</sup>, Mukhin A.<sup>1</sup>, Oshchepkov D.<sup>1</sup>, Zemlyanskaya E.<sup>1,2</sup>, Lashin S.<sup>1,2</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* [levitsky@bionet.nsc.ru](mailto:levitsky@bionet.nsc.ru)

A DNA sequence element possessing transcriptional regulatory activity typically consists of multiple binding sites of several TFs. The analysis of massive ChIP-seq data provides the lists of overrepresented motifs usually interpreted as potential binding sites of TFs. Since commonly TFs bind DNA cooperatively, the first step in this analysis is the search of enriched co-occurred motifs. Here we present the web server Motifs Co-Occurrence Tool (Web-MCOT) that for a single ChIP-seq dataset detects the composite elements (CEs) or overrepresented homo- and heterotypic pairs of motifs with spacers and overlaps, with any mutual orientations, uncovering various similarities to recognition models within pairs of motifs. The first (Anchor) motif in CEs respects the target transcription factor of ChIP-seq experiment, while the second one (Partner) can be defined either by a user or a public library of partner motifs is processed. The public libraries of motifs for TFs of human/mouse and plants can be applied in analysis. The web server tests the significance of motifs similarity to filter out possible false predictions related with a significant match between the Anchor and Partner motifs. Web-MCOT computes significances of CEs without reference to motifs conservation, including separate analysis of CEs with overlaps of motifs and those with spacers. The web server computes the significances of CEs with more conserved Partner and Anchor motifs. Graphic results show histograms of CEs abundance depending on orientations of motifs, their full or partial overlap and spacer lengths. For each pair of Anchor and Partner motifs the web server shows the logo of the most common CE structural type with an overlap of motifs, and heatmaps depicting abundance of CEs with one motif possessing higher conservation than another. Additionally, the web interface allows sorting of CEs with partner motifs of different TFs according to the significance of the co-occurrence. Web-MCOT is available at <https://webmcot.sysbio.cytogen.ru/> and <https://webmcot-euro.sysbio.cytogen.ru/>.

*Acknowledgements:* The study was supported by Russian Science Foundation project No. 21-14-00240.

## Alternative transcription start sites contribute to acute-stress-induced transcriptome response in human skeletal muscle

Makhnovskii P.<sup>1\*</sup>, Gusev O.<sup>2,3</sup>, Bokov R.<sup>1</sup>, Gazizova G.<sup>3</sup>, Vepkhvadze T.<sup>1</sup>,  
Lysenko E.<sup>1,4</sup>, Vinogradova O.<sup>1,4</sup>, Kolpakov F.<sup>5</sup>, Popov D.<sup>1,4</sup>

<sup>1</sup> *Institute of Biomedical Problems, RAS, Moscow, Russia*

<sup>2</sup> *RIKEN Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan*

<sup>3</sup> *Institute of Fundamental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russia*

<sup>4</sup> *Faculty of Fundamental Medicine, Lomonosov Moscow State University, Moscow, Russia*

<sup>5</sup> *Institute of Computational Technologies, SB RAS, Novosibirsk, Russia*

\* *maxpauel@gmail.com*

More than half of human protein-coding genes have an alternative transcription start site (TSS). We aimed to investigate the contribution of alternative TSSs to the acute-stress-induced transcriptome response in human tissue (skeletal muscle) using the cap analysis of gene expression approach. TSSs were examined at baseline and during recovery after acute stress (a cycling exercise). We identified 44,680 CAGE TSS clusters (including 3,764 first defined) belonging to 12,268 genes and annotated for the first time 290 TSSs belonging to 163 genes. The transcriptome dynamically changes during the first hours after acute stress; the change in the expression of 10 % of genes was associated with the activation of alternative TSSs, indicating differential TSSs usage. The majority of the alternative TSSs do not increase proteome complexity suggesting that the function of thousands of alternative TSSs is associated with the fine regulation of mRNA isoform expression from a gene due to the transcription factor-specific activation of various alternative TSSs. We identified individual muscle promoter regions for each TSS using muscle open chromatin data. Then, using the positional weight matrix approach we predicted time course activation of “classic” transcription factors involved in response of skeletal muscle to contractile activity, as well as diversity of less/un-investigated factors. In conclusion, transcriptome response induced by acute stress strongly related to activation of the alternative TSSs indicates that differential TSSs usage is an essential mechanism of fine regulation of gene response to stress stimulus.

*Acknowledgements:* This work was supported by the Russian Foundation for Basic Research (RFBR) grant No. 20-015-00415A.

## Amplification of the release factor genes as a mechanism of adaptation in yeast

Matveenko A.G.<sup>1\*</sup>, Maksiukenko E.M.<sup>2,1</sup>, Barbitoff Y.A.<sup>1,3</sup>, Moskalenko S.E.<sup>2,1</sup>, Zhouravleva G.A.<sup>1,4</sup>

<sup>1</sup>Department of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia

<sup>2</sup>St. Petersburg Branch, Vavilov Institute of General Genetics, RAS, St. Petersburg, Russia

<sup>3</sup>Bioinformatics Institute, St. Petersburg, Russia

<sup>4</sup>Laboratory of Amyloid Biology, St. Petersburg State University, St. Petersburg, Russia

\* a.matveenko@spbu.ru

**Key words:** release factors, yeast, eRF1, eRF3, gene amplification, copy number variation, chromosome instability, WGS

*Motivation and Aim:* Termination of translation in eukaryotes is conferred by two release factors, eRF1 and eRF3. In baker's yeast these factors are encoded by essential genes *SUP45* and *SUP35*. Despite being essential, several viable nonsense mutations in these genes have been discovered [1, 2]. The survival of such mutants involves feedback control of stop codon readthrough, however, the exact molecular basis remains unclear.

*Methods and Algorithms:* To find out the mechanisms of yeast adaptation to nonsense mutations in the release factor genes, we performed whole genome sequencing of 100 yeast strains carrying various mutant alleles of the *SUP35* and *SUP45* genes (*sup35-n/sup45-n*) on centromeric plasmids using Illumina HiSeq 2500. Additionally, 13 strains carrying *sup35-n/sup45-n* mutant alleles on the chromosomes were sequenced and analyzed. The sequencing data were aligned to the reference genome of the U-1A-D1628 strain, obtained previously [3], after which the search for genetic variations and copy number analysis were performed. The number of copies of the release factor genes was additionally confirmed with qPCR.

*Results:* SNV analysis of WGS data did not show any frequent changes in the yeast genome, which could potentially be adaptive to translation termination malfunction. At the same time, we found a significant increase in the number of copies of plasmids carrying *sup35-n/sup45-n* mutant alleles compared to wild-type alleles, which was confirmed using qPCR. In four sequenced strains carrying the *sup35-n/sup45-n* alleles as a single chromosomal copy, we found a duplication of the entire chromosome II (containing the *SUP45* gene) or a region on the chromosome IV, bearing the *SUP35* gene [4]. Interestingly, the borders of the duplicated region on chromosome IV are exactly the same as in previously reported duplication in the *sup35-222* strain, containing a downregulating single-nucleotide substitution in the *SUP35* promoter [5]. In addition, three of the 13 strains were found to have aneuploidies on chromosomes XI and XIII, which do not contain alleles of the release factor genes.

*Conclusion:* We thus conclude that increasing gene dosage of the *SUP45* and *SUP35* genes is a primary mechanism for increasing the survival of cells with impaired termination of translation. The observed genomic instability which did not alter the release factor genes copy number indicates, that other mechanisms might also be involved, which, however, are less wide-spread.

*Acknowledgements:* The study is supported by the RSF grant 18-14-00050, by the State research program 0112-2016-0015, as well as by the resource centres “Development of Molecular and Cell Technologies” and “Biobank” of SPbSU.

### *References*

1. Moskalenko S.E., Chabelskaya S.V., Inge-Vechtomov S.G., Philippe M., Zhouravleva G.A. Viable nonsense mutants for the essential gene *SUP45* of *Saccharomyces cerevisiae*. *BMC Mol Biol.* 2003;4:2.
2. Chabelskaya S., Kiktev D., Inge-Vechtomov S., Philippe M., Zhouravleva G. Nonsense mutations in the essential gene *SUP35* of *Saccharomyces cerevisiae* are non-lethal. *Mol Genet Genomics.* 2004;272(3):297-307.
3. Barbitoff Y.A., Matveenko A.G., Matiiv A.B., Maksiutenko E.M., Moskalenko S.E., Drozdova P.B., ..., Zhouravleva G.A. Chromosome-level genome assembly and structural variant analysis of two laboratory yeast strains from the Peterhof Genetic Collection lineage. *G3 (Bethesda).* 2021;11(4):jkab029.
4. Maksiutenko E.M., Barbitoff Y.A., Matveenko A.G., Moskalenko S.E., Zhouravleva G.A. Gene amplification as a mechanism of yeast adaptation to nonsense mutations in release factor genes. *Genes (Basel).* 2021;12(12):2019.
5. Matveenko A.G., Drozdova P.B., Moskalenko S.E., Tarasov O.V., Zhouravleva G.A. Whole genome sequencing data and analyses of the underlying *SUP35* transcriptional regulation for a *Saccharomyces cerevisiae* nonsense suppressor mutant. *Data Brief.* 2019;23:103694.

## **Functional annotation of lncRNAs involved in epigenetic regulation**

Mazurov E., Antonov I., Zubritskiy A., Budkina A., Matveishina E., Makeev V.J., Kulakovskiy I.V., Medvedeva Y.A.

*The Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia*

**Key words:** long non-coding RNA annotation, RNA-DNA interaction, chromatin regulation

Many long non-coding RNAs have been experimentally associated with chromatin. Yet, for the majority of chromatin-bound lncRNAs, the regulatory mechanism remains unclear. We developed a computational pipeline to identify lncRNAs that could affect DNA methylation, histone modifications, and, consequently, gene expression. Briefly, the pipeline consists of three major steps: (1) correlation analysis between a lncRNA expression and a presence of an epigenetic mark in a specific genomic location; (2) functional analysis of gene expression change in response to a lncRNA knockdown; (3) colocalization analysis of a lncRNA and target genomic region. We applied the pipeline to ten histone modifications and DNA methylation profiles in multiple cell types available from ENCODE, for functional analysis we used NNN lncRNA knockdown experiments from FANTOM6 Pilot. We detected several dozens of lncRNAs potentially involved in writing or erasing a specific histone modification, as well as putative genomic regions where the respective modifications are regulated by particular lncRNAs, both previously reported and unknown.

For several detected lncRNA-histone modification pairs we performed ChIP-seq analysis in the knockdown system to validate the link. At least three lncRNAs are extremely promising candidates for histone modification regulators.

## Evolutionary role of genomic DNA physical properties in the realization of its elements' biological functions

Mukhina K.<sup>1</sup>, Osypov A.<sup>1,2\*</sup>

<sup>1</sup> Institute of Theoretical and Experimental Biophysics of RAS, Pushchino, Russia

<sup>2</sup> Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia

\* aosypov@gmail.com

**Key words:** DNA electrostatics, transcription regulation, genome evolution

*Motivation and Aim:* DNA is highly charged and thus electrostatic interactions with proteins are important in the regulation of transcription. The electrostatic potential is distributed unevenly along the DNA, correlates with GC %, the sequence and its context [1]. Transcription factor binding sites are located in regions with a high electrostatic potential, in areas many times the size of the protein. The distribution of the electrostatic potential on the surfaces of transcription factor molecules reflects the structure of binding sites. Promoters on average have a significantly increased value of the electrostatic potential, the distribution of which correlates with their biological properties. Rho-independent terminators show the same increase in EP, apparently serving to slow down the polymerase to form the terminator hairpin on RNA. Thus, the distribution of the electrostatic potential is an important factor of natural selection that affects the evolution of DNA regulatory regions [1, 2].

Hundreds of transcription factor binding sites lie in protein coding regions, their percentage is not very high, but significant for cell functioning. Some protein coding regions also contain promoters for genes mostly located on the opposite strand.

*Methods and Algorithms:* DEPPDB and its tools [1, 2] were used to carry out the analysis.

*Results:* Bacteria belonging to such different evolutionary lineages as Proteobacteria (*E. coli*, representatives of the genera *Salmonella*, *Shigella*, *Yersinia*, *Shewanella*), Firmicutes (*B. subtilis*), and Actinobacteria (*Corynebacterium glutamicum*) were found to have a bias relative to the average distribution of codons and even amino acids around promoters and transcription factor binding sites spanning several hundred codons. This shift is due to the physical properties of the codons in question, providing the correct electrostatic attractors for proteins that regulate transcription (transcription factors and RNA polymerase). The deviation of occurrence from the average level reaches 300 % for codons and 200 % for amino acids. The picture is especially pronounced for such amino acids as glycine, serine, lysine and aspartic acid.

Synonymous codons have, as a rule, the opposite tendency to change in frequency depending on the calculated value of the electrostatic potential they form, the more pronounced, the higher the potential deviation and the frequency of occurrence of the codon and the amino acid itself (leucine is a striking example).

*Conclusion:* The data obtained require a serious revision of the concept of molecular evolution. The revealed non-synonymity of synonymous substitutions can lead to incorrect assessment of the evolutionary fate of sequences, including incorrect use of the molecular clock and erroneous estimates of the specificity of mutations. The shift in amino acid preferences leads to an even more important revision of the concept of the



proteins natural selection. It substantiates the idea of DNA not only as a text for the implementation of the first step of the Central Dogma of Molecular Biology, but as a complex organ of heredity that fulfills different, sometimes contradictory, requirements, such as packaging and storage, regulation of the implementation of hereditary information, and, finally, encoding proteins and functional non-coding RNAs, whose properties, in turn, depend on the primary structure. These ideas correspond to the ideas of the great evolutionist I.I. Schmalhausen about non-redundant duplication, which he developed in relation to the structures and functions of the animal organism.

Apparently, DNA electrostatics plays an important and universal role in the regulation of prokaryotic transcription, affecting the probability of protein binding and their positioning. The effect of EP on the functioning of the genome is closely related to the effects of other physical and structural properties, first of all, thermal stability, bending, and supercoiling. They can interact both at the stage of their formation and in the process of functional action. Since the physical properties of DNA often depend on the sequence on a statistically averaged basis, information is lost at the stage of their formation. This can work in coding regions due to the redundancy of the genetic code, and different physical properties can operate without interfering with each other.

All this can affect horizontal gene transfer, the evolution of transcription regulation systems, and contribute to AT enrichment of regulatory regions of the genome, and also allows us to put forward a hypothesis about the evolutionary originative role of DNA physical properties in the formation of the transcription regulation system. The regularities of the formation of EP and other physical properties affect such fundamental problems as the 2nd Chargaff rule, the redundancy of the genetic code, the evolutionary non-neutrality of synonymous substitutions and serve as a justification for the fundamental idea of the DNA phenotype, defining the principles of biophysical bioinformatics.

Thus, the electrostatic properties of genomic DNA in combination with other physical properties are very important for the evolution of the realization of its biological functions, primarily for the regulation of transcription. The task of presenting the relevant information is designed to be solved by the database DEPPDB developed by us, available for academic use at <http://deppdb.iteb.ru>.

*Acknowledgements:* The study is supported by the RFBR (16-04-01865).

### *References*

1. Osypov A.A. et al. DEPPDB – DNA electrostatic potential properties database. electrostatic properties of genome DNA. *JBCB*. 2010;8(3):413-425.
2. Osypov A.A. et al. DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA elements. *JBCB*. 2012;10(2):1241004.

# Massive comparison of the ‘genomic’ and ‘shuffled’ background set generation approaches for efficiency of *de novo* motif search in *A. thaliana* ChIP-seq data

Raditsa V.V.\*, Tsukanov A.V., Levitsky V.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *radicav06@gmail.com*

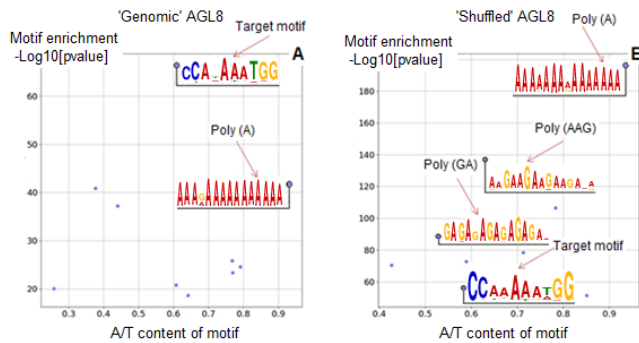
**Key words:** background data, transcription factor binding sites, genome sequence bias

*Motivation and Aim:* Transcription factors (TFs) are proteins that control gene transcription by sequence-specific DNA binding. Chromatin immunoprecipitation (ChIP)-based high throughput technique ChIP-seq allows genome-scale mapping of TF binding sites (TFBS). Motif enrichment analyses in the foreground data of ChIP-seq loci depend on the choice of the approach of background data generation. The task of this analysis consists in detection of motifs specific for biological function of a ChIP-seq target TFs. Hence, it is critically important to pay attention to false positive genomic sequence biases, such as polyA tracts, which most probably do not respect to interactions of target TFs with genomic DNA [1, 2]. However, up to now a systematic massive analysis of efficiency of various approaches of the background set generation for motif enrichment analyses in ChIP-seq data has not been performed.

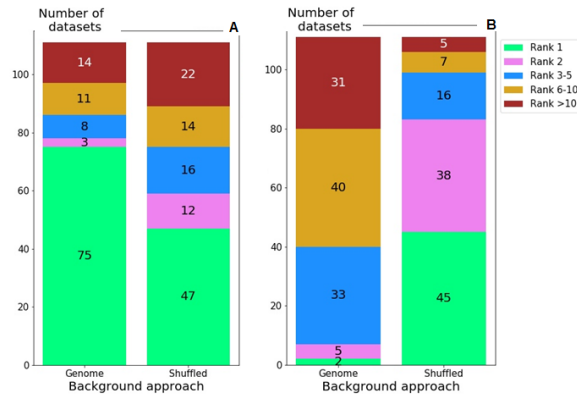
*Methods and Algorithms:* We used in analysis 111 ChIP-seq datasets from GTRD database [3]. We applied the STREME [4] tool for *de novo* motif search. To generate background data for this search we used two approaches, each of them presumes the same nucleotide content and lengths as those from peaks of foreground data. The first approach ‘Shuffled’ is the standard shuffling of nucleotides, i.e. the background dataset includes artificial sequences, which were generated according to the Markov chain of 0<sup>th</sup> order. The second approach ‘Genomic’ compiles randomly chosen genomic sequences [5]. We considered in analysis ten top ranked enriched motifs for each ChIP-seq dataset. The motif enrichment we estimated with the Fisher’s exact test comparing fractions of recognized sequences in the foreground and background datasets (the same recognition threshold for all motifs were applied, see [6], the recognition rate  $2E-4$  for the whole genome set of promoter respected this threshold). The TOMTOM tool [7] we used to prove the similarity between motifs. In particular (a) between results of *de novo* motif search and known motifs of target TFs and (b) between results of *de novo* motif search and the simple repeats (10 (XY)<sub>4</sub> and 32 (XYZ)<sub>3</sub> motifs). Here 10/32 imply the number of distinct possible di-/tri-nucleotides with respect to DNA strand.

*Results:* We compiled top ten enriched motifs from *de novo* motif search [4] for all ChIP-seq datasets and compared these motifs with (a) known motifs of target TFs and (b) motifs of simple repeats [7]. Analysis of the example GSM2898772 dataset for TF AGL8 (Fig. 1) shows that the background ‘Genomic’ has the better efficiency in the search for target motifs than the ‘Shuffled’ one. We may conclude that the first ranked motifs are the motif of target TF and polyA for the ‘Genomic’ and ‘Shuffled’ background datasets respectively. The distribution of ranks of ChIP-seq target motifs for all 111 ChIP-seq datasets confirms this conclusion (Fig. 2, A). Thus, among all 111 datasets in 75 cases (68 %) the target motifs have the first rank, in 97 (87 %) cases ranks are from one to ten. For the background ‘Shuffled’, 47 (42 %) target motifs possess the first rank and 89 (80 %) have ranks from one

to ten ranks. Figure 2, B shows respective distributions for motifs of simple repeats. Here we observe the opposite trend to that for motifs of target TFs.



**Fig. 1.** The nucleotide content bias in motif representation analysis on the example ChIP-Seq dataset for TF AGL8 (GSM2898772). Axes X imply A/T content of motif. Axes Y imply the motif enrichment  $-\text{Log}_{10}[\text{pvalue}]$ . Panels A and B shows the results of de novo motif search for the background datasets ‘Genome’ and ‘Shuffled’



**Fig. 2.** The distribution of ranks of enriched motifs in the results of de novo motif search for all 111 ChIP-seq datasets. Panels A and B depict motifs of target TFs and simple repeats. Axes X shows the background dataset approaches ‘Genomic’ and ‘Shuffled’. Axes Y marks the number of ChIP-seq datasets respecting detection of motifs of certain ranks

Hence, application of the genomic background dataset avoids in the results of de novo motif search unspecific genome sequence bias motifs such as polyA. Overall, the approach for genomic background is more efficient than the approach, of nucleotide shuffling.

*Acknowledgements:* The study was supported by Russian Science Foundation project 21-14-00240.

## References

1. Khan A. et al. BiasAway: command-line and web server to generate nucleotide composition-matched DNA background sequences. *Bioinformatics*. 2021;37(11):1607-1609.
2. Worsley Hunt R. et al. Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC Genomics*. 2014;15(1):472.
3. Kolmykov S. et al. *Nucl Acids Res*. 2021;49(D1):D104-D111.
4. Bailey T.L. *Bioinformatics*. 2021;37:2834-2840.
5. The software package SiteGA for de novo motif search in ChIP-seq data (<https://github.com/parthiansterlet/sitega>).
6. Levitsky V. et al. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res*. 2019;47:e139.
7. Gupta S. et al. Quantifying similarity between motifs. *Genom Biol*. 2007;8(2):R24.

# New insights on genome organization, gene expression regulation and evolution of the anhydrobiotic midge *Polypedilum vanderplanki*

Shagimardanova E.<sup>1\*</sup>, Kozlova O.<sup>1</sup>, Nesmelov A.<sup>1</sup>, Cherkasov A.<sup>1</sup>, Deviatiiarov R.<sup>1</sup>, Shaikhutdinov N.<sup>1</sup>, Gazizova G.<sup>1</sup>, Ryabova A.V.<sup>1</sup>, Cornette R.<sup>2</sup>, Kikawada T.<sup>2</sup>, Gusev O.<sup>1,3</sup>

<sup>1</sup>Extreme biology laboratory, Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia

<sup>2</sup>Institute of Agrobiological Sciences, National Agriculture and Food Research Organization (NARO), Tsukuba, Ibaraki, Japan

<sup>3</sup>Department of Regulatory Transcriptomics for Medical Genetic Diagnostics, Graduate School of Medicine, Juntendo University, Tokyo, Japan

\* ryukula@gmail.com

**Key words:** anhydrobiosis, *Polypedilum vanderplanki*, stress response, genome assembly, gene expression

**Motivation and Aim:** Extreme environmental conditions, including physical (temperature, high radiation), chemical (salinity, toxins), and biological (infection agents) factors, are the major engine of evolutionary changes. Organisms adapted to tolerate conditions in which others would not survive are referred to as extremophiles. A remarkable example of an extremophile is the sleeping chironomid midge *Polypedilum vanderplanki*. The midge tolerates complete desiccation by entering an ametabolic state, anhydrobiosis. The phenomenon of anhydrobiosis in *P. vanderplanki* was first discovered in 1951 by Howard Hinton [1]. Further studies in the early 2000s revealed basic biochemical mechanisms underlying tolerance to complete desiccation [2–4]. The whole genome of *P. vanderplanki* was decoded in 2014 [5]. The comparative genome study with sister species *P. nubifer* sensitive to desiccation showed that the genome of *P. vanderplanki* contains multicopy clusters of genes contributing to the formation of molecular shield. Such clusters include genes encoding antioxidants, chaperone-like Lea proteins, DNA and protein repair molecules, trehalose synthase and trehalose transporter. *P. vanderplanki* and recently discovered *P. pembae* are the only known species of chironomids capable of inducing anhydrobiosis, suggesting the recent origin and rapid adaptive evolution of this extreme adaptation in insects. **Methods and Algorithms:** Libraries of Illumina short pair-end reads, PacBio long reads, and Hi-C reads were used for genome assembly RNA-seq, Cage-seq, and CTR-seq for gene prediction and differential gene expression, as well as for promoter identification. Metabolites were measured using flight mass spectrometry (CE-TOFMS; Agilent Technologies). Single cell libraries were obtained using the Chromium platform (10xGenomics). For intraspecies comparative study genomes of other organisms were obtained from ENSEMBL, NCBI, and other databases. **Results:** To further investigate genetic mechanisms of adaptations to complete desiccation, we obtained a high-quality chromosome-level assembly of the *P. vanderplanki* genome. RNA-seq and CTR-seq data allowed the annotation of additional genes so that their number increased to 18990 compared to 17852 in the previous genome assembly. Anhydrobiosis-associated genes were found to preferably be localized in Chromosome IV, and most of the genes localized on this chromosome are highly specific for *P. vanderplanki* and do not have orthologs in other Dipteran insects [6]. Such genes have been extensively multiplied, forming specific loci, anhydrobiosis-related islands (ARIDs). One of the most extensively studied ARIDs is a

group of 27 Lea genes that encode intrinsically disordered proteins. It has been suggested that the evolution of Lea went toward subfunctionalization: The paralogs differ in both expression patterns in response to stress and cellular localization [7]. The Lea Island-Located (LIL) orphan genes form another important ARID. 14 Lil protein-coding genes were shown to be upregulated during desiccation. Lils do not possess any known protein domains, so their function is not fully clear. Lils are localized in the membrane, which may suggest their role in the transport of molecules during water loss [8]. Both the Lea and Lil genes are highly up-regulated during desiccation. We showed that the transcription of these genes is regulated by the heat shock factor (Hsf). Hsf controls many other protective genes related to anhydrobiosis, as they harbor Hsf binding sites in their promoter regions. Thus, Hsf is the main regulator for controlling and coordinating drastic transcriptional changes during anhydrobiosis [9]. The main hypothesis of multiple gene duplications in the *P. vanderplanki* genome is the specialization of the paralogs. To verify this hypothesis, we performed a CAGE-seq transcriptome analysis of several tissues of larvae during anhydrobiosis. This method allows to estimate both gene expression and reveal promoter switches suggesting gene regulation mechanisms. We found that some protective genes, including Lea, are expressed in a tissue-specific manner under desiccation. Most of these genes are located on Chromosome IV. Chromosome IV has a lower GC ratio and highly biased nucleotide composition and accumulation of variants compared to other chromosomes, suggesting an increase in the genetic diversity of Chromosome IV, possibly caused by relaxed selective pressure. Furthermore, the synteny block assay detected the absence of collinear blocks in Chromosome IV with any other species of dipteran insect except *P. pembae*. On the other hand, genes located in Chromosome IV of *P. vanderplanki* have some orthologous genes in other dipteran genomes. Interestingly, seven out of nine ARIDs are located in Chromosome IV. We further performed metabolome profiling of the larvae during the dehydration-rehydration cycle to find that, in addition to the well-known protective role, trehalose acts as a major source of energy for the initial steps of rehydration in the larvae. Citrate and adenosine monophosphate (AMP), accumulated in the dry state, allow rapid resumption of metabolism during the recovery phase [10]. Finally, we ask if all protective molecules provide protection to all cell populations in the larvae body or if there are specific cell types that survive and then proliferate during recovery. We conducted a pilot single-cell gene expression profiling of the brain of the larvae during the desiccation-rehydration cycle. Preliminary data showed strong fluctuation associated with anhydrobiosis of certain cell types. We hypothesized that some highly differentiated neural cells do not survive water loss, but stem cells proliferate during rehydration. This suggestion still needs to be validated by further studies. **Conclusions:** Using the multi-omics approach, we made several important steps in the understanding of molecular mechanisms of adaptations to extreme desiccation. Chromosome-level genome assembly allowed us to obtain evidence of the specific role of a single chromosome in the evolution of anhydrobiosis

*Acknowledgements:* This study was supported by Russian Science Foundation Grant No. 20-44-07002.

## References

1. Hinton H.E. *Proc Zool Soc Lond.* 1951;371-380
2. Watanabe M. et al. *J Exp Biol.* 2002;205:2799-2802.
3. Kikawada T. et al. *Integr Comp Biol.* 2005;45(5):710-714.
4. Sakurai M. et al. *Proc Natl Acad Sci U S A.* 2008;105(13):5093-5098.
5. Gusev O. et al. *Nat Commun.* 2014;12(5):4784.
6. Yoshida Y. et al. *NAR Genom Bioinform.* 2022;4(2):lqac029.
7. Kondratyeva S.A. et al. *Biology.* 2022;11(4):487.
8. Voronina T.A. et al. *Sci Rep.* 2020;10(1):11633.
9. Mazin P.V. et al. *Proc Natl Acad Sci U S A.* 2018;115:2477-2486.
10. Ryabova A. et al. *Proc Natl Acad Sci U S A.* 2020;117:19209-19220.

## Heterogeneity of transcription factor binding sites within ChIP-Seq datasets

Sharipov R.<sup>1,2,3\*</sup>, Kondrakhin Y.<sup>2,4</sup>, Kolmykov S.<sup>1,2</sup>, Yevshin I.<sup>1,2</sup>, Ryabova A.<sup>1,2</sup>, Kolpakov F.<sup>1,2,4</sup>

<sup>1</sup> *Sirius University of Science and Technology, Sochi, Russia*

<sup>2</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup> *Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia*

\* *shruss79@gmail.com*

**Key words:** transcription factor binding sites, ChIP-Seq, heterogeneity of binding sites

*Motivation and Aim:* Despite wide and useful application of the powerful ChIP-Seq technology for experimental identification of transcription factor (TF) binding sites, the computational prediction of site motifs is also relevant. There are many distinct methods and approaches to predict site motifs, however position weight matrix (PWM) approach remains to be the most common and widely used. Several well-known repositories contain PWMs for many distinct TFs and species, in particular HOCOMOCO, JASPAR and UniPROBE databases are among them.

In general, the main disadvantages of PWM methods are the ignorance of possible dependencies between nucleotides at different site positions as well as ignorance of information about nucleotide content at both flanks of site cores. To avoid these disadvantages, we proposed novel method for site prediction.

*Methods and Algorithms:* The novel method works as follows:

- at first, it assigns individual probability scores (IPSs) to each motif in order to take into account nucleotide content at both site flanks;
- at second, several PWMs are simultaneously used for prediction of binding sites for a single TF. In other words, the existence of such matrices indicates heterogeneity of total set of binding sites. Thus, the matrices represent distinct homogeneous subsets of total heterogeneous set of binding sites. These matrices were derived beforehand by application of the special IPS-alignment method to the training sample of most reliable binding regions.

The key point of alignment method is successive application of two-component normal mixture method to training sequence sample. The training sample, in turn, was composed by meta-analysis of ChIP-Seq datasets available in a GTRD database. It consists of sequences that are the most reliable binding regions. The reliability of training sequences was measured using special scores obtained by three-steps meta-method based on Rank Aggregation approach.

*Results:* Currently, using the developed method, we have composed three libraries of TF matrices for *Drosophila melanogaster*, *Danio rerio* and *Arabidopsis thaliana*. The corresponding site motifs are available in the GTRD database (<http://gtrd.biouml.org/>).

*Acknowledgements:* The study was supported by the Ministry of Science and Higher Education of the Russian Federation, grant 075-15-2021-1344.

## Capsid modifications of adeno-associated virus serotype 2

Shitik E.M.\*, Shalik I.K., Yudkin D.V.

*State Research Center of Virology and Biotechnology “Vector”, Rospotrebnadzor, World-Class Genomic Research Center for Biological Safety and Technological Independence, Federal Scientific and Technical Program on the Development of Genetic Technologies, Koltsovo, Novosibirsk, Russia*

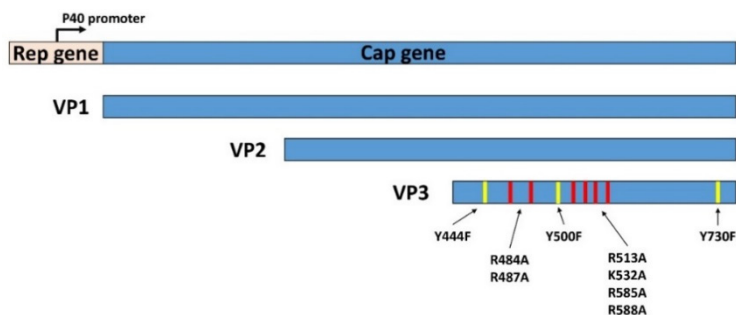
\* [shitik.ekaterina@gmail.com](mailto:shitik.ekaterina@gmail.com)

**Key words:** adeno-associated virus, VP3 protein, VP2 protein, viral tropism

*Motivation and Aim:* Adeno-associated virus (AAV) is a non-enveloped virus of size 20–25 nm belonging to the parvovirus family that becomes a leading platform for the delivery of transgenes *in vivo*. In 2022 263 clinical trials of drugs based on recombinant AAV (rAAV) are being carried out, 32 of which have reached phase III [1]. rAAV is not a pathogen for human cells and does not have the ability to maintain its own life cycle in the absence of helper viruses – adenovirus or herpesvirus. At the same time, rAAV entry does not depend on the cell division cycle and the virus maintains the expression of the transgene for a long time in the form of an episome. At the beginning of the virus study, the nonspecific affinity of rAAV for different target cells was also highlighted as an additional advantage because of the possibility to develop vectors for the treatment of various diseases. However, with the beginning of clinical trials, it became clear that the ability of the virus to enter all tissues of the body significantly limits the possibility of its systemic administration. Hence, the aim of the study is to develop a versatile vector based on AAV serotype 2 that can be used to create a highly tropic delivery platform to target cells.

*Methods and Algorithms:* A capsid of adeno-associated virus is composed of three main proteins VP1, VP2 and VP3 that are translated through splicing from one mRNA (Fig. 1). The VP3 sequence is a common part of all of three proteins that determines the tropism of the virus [2]. Moreover, it determines the transduction efficiency because of tyrosine residues that can be phosphorylated in the cell cytoplasm [3]. The VP2 protein presented in a small amount has a free N-terminus in the region of which rather long sequences of amino acids can potentially be introduced without affecting the capsid assembly [4].

In this study the DNA sequence of VP3 protein of AAV serotype 2 in pDp2 plasmid [5] has been modified. To increase the transduction efficiency of AAV2 three codon substitution Y444F, Y500F and Y730F were made with overlap extension PCR. For all DNA regions to be changed it has been chosen a pair of forward and reverse primers complemented to one region. To eliminate natural viral tropism six codon substitution were also made with overlap extension PCR. Codons 484, 487, 532, 585 and 588 were changed to codons of aliphatic amino acid residues to eliminate the viral tropism to heparan sulfate, and codon 513 – to eliminate the tropism to  $\alpha 5\beta 1$  integrin. Furthermore, the VP2 protein sequence has been cloned to pHMGFP plasmid (E642A, Promega) to express separately from other capsid proteins. In order to avoid expression from original pDp2 plasmid the start-codon of VP2 was destroyed without alteration of other proteins sequences also using overlap extension PCR.



**Fig. 1.** Scheme of the Cap gene encoding all proteins of the AAV capsid. Proteins VP1, VP2 and VP3 are transcribed from p40 promoter. Yellow markers indicate codons to be changed to increase the efficiency of virus transduction. Red markers indicate codons to be changed to destroy viral natural tropism

To receive viral particles cell line HEK293A was transfected with the plasmids obtained. The titer of viral particles was analyzed using real-time PCR. All of viral particles contained sequence of EGFP protein as a transgene. An efficiency of transduction was analyzed with fluorescence microscopy on HeLa cell line and on a line of fibroblasts that had been immortalized in our laboratory.

*Results:* Three types of modified rAAV serotype 2 have been obtained. One type contains replacement tyrosine residues with phenylalanine. With fluorescence microscopy it has been shown more effective transduction of this type in comparison with wild type of rAAV2. The second type in addition to previous modification has been obtained using plasmid that contains the sequence of VP2 protein regulated with its own promoter. With real-time PCR and fluorescence microscopy it has been shown that this modification does not affect the assembly of the viral particles and can be used to introduce proteins of interest. Apart from two previous modifications, the third type contains substitutions R585A, R588A, R484A, R487A, K532A and R513A. An almost absence of fluorescence on cell lines compared to the second modified type has shown that these substitutions determine AAV serotype 2 tropism.

Further studies on the introduction of proteins of interest into the N-terminus of VP2 protein to increase the tropism of viral particles will allow us to evaluate the effectiveness and versatility of the platform.

*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2019-1665).

## References

1. Gene Therapy Clinical Trials Worldwide. Wiley. <https://a873679.fmphost.com/fmi/webd/GTCT>.
2. Sha S. et al. Cellular pathways of recombinant adeno-associated virus production for gene therapy. *Biotechnol Adv.* 2021;49:107764.
3. Li et al. High-efficiency transduction of fibroblasts and mesenchymal stem cells by tyrosine-mutant AAV2 vectors for their potential use in cellular therapy. *Hum Gene Ther.* 2010;21(11):1527-1543.
4. Warrington et al. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. *J Virol.* 2004;78:6595-6609.
5. Grimm et al. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. *Mol Ther.* 2021;7:839-850.



## Potential bacteriophage recombination sites inside genes containing cysteine repeats

Sorokina N.<sup>1</sup>, Daugavet M.<sup>2</sup>

<sup>1</sup> Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia

<sup>2</sup> Laboratory of Noncoding DNA, Institute of Cytology, RAS, St. Petersburg, Russia

\* sorokina@scamt-itmo.ru

**Key words:** cysteine repeats, genes, horizontal transfer

*Motivation and Aim:* Prokaryotic-specific domains have been found in eukaryotic organisms frequently. This event is often associated with horizontal gene transfer and some studies have considered viruses as putative vectors for gene transfer. A previous study identified specific eukaryotic cysteine repeats adjacent to bacteriophage recombination sites (Atp) in the ascidian genome [1, 2]. A pattern of cysteine repeats was retrieved and used to create an experimental set of proteins to search for bacteriophage recombination sites in various eukaryotic genomes.

*Methods and Algorithms:* The work was done based on Python algorithms developed. An experimental set of proteins was created using the cysteine pattern  $C^*\{4,9\}C^*\{4,9\}C^*\{4,9\}C^*\{4,9\}CC$  for the study. All proteins matching the pattern in the Uniprot Trembl database were selected. Also, control dataset was generated: for each species from experimental dataset proteins without specific cysteine repeats were extracted. The second experimental set - sequences of bacteriophage recombination sites were collected earlier from the NCBI database. The resulting proteins were back-translated into nucleotide sequences using the Entrez database. Recombination sites in the nucleotide sequences of proteins with cysteine repeats without mismatches and indels were searched. Atp was also searched in the control dataset.

*Results:* Following a cysteine pattern search, 47,622 proteins were selected from 219,174,960 Uniprot Trembl proteins at the time the database was uploaded. Out of these, for 3,010 proteins nucleotide sequences of the genes were available. The experimental set of bacteriophage Atp recombination sites amounted to 186 sequences ranging from 10 to 491 nucleotides. By searching for Atp in the genes of proteins with cysteine repeats without mismatches and indels the following results were obtained. In 1,294 sequences of 3,010 (43 %) proteins 2,700 recombination sites were found, and in some protein genes several Atp were found. The length of the atp detected was between 10 and 16 nucleotides. Genes of proteins with cysteine repeats were obtained for 122 organisms. The control dataset consisted of one protein without a cysteine repeat for each species. Out of 122 genes of proteins, in 17 sequences (13,9 %) we found 28 short atp of length from 10 to 13 nucleotides.

*Conclusion:* Consequently, potential recombination sites of bacteriophages within genes containing cysteine repeats were found. Compared with the control group, it can be concluded that atp was more frequently found in the dataset containing cysteine repeats. Further, it is necessary to search for bacterial domains in eukaryotic genes containing recombination sites. It is reasonable to continue investigating the hypothesis of horizontal transfer associated with bacteriophages.

*Acknowledgements:* This work was supported by the Russian Science Foundation (grant No. 19–74-20102) and Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1075, signed 28.09.2021).

### *References*

1. Daugavet M.A., Shabelnikov S.V., Podgornaya O.I. Amino acid sequence associated with bacteriophage recombination site helps to reveal genes potentially acquired through horizontal gene transfer. *BMC Bioinformatics*. 2020;21(Suppl 12):305.
2. Daugavet M.A., Shabelnikov S., Shumeev A., Shaposhnikova T., Adonin L.S., Podgornaya O. Features of a novel protein, rusticalin, from the ascidian *Styela rustica* reveal ancestral horizontal gene transfer event. *Mobile DNA*. 2019;10:4.

## GESECA: Gene Set Co-regulation Analysis

Sukhov V.<sup>1</sup>, Artyomov M.<sup>2</sup>, Sergushichev A.<sup>1\*</sup>

<sup>1</sup> ITMO University, St. Petersburg, Russia

<sup>2</sup> Washington University in St. Louis, School of Medicine, St. Louis, MO, USA

\* [alserg@itmo.ru](mailto:alserg@itmo.ru)

**Key words:** pathway enrichment, gene signature, gene expression omnibus

Pathway enrichment is a standard approach to interpret transcriptional data. Commonly, in this approach, two conditions of interest are considered, for which differential gene expression is calculated, followed by one of the existing tools to find pathways that have coordinated positive or negative regulation of gene expression levels. However, currently the experiments contain many conditions and samples, ultimately, single-cell RNA-sequencing datasets contain thousands of individual cells measured without an inherent annotation for which cell corresponds to which cell type. For such datasets, classical pathway enrichment methods cannot be directly applied.

Here we propose GESECA (Gene Set Co-regulation Analysis): a method to find co-regulated gene sets in transcriptomic data without requirement for an explicit sample annotation. The method uses the percentage of variance explained by the queried gene set as an enrichment statistic: for gene sets with high pairwise correlation this value is high, compared to gene sets with uncorrelated genes. Similarly to the FGSEA approach, the statistical significance is assessed with a Multilevel Monte Carlo scheme, which allows empirical estimation of arbitrarily low P-values. Together, this establishes a method that can find co-regulated pathways in the data without explicit knowledge of the dataset annotation, and thus applicable both to bulk RNA-seq datasets with many conditions, as well as to scRNA-seq datasets.

Development of GESECA opens an opportunity to systematically re-analyze/re-utilize *all* of the published transcriptional profiling data. Specifically, we developed a search engine <https://ctlab.itmo.ru/geseca/> that allows to find gene expression datasets in which the query gene set is co-regulated. For this purpose we preprocessed a collection of more than 50000 gene expression datasets from Gene Expression Database (both RNA-seq and microarray). We have benchmarked the search engine against manually annotated signatures from the CRowd Extracted Expression of Different Signatures (CREEDS) web portal, which showed high dataset recovery rate. Furthermore, we have developed a keyword extraction algorithm for the search results, that allowed us to validate the search quality on the collection of MSigDB Hallmark gene sets.

To conclude, here we developed GESECA method to find co-regulated gene sets in transcriptomic data that provides (1) a quick and straightforward approach for annotation-free gene set enrichment, paving a path to (2) unbiased search of biologically relevant datasets based on a specific query gene set.

*Acknowledgements:* This research was supported by the Ministry of Science and Higher Education of the Russian Federation, Priority 2030 Federal Academic Leadership Program.

## **enRest tool for transcription factor binding sites overrepresentation analysis in RNA-seq data**

Tsukanov A.V.\*, Levitsky V.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [tsukanov@bionet.nsc.ru](mailto:tsukanov@bionet.nsc.ru)

Contemporary transcriptomic technology such as RNA-sequencing provides reliable approach to identify differentially expressed genes (DEG) across different cell types, tissues or conditions of stimuli/treatments. This standard analysis reveals many co-expressed genes or even gene networks that coordinately explain the observed biological responses. However, specific mechanisms of DEG regulation still too far from a full clarification. Hence, identifying the transcription factors responsible for observed changes in gene expression patterns is an important step in understanding gene regulatory networks. To determine the overrepresentation of transcription factor binding sites (TFBS) in regulatory regions of DEG we developed the tool *enRest* that took results of the RNA-seq experiment and the list of motifs (positional frequencies matrices, PFMs) from a public library, e.g. HOCOMOCO, CISBP and JASPAR. The tool scans 5' upstream regions (-2000; +1) of transcription start sites of all protein-coding genes of genome, hereafter called promoters. The tool divides promoters into the foreground and background sets. The former consists of promoters of DEG (RNA-seq experiment, adjusted  $p$ -value  $< 0.05$ ). The latter includes the same number of promoters randomly sampled from the rest genes of a genome. To estimate significance of the overrepresentation of potential TFBSs we applied the Monte Carlo approach. We repeated many sampling iterations of the background set to estimate the distributions of the number of sites per sequence in this set. Finally, we computed the enrichment  $p$ -value for each motif. The tool was tested on RNA-seq data of *H. sapiens* and *A. thaliana*. Our new tool can efficiently identify TFBSs that are statistically overrepresented in DEG.

*Acknowledgements:* The study was supported by Russian Science Foundation project No. 21-14-00240.

## FoxO1 may play a role in SREBP1 interaction with DNA

Vishnevsky O.<sup>1,2\*</sup>, Ignatieva E.<sup>1,2</sup>, Bocharnikov A.<sup>2</sup>, Kolchanov N.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *oleg@bionet.nsc.ru*

**Key words:** FoxO1, SREBP1, motif discovery, composite elements, transcription regulation, ChIP-Seq

*Motivation and Aim:* Sterol regulatory element-binding proteins (SREBPs) are a class of transcription factors that regulate lipid metabolism by controlling the synthesis of cholesterol, fatty acids, triglycerides, and phospholipids. The forkhead box-O1 (FoxO1) plays a key role in gluconeogenesis and glycogenolysis by insulin signaling, adipogenesis, cell cycle, proliferation, apoptosis, autophagy, stress resistance, DNA repair, tumor inhibition, and other cellular activities. Both transcription factors are insulin inducible and their mutual inhibition in Hepatoblastoma, leading to the increase of intracellular fatty acid metabolism ultimately facilitated the development of Hepatoblastoma. The aim of our study was to evaluate the mutual role of SREBP1 and FoxO1 in the regulation of mouse gene expression.

*Methods and Algorithms:* We have developed the Argo\_CEL system [this issue], which makes it possible to *de novo* identify significant mutually correlated pairs of IUPAC motifs corresponding to potential composite elements in a set of regulatory regions of genes. It was used to analyze the [−100; 100] sequences relative to the peak maxima of ChIP-Seq experiments with the transcription factor SREBP1 from [1]. The resulting motifs were annotated using the Tomtom system against the Jaspar and Hocomoco databases. Then, the correlation of the mutual presence of the most significant motifs corresponding to the SREBP1 (CSCCRCCS) and FoxO1 (TYRTTDWY) binding sites in the sequences of the ChIP-Seq peaks localized in the core promoter regions and in the intergenic regions was evaluated.

*Results:* From the results of the analysis presented in the table, it can be seen that for the CSCCRCCS (SREBP1) and TYRTTDWY (FoxO1) motifs, a high anti-correlation is observed in almost all orientations in the promoter region, which sharply decreases in intergenic regions. Based on this observation, it can be assumed that, in the absence of a strong motif for binding site of the transcription factor SREBP1, the interaction of this transcription factor with DNA is mediated by the presence of accompanying binding sites for partner transcription factors, such as FoxO1. Using the DAVID system, we performed a functional annotation of genes in the regulatory [−2000; 500] regions of which the ChIP-Seq SREBP1 peaks were localized, in which there were motifs corresponding to the FoxO1 binding sites, but there were no motifs corresponding to the SREBP1 binding sites. It turned out that such genes were significantly ( $p < 0.05$ ) characterized by GO terms “cell cycle” and “cell cycle process”.

**Table 1.** Changes in the correlations of the mutual arrangement of non-overlapping motifs found in ChIP-Seq peaks of SREBP1 in different regions of DNA relative to the closest start of transcription. The red arrows mark the CSCCRCCS motif in all orientations; the blue arrows mark the TYRTTDWY motif. Cells with significant correlations are marked with color ( $p < 0.05$ )

Mutual motifs orientation and arrangement		Regions of DNA relative to the closest start of transcription			
		-500 +500	-2000 +500	-5000 +500	-5000000 -50000
CSCCRCCS	TYRTTDWY	-0.0679311	-0.0788838	-0.0680081	-0.0301905
RWHAAYRA	SGGYGGSG	-0.0629877	-0.0621974	-0.0612617	-0.0285889
TYRTTDWY	CSCCRCCS	-0.0944169	-0.100894	-0.0878829	-0.0301905
SGGYGGSG	RWHAAYRA	-0.0477108	-0.0483002	-0.0496487	-0.0285889
CSCCRCCS	RWHAAYRA	-0.0823771	-0.0827146	-0.0761681	-0.000118922
TYRTTDWY	SGGYGGSG	-0.0713339	-0.0770609	-0.0687663	0.00486136
RWHAAYRA	CSCCRCCS	-0.0823771	-0.0827146	-0.0761681	-0.000118922
SGGYGGSG	TYRTTDWY	-0.0971916	-0.0993576	-0.0885172	-0.0320027

*Conclusion:* Our analysis showed that in the sequences of the ChIP-Seq peaks of experiments with antibodies to SREBP1, context motifs that best correspond to the binding sites of SREBP1 and FoxO1 significantly avoid coexistence. Furthermore, in cell cycle genes, the binding of SREBP1 to its binding site may be mediated by the presence of FoxO1.

*Acknowledgements:* The study is supported by the budget project of Institute of Cytology and Genetics, SB RAS (FWNR-2022-0020).

#### References

1. Gilardi F., Migliavacca E., Naldi A. et al. Genome-wide analysis of SREBP1 activity around the clock reveals its combined dependency on nutrient and circadian signals. *PLoS Genet.* 2014;10(3):e1004155.

## Methylation of p53-responsive microRNA genes in tumor tissue of lymphoma

Voropaeva E.<sup>1\*</sup>, Pospelova T.<sup>2</sup>, Berezina O.<sup>2</sup>, Churkina M.<sup>2</sup>, Gurazheva A.<sup>1</sup>, Maksimov V.<sup>1</sup>

<sup>1</sup>Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup>Novosibirsk State Medical University, Novosibirsk, Russia

\*vena.81@mail.ru

**Key words:** methylation, p53, microRNA, lymphoma

*Motivation and Aim:* The p53-mediated response to genotoxic cellular stress is largely carried out by activation of oncosuppressive and suppression of oncogenic microRNA [1, 2]. The aim of our study was to identify the frequency, specificity and clinical significance of *MIR-34B/C*, *MIR-34A*, *MIR-203* and *MIR-129-2* genes methylation in Diffuse Large B-cell Lymphoma (DLBCL).

*Material and Methods:* DNA was extracted from primary FFPE-samples and treated to convert unmethylated cytosine to uracil. Methylation status was analyzed by methyl-specific PCR and methyl-sensitive HRM in 73 tumor samples and 11 samples with reactive polyclonal B-cell proliferation. We evaluated the combined methylation of genes in pairs using the one-sided Fisher exact criterion ( $p$ -value) and multiple testing correction with Benjamin-Hochberg procedure ( $q$ -value).

*Results:* Methylation of *MIR-129-2*, *MIR-203*, *MIR-34A* and *MIR34B/C* in tumor samples occurred with a frequency of 67, 66, 27 and 62 % respectively. Methylation of the analyzed genes in the tissue of reactive lymph nodes was not detected. Combined methylation of *MIR-203*, *MIR-129-2* and *MIR-34B/C* genes ( $p = 0.013$ ,  $q = 0.020$ ), as well as pairs of *MIR-34B/C* and *MIR-34A* genes ( $p = 0.010$ ,  $q = 0.029$ ) has been shown. An assessment of the relationship between studied microRNA genes methylation and clinical and laboratory features of the disease showed that 18/20 (90 %) patients in the subgroup with the *MIR-34A* gene methylation had a high and intermediate/high risk according to International Prognostic Index against 26/53 (49.1 %,  $p = 0.002$ ) in the subgroup of patients without this gene methylation. Methylation of *MIR-34A* was associated with reduced frequency of remission ( $p = 0.060$ ) and decreased overall survival ( $p = 0.162$ ). Methylation of *MIR-34B/C* ( $p = 0.026$ ) and *MIR-203* ( $p = 0.011$ ) was associated with Ki-67 expression level of more than 45 %.

*Conclusions:* Tumor-specific methylation of gene promoters can serve as a significant mechanism for reducing the miR-34B/C, miR-34A, miR-203 and miR-129 expression in DLBCL. Aberrant methylation of oncosuppressive microRNA genes associated with underlying p53 signalling pathways is a potentially useful molecular biomarker in the lymphoma diagnosis. *MIR-34A* gene methylation is helpful in prognosis and targeted therapy strategy development of DLBCL.

*Funding:* This work was supported by grant of Russian Science Foundation No. 22-25-00222.

## mitomiRs as the common regulators of gene silencing

Vorozheykin P.<sup>1\*</sup>, Vishnevsky O.<sup>2</sup>, Titov I.<sup>1,2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* Pavel.Vorozheykin@gmail.com

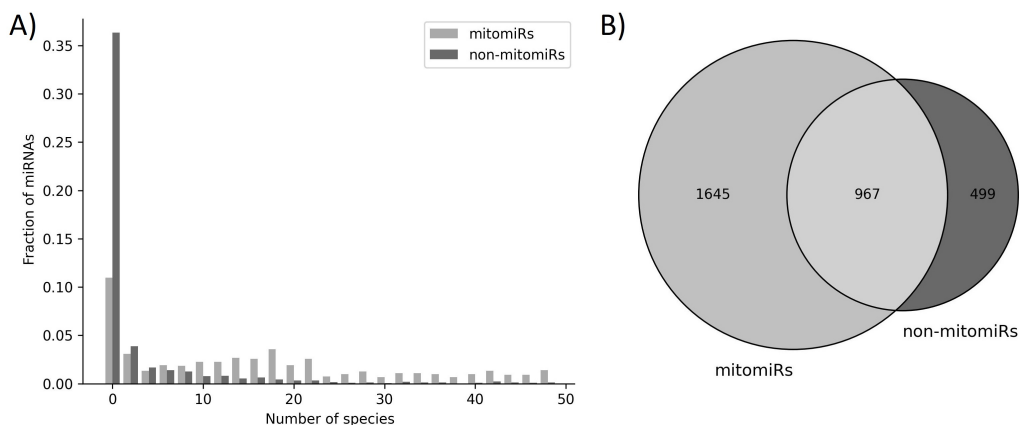
**Key words:** miRNA, mitomiR, mitochondria, evolution, targets

*Motivation and Aim:* To date it is well known that fundamental pathways of the miRNA biogenesis start in nucleus and end into cytoplasm. However, there is evidence that these short RNA sequences are also presented in organelles, for example in mitochondria. These observations may reveal both the nucleus miRNA translocation to the mitochondria and the existence of the miRNA maturation process within the mitochondria. The miRNA machinery proteins (Argonaute/AGO and Dicer) have been detected in mitochondria and the mitochondrial gene expression can be regulated by the miRNAs. The existence of such mitochondria associated miRNAs (the so-called mitomiRs) raises questions about their function and origin. It is still unknown, is there a specific functions of different miRNA classes according to their cell location? Are mitomiRs originally evolved or have appeared only recently? Is there a sequence or structure difference between the mitomiRs and non-mitomiRs that provides a mitomiR translocation to the mitochondria?

*Methods and Algorithms:* The dataset miRNA sequences were downloaded from the miRBase database (miRBase.org, release 22.1). We considered the experimental studies that demonstrate the presence of the miRNAs within mitochondria isolated from various organisms, cell types and tissue. According to these studies we construct two sets with human, mouse and rat miRNAs: mitomiRs (observed within mitochondria, 652 sequences) and non-mitomiRs (others, 4766 sequences). The homologous search was carried out by comparing Hamming distance of the aligned miRNAs and selecting those miRNA pairs that differ less than 5 % of the sequences length. To investigate the sequence and structure differences of the miRNA classes we searched for statistically overrepresented oligonucleotide motifs by ARGO-program [2]. To compare functions of the miRNAs, we explore the mitomiR and non-mitomiR targets from the experimentally validated miRNA-associated gene database miRTarBase (release 8.0) [3] and investigated the relationship of the miRNA targets with the gene ontology (GO) terms from the GeneOntology.org database (on the date 2022-03-2) [4].

*Results:* Comparing the miRNA homolog sequences from the miRBase we concluded that the mitomiRs are used to be more widely distributed among species (Fig. 1, A). This supports the hypothesis that mitomiRs may be recruited into mitochondria during their domestication. In addition to that, the mitomiRs are associated with a greater number of target genes than the non-mitomiRs (see Fig. 1, B), even though the less mitomiR sequences were discovered.





**Fig. 1.** A) The fraction of the miRNA sequences corresponding to the number of the species in which miRNAs have homologs; B) The number of the target genes for the mitomiRs and non-mitomiRs

Comparing the sequences of mitomiRs and non-mitomiRs we discovered several statistically overrepresented oligonucleotide motifs in mitomiRs: CAKTSHAN, KTGCANDK, HASHWSBD and others. These motifs may be signals for a specific miRNA processing (for example, a mitomiR translocation) or may directly affect the gene regulation. To explore the functions of the considered miRNAs, we performed the GO enrichment analyses for the targets lists of human mitomiRs and non-mitomiRs in three general categories “biological process”, “molecular function”, and “cellular component”. About 889 GO terms are related only to the mitomiRs targets which indicates the unique features of mitomiRs. For the common GO terms of mitomiRs and non-mitomiRs targets we calculated the Fisher’s exact test ( $p$ -value  $\leq 0.05$ ) and received 290 GO terms which are highly associated with one list of genes comparing to the other. All discovered GO terms are specifically enriched in list of non-mitomiR target genes.

Therefore, the mitomiR sequences can be identified as the common regulators while the non-mitomiRs appear to be specific ones.

*Acknowledgements:* The work of OV and IT was supported by the budget project No. FWNR-2022-0020.

### References

1. Kozomara A., Birgaoanu M., Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1):D155-D162.
2. Vishnevsky O., Kolchanov N. ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters. *Nucleic Acids Res.* 2005;1(33):W417-W422.
3. Huang H.Y. et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020;48(D1):D148-D154.
4. Ashburner et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25-29.

## BacRegDB: a database of bacterial regulatory elements with structural evidence

Vychyk P.<sup>1\*</sup>, Duvalov E.<sup>2</sup>, Digris A.<sup>2</sup>, Skakun V.<sup>2</sup>, Nikolaichik Y.<sup>1</sup>

<sup>1</sup> Department of Molecular Biology, Belarusian State University, Minsk, Belarus

<sup>2</sup> Department of Systems Analysis and Computer Modelling, Belarusian State University, Minsk, Belarus

\* p.vychik@gmail.com

**Key words:** genome annotation, transcription factor binding sites

*Motivation and Aim:* None of the current genome annotation pipelines handles regulatory elements like promoters and operators. Although the information about bacterial regulatory elements is collected in several specialized databases, only part of it can be used to annotate regulatory elements, and only in a very limited range of bacteria. Most importantly, no reliable criteria is used for the transfer of regulatory information between genomes. Such transfer is usually based on thorough homology evaluation and heavily depends on human judgement.

Our aim in this work was to develop a new transcription factor binding site database permitting reliable automated transfer of regulatory information between bacterial genome sequences.

*Methods and Algorithms:* We have previously proposed a 3D structure based strict formal criterion for applying regulatory information to any bacterial genome [1]. This criterion is the CR-tag: the amino acid residues of a transcriptional regulator that specifically contact the nitrogenous bases of the regulatory element in genomic DNA. A motif associated with the CR tag can be correctly applied to annotate such elements in genomes encoding transcriptional regulators with identical CR tags.

The accumulation of motifs associated with CR tags demanded their ordered storage for the convenience of subsequent use in genome annotation. Since none of the known databases use the CR tag concept, a new relational database was developed and implemented in Microsoft SQL Server.

Original regulatory motif information was collected either from the literature or from the regulatory motif databases RegulonDB, CollecTF, Prodoric2, CoryneRegNet [2–5]. For each transcription factor present in these databases, our automated *de novo* transcription factor binding site inference pipeline [1] was run to collect transcription factor gene-linked operators from all genomes encoding TFs with the same CR tag. If the resulting *de novo* inferred motif matched the one present in the databases or described in the literature, automatically collected and experimentally characterised operators were combined in the final profile. Hidden Markov models for each motif were built with threshold cutoff scores manually set to find known operators with minimum or no false positive hits.

A fully automated import procedure was developed within the framework of our Sigmoid application [6] for RegPrecise, the largest bacterial regulatory database [7]. Little or no manual curation was performed for the resulting motifs at this time. Cutoff scores for RegPrecise-derived operator profiles were calculated based on manually curated data via regression curves of motif information content versus bit score thresholds.

**Results:** The database currently covers only transcription factor binding sites and has two divisions: the core and the extended collection. The core includes 237 regulatory elements and has undergone full manual curation including verification of experimental evidence and determination of threshold scores for HMM models of each element. Each core record has manually assigned experimental evidence codes and is linked to the corresponding literature. The extended collection includes information on over 3000 regulatory elements exported from various databases with majority of information coming from RegPrecise. We noted that RegPrecise regulons quite often combine regulons of TFs with different CR tags. On the other hand, many regulons of TFs with identical CR tags are present in RegPrecise as separate entries. Our RegPrecise import procedure takes special care to correctly split mixed regulons and combine regulons with the same CR tag, thus improving the quality of the resulting regulatory motifs. The full BacRegDB content is currently accessible via the standalone Sigmoid application (<https://github.com/nikolaichik/Sigmoid>) and can be used for annotating regulatory elements in bacterial genomes in semi-automated fashion. We are also completing the development of the web application interfacing BacRegDB and allowing regulatory element annotation in user-submitted genomes.

**Conclusion:** The main advantage of BacRegDB is the CR-tag concept – a fingerprint uniquely matching transcription factors with their operators. All regulatory motif records in the database are associated with a CR tag and, therefore, can be correctly used to annotate similar elements in any genomes encoding a transcriptional regulator with an identical CR tag.

## References

1. Nikolaichik Y., Vychik P. New approach to genome-wide automated inference of bacterial transcription factor binding sites. In: *Bioinformatics of Genome Regulation and Structure/Systems Biology*. 2020;75-76. DOI: 10.18699/BGRS/SB-2020-046.
2. Santos-Zavaleta A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* 2019;47(D1):D212-D220.
3. Kılıç S. et al. CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res.* 2013;42(D1):D156-D160.
4. Dudek C.-A., Jahn D. PRODORIC: state-of-the-art database of prokaryotic gene regulation. *Nucleic Acids Res.* 2022;50(D1):D295-D302.
5. Parise M.T.D. et al. CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Sci Data.* 2020;7(1):142.
6. Nikolaichik Y., Damienikan A.U. Sigmoid: a user-friendly tool for improving bacterial genome annotation through analysis of transcription control signals. *PeerJ.* 2016;4:e2056.
7. Novichkov P.S. et al. RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics.* 2013;14:745.

## Actin genes of the White Sea sponge *Halisarca dujardini*: structure and regulation

Zhurakovskaya A.<sup>1\*</sup>, Kravchuk O.<sup>2</sup>, Adameyko K.<sup>2</sup>, Lyupina Y.<sup>2</sup>

<sup>1</sup> Faculty of Biology, Department of Genetics, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> N.K. Koltzov Institute of Developmental Biology, RAS, Moscow, Russia

\* anne.zhurakovskaya@gmail.com

**Key words:** actin, beta-actin, reaggregation, sponges, *Halisarca dujardini*, invertebrates

**Motivation and Aim:** Sponges (Porifera) are the oldest living multicellular organisms. The body of sponges is formed by several different cell types that do not form tissues *sensu stricto*. Cell plasticity is one of the key features of sponges: their cells constantly perform movement, remodeling, and reorganization. Most of these cells have ability to transdifferentiate. Such plasticity of sponge cells is fully manifesting after dissociation of the sponge body, when cells are actively reaggregating in order to form multicellular aggregates and eventually rebuild the body.

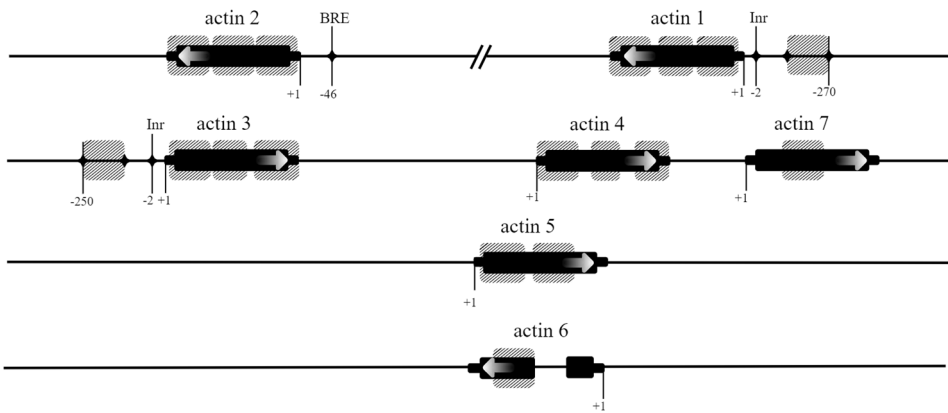
Actin is one of the main components of the cytoskeleton of eukaryotic cells. It creates mechanical support, participates in cell division, intracellular transport, and cell movement. The contribution of actin to the maintenance of plasticity of sponge cells, especially during reaggregation is undeniable. In the present work, we studied the structure of actin genes, the regulation of their expression and the functional features of protein products in the White Sea sponge *H. dujardini*.

**Methods and Algorithms:** Whole-genome sequencing of the *H. dujardini* was performed using Oxford Nanopore and Illumina. A draft assembly was made and contigs containing actin genes were identified and examined. Expression in different physiological states was studied using RNA-seq of intact sponges, cell suspension, and aggregates at the 24 h after reaggregation. 5'UTRs of actin genes were determined by RACE analysis. Protein products were separated in the native or SDS-containing polyacrylamide gels and stained with Coomassie blue. Mass spectra of tryptic peptides were obtained on the matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometer Ultraflexxtreme (Bruker), equipped with UV laser (Nd) and reflectron.

**Results:** Whole-genome study of the *H. dujardini* sponge revealed seven actin genes. Six out of seven genes lacked introns, some of the genes are likely formed as a result of a recent duplication. Phylogenetic analysis classified the actins of *H. dujardini* as beta-actins. Transcriptomic data confirmed the expression of products of all genes and mass spectrometry showed the presence of their protein products. Actins 1–5 have high sequence similarity to human beta-actin (Table 1). The most variable actin 7 lacks the sequence required for polymerization, which may indicate its special molecular function. All *H. dujardini* actins have binding motifs for profilin and gelsolin – the key protein regulators of actin filament assembly and disassembly. The analysis of the upstream genomic regions of the actin genes revealed the diversity of actin gene promoters in *H. dujardini*, which may indicate the cell type-specific mode of expression for these genes. This is indirectly confirmed by the changes in actin gene expression during reaggregation, when numerous events of dedifferentiation and transdifferentiation occur.

**Table 1.** GenBank accession numbers (ANs) for mRNA sequences of actins in *H. dujardinii* and accession numbers of their homologs / best blastp hits in *Homo sapiens* and *Amphimedon queenslandica*

Protein name	<i>H. dujardinii</i>	Homologs / best blastp hits			
		<i>H. sapiens</i>		<i>A. queenslandica</i>	
		AN	Identity, %	AN	Identity, %
actin 1	MT451954	<a href="#">NP_001092.1</a>	97.33	<a href="#">XP_019853827.1</a>	62.13
actin 2	MT451955	<a href="#">NP_001092.1</a>	97.33	<a href="#">XP_019853827.1</a>	62.13
actin 3	MT451956	<a href="#">NP_001092.1</a>	97.33	<a href="#">XP_019853827.1</a>	62.13
actin 4	MT451957	<a href="#">BAD96645.1</a>	97.06	<a href="#">XP_019853827.1</a>	61.60
actin 5	MT451958	<a href="#">NP_001092.1</a>	96.26	<a href="#">XP_019853827.1</a>	61.33
actin 6	MT451959	<a href="#">NP_005150.1</a>	73.60	<a href="#">XP_019853827.1</a>	51.06
actin 7	MT518195	<a href="#">7P1H_B</a>	88.20	<a href="#">XP_003384376.1</a>	59.26



**Fig. 1.** Actin genes of *H. dujardinii*. Arrows indicate the direction of the transcription; CpG islands are marked with grey shading; Inr, initiator of transcription; BRE, B recognition element. The representation of the regulatory elements and genes is not to scale

**Conclusion:** High level of identity between sponge and human actins indicates that the structure of these proteins was formed in great antiquity. Despite its simple organization, *H. dujardinii* has seven actin genes featuring a variety of regulatory elements. Five of the actins are conserved, while two other have distinct structural features. We hypothesize that the differential expression of actin genes contributes to the distinct features of sponge cell types.

**Acknowledgements:** We are grateful to the Polar Circle tourist center for the help in collecting the material.

## 1.3 Section “Functional and applied 3D genomics”



# Maps of chromatin conformation in the cerebral cortex

Cherkasov A.<sup>1</sup>, Bazarevich M.<sup>2</sup>, Pletenev I.<sup>1</sup>, Efimova O.<sup>1</sup>, Kononkova A.<sup>1,3</sup>, Khrameeva E.<sup>1\*</sup>

<sup>1</sup> Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> A.A. Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia

\* e.khrameeva@skoltech.ru

**Key words:** chromatin, brain, Hi-C, RNA-seq, regulation of gene expression

*Motivation and Aim:* Chromosome conformation capture (3C) methods have launched tremendous progress in our understanding of chromatin organization. At the higher end of chromatin organization, interphase chromosomes form distinct territories called chromosome territories; small and active chromosomes tend to locate at the center of the nucleus, whereas inactive and heterochromatic regions are located in the nuclear periphery [1–3]. Using the Hi-C technique, Lieberman-Aiden et al. demonstrated that chromatin is spatially segregated into two compartments [4]. Follow-up studies with higher data resolution revealed that individual chromosomes are partitioned into topologically associating domains (TADs), which are highly conserved across cell types and species. At the lower level of chromatin organization, long-range chromatin loops were identified within TADs from super-resolution 3C-based sequencing data [5, 6]. Despite this progress, Hi-C experiments in the brain tissue remain challenging, which is why, among hundreds of publicly available Hi-C datasets, only a few papers describe chromatin architecture in the human brain [7–12], two of these papers still being on bioRxiv. Even fewer studies analyze the adult brain, as the fetal brain is easier to work with. In this work, we aim to comprehensively describe chromatin architecture in the adult brain and link chromatin organization features with functional properties of the brain at the molecular level.

*Methods and Algorithms:* We used FACS flow cytometry to sort neurons (NeuN+ cells) and glia (NeuN- cells), two main types of cells in the brain, allowing to determine the chromosome conformation separately for each cell type. We then applied Hi-C method to produce genome-wide chromatin conformation maps separately in neurons and glial cells, according to the Hi-C protocol that we have recently optimized for the mouse brain tissue [13]. Each Hi-C library was sequenced to 150–200 mln. paired-end reads per replicate and the resulting Hi-C data were binned at 50, 25, and 20-kb resolution to choose the most suitable Hi-C map resolution for each level of chromatin architecture. We also performed RNA-seq experiments in the same samples.

*Results:* In agreement with current understanding of chromatin organization in mammals, our Hi-C interaction maps of cerebral cortex demonstrate multiple architectural levels: interchromosomal contacts, A/B compartments, TADs, and point-scale interactions. Further, we comprehensively analyze chromatin organization at all levels of chromosome architecture, starting from interchromosomal interactions and finishing by chromatin loops. In addition, we analyze gene expression data accessed using RNA-seq experiments in the same cortex area to obtain integrated information about transcription and chromatin architecture, and subsequently suggest possible

mechanisms linking chromatin conformation features with molecular functioning of the brain.

*Conclusion:* Our new data on the 3D organization of chromatin in the brain allows filling one of the few remaining gaps in an overall picture of the molecular organization of the brain.

*Acknowledgements:* The study is supported by the RSF grant 21-74-10102. This is a joint study with Dr. Sergey Ulianov (Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia) and Dr. Nikolay Kondratyev (Federal State Budgetary Scientific Institution the Mental Health Research Center, Moscow, Russia).

## References

1. Croft J.A., Bridger J.M., Boyle S., Perry P., Teague P., Bickmore W.A. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol.* 1999;145:1119-1131.
2. Tanabe H., Müller S., Neusser M., von Hase J., Calcagno E., Cremer M., Solovei I., Cremer C., Cremer T. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci USA.* 2002;99:4424-4429.
3. Bolzer A., Kreth G., Solovei I., Koehler D., Saracoglu K., Fauth C., Müller S., Eils R., Cremer C., Speicher M.R. et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 2005;3:e157.
4. Lieberman-Aiden E., Van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326:289-293.
5. Rao S.S.P., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014;159:1665-1680.
6. Tang Z., Luo O.J., Li X., Zheng M., Zhu J.J., Szalaj P., Trzaskoma P., Magalska A., Włodarczyk J., Ruszczycki B. et al. CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell.* 2015;163:1611-1627.
7. Won H., de la Torre-Ubieta L., Stein J.L., Parikshak N.N., Huang J., Opland C.K. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature.* 2016;538(7626):523-527.
8. Wang D., Liu S., Warrell J., Won H., Shi X., Navarro F.C.P. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science.* 2018;362(6420):eaat8464.
9. Espeso-Gil S., Halene T., Bendl J., Kassim B., Ben Hutta G., Iskhakova M. et al. A chromosomal connectome for psychiatric and metabolic risk variants in adult dopaminergic neurons. *Genome Med.* 2020;12(1):1-19.
10. Lu L., Liu X., Huang W.K., Giusti-Rodríguez P., Cui J., Zhang S. et al. Robust Hi-C chromatin loop maps in human neurogenesis and brain tissues at high-resolution. *bioRxiv.* 2019;744540. doi: 10.1101/744540.
11. Giusti-Rodríguez P., Lu L., Yang Y., Crowley C.A., Liu X., Bryois J. et al. Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex. *bioRxiv.* 2018;406330. doi: 10.1101/406330.
12. Hu B., Won H., Mah W. et al. Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat Commun.* 2021;12(1):3968.
13. Eremenko E., Golova A., Stein D., Einav M., Khrameeva E., Toiber D. FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C. *STAR Protoc.* 2021;2(3):100643.



# Comparative bioinformatic analysis of Hi-C chromosome-level scaffolds and chromosome painting maps of *Vulpes vulpes* and *Canis familiaris* (Canidae, Carnivora, Mammalia)

Fofanov M.V.<sup>1\*</sup>, Prokopov D.Yu.<sup>1</sup>, Trifonov V.A.<sup>1</sup>, Beklemisheva V.R.<sup>1</sup>, Khan R.<sup>2</sup>, Aiden E.<sup>2</sup>, Graphodatsky A.S.<sup>1</sup>, Perelman P.L.<sup>1\*</sup>

<sup>1</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>2</sup> Baylor College of Medicine, Houston, TX, USA

\* fofanov@mcb.nsc.ru, polina.perelman@gmail.com

**Key words:** canids, rearrangements, chromosome evolution, genome assembly

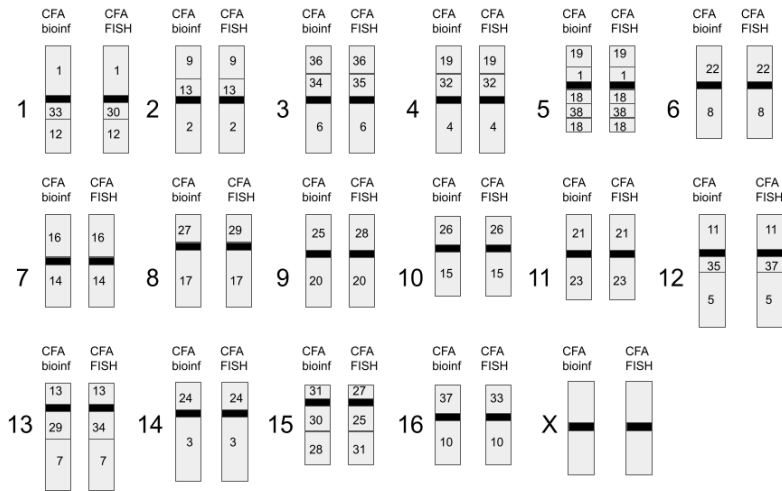
**Motivation and Aim:** For the past decades the main source of information about the chromosome-level rearrangements underlying speciation was cross-species chromosome painting. Efficient comparison of chromosomes between species yielded a large bolus of data on chromosome evolution trends and rates for all mammalian orders. Recently, chromosome level genome assemblies became available for a variety of animal species, with the potential to broaden such comparisons including detecting “smaller” chromosomal events, identifying precise breakpoint coordinates, involving species that do not have painting data. There is a need for a translation pipeline so that the data coming from both methods can be used in a mutually beneficial way. Here we attempt to set up the pipeline to evaluate cross-species evolutionary chromosome rearrangements based on chromosome level assemblies.

We chose the family Canidae (wolves, foxes, raccoon dogs etc.) that is characterized by remarkably high rates of evolutionary chromosomal rearrangements compared to other mammalian groups. Among canids multiple breeds of domestic dog (*Canis familiaris*) now have their genomes assembled to chromosome level [1], and red fox (*Vulpes vulpes*) has a pre-publication of chromosome-length genome assembly has become recently available on DNA Zoo (www.dnazoo.org) [2, 3].

**Methods and Algorithms:** *Vulpes vulpes* (VulVul2.2\_HiC) and *Canis familiaris*, golden retriever breed (canFamDis\_HiC), assemblies were retrieved from the DNA Zoo website (<https://www.dnazoo.org/assemblies>). These genome assemblies were soft-masked using RepeatMasker [4] with the following options: -engine hmmer -s -nolow -norna -species canidae. Then chromosome-length scaffolds from these assemblies (17 scaffolds for fox and 39 for dog) were aligned using the LASTZ [5] with several non-default options (--format=maf --notransition --nogapped --seed=match15 --step=10 --hspthresh=100000 --ambiguous=iupac --allocate:traceback=200M). Next, chromosome-level alignments were visualized as dot-plots via D-GENIES [6]. Finally, we compared dot-plots and chromosome maps from the Atlas of Mammalian Chromosomes [7] and visualized them manually in one image.

**Results:** Chromosome-length scaffolds (17 comprising 96.54% of the fox genome assembly and 39 (89.51% of the dog genome assembly) were selected for the LASTZ. Alignments were visualized as dot-plots and used to build chromosome maps (Fig. 1). All large genomic rearrangements observed during our bioinformatic analysis were previously detected using FISH with chromosome-specific probes. There were some

small genomic rearrangements involving genome fragments ranging from 1100 to 1600 bp which could represent an artifact of genome data preprocessing. On comparative dot-plots we observed a large number of gaps that could be caused by the use of only chromosome-length scaffolds, evolutionary distance between two species (9–10 million years), and soft-masking of genome assemblies prior to the analysis.



**Fig. 1.** Comparison of dog-fox chromosome maps derived from Hi-C scaffold alignments (bioinf (*left*)) and reciprocal chromosome painting (FISH (*right*)). Fox chromosome numbers are indicated on the left; corresponding dog (CFA) chromosome segments are indicated inside of chromosome outlines (note painting CFA nomenclature correspondence with the dog reference karyotype is in [8])

**Conclusion:** The approach of cross-species comparison of chromosome-level scaffolds can be reliably applied to species that do not have chromosome painting maps for the identification of the evolutionary chromosome rearrangements because here we show that all rearrangements previously detected by chromosome painting were also identified in the pairwise analysis of Hi-C scaffolds.

**Acknowledgements:** This study is supported by RSF grant No. 22-24-01076.

### References

1. Field M.A. et al. The Australian dingo is an early offshoot of modern breed dogs. *Sci Adv.* 2022;8(16).
2. Dudchenko O. et al. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv.* 254797.
3. Kukekova A.V. et al. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat Ecol Evol.* 2018;2(9):1479-1491.
4. Tempel S. Using and Understanding RepeatMasker. *Methods Mol Biol.* 2012;859:29-51.
5. Schwartz S. et al. Human–Mouse Alignments with BLASTZ. *Genome Res.* 2003;13(1):103-107.
6. Cabanettes F., Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 2018;6:e4958.
7. Graphodatsky A.S., Perelman P.L., O’Brien S.J. Atlas of Mammalian Chromosomes. Wiley, 2020.
8. Breen M Canine cytogenetics – from band to basepair. *Cytogenet Gen Res.* 2008;120(1-2):50-60.

# Influence of TADs boundary disruption on 3D genome organisation and genes expression at the mouse *Pdgfra/Kit/Kdr* locus

Kabirova E.<sup>1,2\*</sup>, Lukyanchikova V.<sup>1,2</sup>, Ryzhkova A.<sup>1</sup>, Khabarova A.<sup>1</sup>, Battulin N.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *kabevelyn@gmail.com*

**Key words:** development genes, topologically associated domains, 3D genome

**Motivation and Aim:** It is generally accepted that topologically associated domains (TADs) participate in genes expression regulation by bringing enhancer and promoter into proximity and maintaining their interaction. Whether TADs restrict ectopic enhancer interactions, thus exhibiting an insulation function, remains unclear. To explore the question we studied the role of 3D genome organization in the regulation of the *Pdgfra/Kit/Kdr* locus. The locus encodes tyrosine-kinase receptors, which are involved in cell differentiation processes. The locus consists of three TADs, so each gene is located within a corresponding TAD. Of note, the genes possess a distinct contrasting tissue-specific expression pattern: *Pdgfra* — fibroblasts; *Kit* — melanocytes, mast cells; *Kdr* — endotheliocytes (*Kdr* is also expressed in melanocytes, albeit weakly). We hypothesized that removal of TADs boundaries would provide ectopic enhancer interactions leading to aberrant genes expression in an unfamiliar cell type.

**Methods and Algorithms:** The mouse strains, carrying CRISPR/Cas9-mediated deletions of the key boundary CTCF-binding sites, and C57BL/6 mice as a control were used to obtain cell cultures (fibroblasts, mast cells, melanocytes). In order to analyse 3D genome organisation of the locus in the cell cultures we performed capture Hi-C and CTCF ChIP-seq, as well as H3K27Ac ChIP-seq to distinguish the enhancer pattern. To analyse a possible transcriptional change we obtained RNA from the cell cultures and proceeded with RNA-seq.

**Results:** For the boundary encompassing *Pdgfra* and *Kit* TADs, the deletion was not enough to disrupt the TADs insulation, as well as to activate *Kit* expression in fibroblasts. For the boundary encompassing *Kit* and *Kdr* TADs, the deletion revealed a tissue-specific influence. In mast cells the deletion led to a gain of inter-TADs contacts, but the *Kit* remained insulated from *Kdr* and no *Kdr* activation was detected. Conversely, in melanocytes the deletion did lead to TADs fusion and, most interestingly, to a noticeable activation of *Kdr* expression. There are at least two factors for such a difference. Firstly, unlike the mast cells, the melanocytes do express *Kdr* slightly. Secondly, in the mast cells *Kit* enhancers locate upstream of the *Kit*, while in the melanocytes — downstream, closer to *Kdr*. We presume that an epigenetic status of the locus and an enhancer location might contribute to TADs fusion restriction. In addition, the *Kit/Kdr* mutant mice demonstrated a phenotypic change, mutant mice being brown compared to the black control mice. The change putatively implies an extent of the *Kit/Kdr* TADs boundary evolutionary conservatism.

*Conclusion:* Considering these findings, we suggest that the TADs boundaries deletions are insufficient to change the regulatory pattern, unless the TADs fusion is allowed by other factors at least in the mouse *Pdgfra/Kit/Kdr* locus. Thus, the result provides an insight in a more complex TADs role requiring a comprehensive investigation of the factors participating in TADs formation.

*Acknowledgements:* The Russian Foundation for Basic Research supported the study (No. 18-29-07022).

# Insights into the 3D-genome organization in malaria mosquitoes

Lukyanchikova V.<sup>1,2,3,4\*</sup>, Nuriddinov M.<sup>3,4</sup>, Belokopytova P.<sup>3,4</sup>, Sharakhov I.<sup>1,2</sup>, Fishman V.<sup>3,4</sup>

<sup>1</sup> Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, USA

<sup>2</sup> Fralin Life Science Institute, Virginia Polytechnic Institute and State University, Blacksburg, USA

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

\* varvara\_lukyanchikova@mail.ru

**Key words:** 3D genome organization, long-distance chromatin loops, Hi-C technology, mosquito genomics, chromatin, malaria mosquitoes

Anopheles mosquitoes are the only carriers of malaria disease across the world. Their quick spreading all over the globe and devastating effect on human health raises the dramatic necessity to investigate their genomes in details, especially knowing that gene expression and regulation is tightly linked with 3D chromatin interactions. Comparative studies performed in vertebrates species revealed that genome architecture is evolutionarily conserved and could be explained by the dynamic interplay between processes of cohesin-mediated loop extrusion and chromatin compartmentalization. In insects, comprehensive analyses and cross-species comparisons of genome architecture have to date focused on *Drosophila* species mostly. Using Hi-C technology, we characterized genomic landscape of chromatin in five species of *Anopheles* mosquitoes: *An. coluzzii*, *An. merus*, *An. stephensi*, *An. atroparvus*, and *An. albimanus*. We comprehensively profiled all levels of chromatin organization using Hi-C, ChIP-seq, RNA-seq, and FISH methodologies, and studied connections between the distribution of epigenetic marks and 3D-interactions to provide mechanistic explanations underlying observed chromatin structures (Lukyanchikova et al., 2022). We showed that most 3D-structures can be explained by principles described previously for *Drosophila* species. However, at the level of individual loci, we identified specific, extremely long-ranged looping interactions, which are strikingly conserved across *Anopheles* genera for ~100 million years. We detected specific looping interactions, sometimes spanning several dozens of megabases (Mb) (up to 31-Mb) in all studied *Anopheles* species, and validated those interactions by independent FISH approach. Anchors of those loops were represented by large (~200-300-kb) loci, interacting significantly more than expected at their genomic distance (top 1–2 % of all interactions). We compared the gene content located within the loops anchors and, surprisingly, all species in comparison shared orthologous genes. Hi-C contact maps obtained for adults and embryos of *An. merus* demonstrated that the long-distance loops are present at both developmental stages, indicating that they are not developmental-stage specific. Analysis of gene expression profiles within loop anchors resulted in finding several genes with moderate expression. In accordance with this observation, cePC1 analysis showed that loop anchors were enriched in B-compartment regions, so their formation could not be explained by the clustering of active chromatin. Analysis of ChIP-seq results showed moderate H3K27me3 and H2AK119Ubiq enrichment within our loop anchors while some other regions highly enriched in H3K27me3 and H2AK119Ubiq histone marks and located between the loop anchors were not involved in long distant looping. That fact indicates that long-distance chromatin loops are probably not formed as a result of Polycomb proteins activity (PRC1 and PRC2) but by other molecular mechanism.

*Acknowledgements:* The research was supported by RSF No. 22-14-00247 grant.

# Chromosome structural variations of the chicken erythroblast HD3 cell line identified by Hi-C

Maslova A.<sup>1\*</sup>, Plotnikov V.<sup>1</sup>, Nuriddinov M.<sup>2</sup>, Gridina M.<sup>2</sup>, Fishman V.<sup>2</sup>, Krasikova A.<sup>1</sup>

<sup>1</sup> Saint-Petersburg State University, St. Petersburg, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* antonia.maslova@gmail.com

**Key words:** chicken cell lines, chromosome structural variations, HD3, Hi-C, karyotype abnormalities

*Motivation and Aim:* Karyotype abnormalities are frequent in immortalized continuous cell lines, derived from either primary tumors or *in vitro* transformed cells. Structural variations in chromosomes (SVs) can cause dramatic changes in gene expression patterns through direct gene fusion or gene breakage or through disturbance of the spatial organization of gene regulatory framework. This, in turn, can affect cellular phenotype and behavior. While for many continuous mammalian cell lines chromosome SVs are systematically studied and documented, cell lines from other vertebrate models often lack such information. The chicken LSCC-HD3 cell line (HD3), generated from erythroid precursors by transformation with avian erythroblastosis virus (AEV), has been used as a chicken model for erythroid differentiation and AEV-induced hematologic malignancies since 1982 [1]. However, karyotype abnormalities in HD3 cell line have not been assessed. In the present study we analyzed genome spatial organization of HD3 cells by high throughput chromosome conformation capture (Hi-C), which proved to be an effective tool for delineating chromosome rearrangements in mammalian cells.

*Methods and Algorithms:* To detect large-scale chromosome SVs, we obtained Hi-C heatmaps of genomic interactions for HD3 cells and looked for aberrant patterns of high density interchromosomal contacts, which are the hallmarks of interchromosomal translocations. Intrachromosomal inversions, deletions and duplications were revealed by analyzing intrachromosomal contacts in HD3 Hi-C heatmap compared to Hi-C map of chicken embryonic fibroblasts with normal female karyotype (ZW), obtained earlier [2]. We also compared the profiles of A/B compartments and topologically-associating domains (TADs) between the two chicken cell lines. To verify our data, we utilized the computational algorithm HiCtrans [3] for identification of translocations and chromosome breakpoints, as well as FISH with probes to rearranged regions.

*Results:* By visual inspection of Hi-C heatmaps of HD3 cells we identified more than 25 interchromosomal translocations, involving regions of  $\geq 200$  Kb on both micro- and macrochromosomes, which were confirmed by HiCtrans algorithm. Interestingly, HiCtrans tool revealed dozens of additional smaller scale translocations, which were hard to detect by visual analysis of Hi-C maps alone. Based on the Hi-C heatmap pattern we classified most translocations as unbalanced, leading to the formation of heteromorphic chromosomes. In many cases of microchromosome rearrangements, a whole microchromosome participated in the emergence of derivative chromosome with other macro- and microchromosomes, resembling “chromosomal fusions” or Robertsonian translocations between acrocentric microchromosomes. Several *de novo* identified simple and complex chromosomal rearrangements, such as between GGA2

and GGA1qter, between GGA5, GGA4p and GGA7p, between GAA4q, GGA6 and GGA19, were confirmed by FISH with BAC-probes, GGA4p-q paints, centromere-specific probes and probes to different types of chicken tandem repeats. By using paint probe to sex chromosomes, we also validated the GGAW duplication in HD3 cells, which thus have a ZWW karyotype.

*Conclusion:* The HD3 chicken cell line has a severely rearranged karyotype with most of the chromosomes engaged in translocations. Hi-C allows to simultaneously assess the spatial genome organization and chromosomal aberrations in the karyotype of birds with a large number of microchromosomes, which can hamper the investigation of karyotype abnormalities by classical cytogenetic approaches. Our Hi-C data of genomic interactions for the HD3 cell line can be used for comparative studies, analysis of affected genes at breakpoint regions, and general mechanisms leading to karyotype instability.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant No. 19-74-20075) and was performed using the equipment of Resource Center “Molecular and Cell Technologies” (Research Park of Saint-Petersburg State University) and Skoltech Genomics Core Facility.

### *References*

1. Beug H. et al. Erythroblast cell lines transformed by a temperature-sensitive mutant of avian erythroblastosis virus: a model system to study erythroid differentiation *in vitro*. *J Cell Physiol Suppl.* 1982;1:195-207.
2. Fishman V. et al. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* 2019;47(2):648-665.
3. Chakraborty A., Ay F. Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics.* 2018;34(2):338-345.

## Development of a mESC line for studying the condensin II influence on interphase 3D genome organization

Myakinkov I.<sup>1,2\*</sup>, Yunusova A.<sup>1</sup>, Gridina M.<sup>1</sup>, Battulin N.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *i.myakinkov@g.nsu.ru*

**Key words:** mouse embryonic stem cells, cell line, condensin, microcephalin, cohesin, 3D genome

*Motivation and Aim:* Study of the three-dimensional (3D) organization of chromatin became highly relevant recent years as we develop awareness of its great influence on the gene epigenetic regulation and the maintenance of a genome. The way in which DNA folded inside the nucleus heavily affects gene regulation and the development [1].

One of protein complexes responsible for the regulation of 3D genome is cohesin [2]. Besides the cohesion of sister chromatids during cell division, cohesin forms loops of chromatin and larger structures called topologically associating domains (TADs) in interphase. Both chromatin structuration and sister chromatids connection achieved due to ATP-dependent mechanism named loop extrusion (LE), which is literally a threading of DNA through the ring shaped cohesin complex.

Cohesin belongs to the structural maintenance of chromosomes (SMC) complexes family. Besides cohesin, it includes condensin I and condensin II. All of them have a similar structure: the core frame is a triangle “ring” made of two SMC proteins and one kleisin subunit. Condensins known as agents of chromosome condensation before cell division. Strikingly, mechanism of their work is the same as in case of cohesion – LE [3]. A condensed chromosome is a dense thread of loops which were stretched both by condensin I and II. The CTCF protein sets borders for loops made by cohesin. There are no such factors known for condensins, which gives reason to think that other agents regulate the work of these complexes.

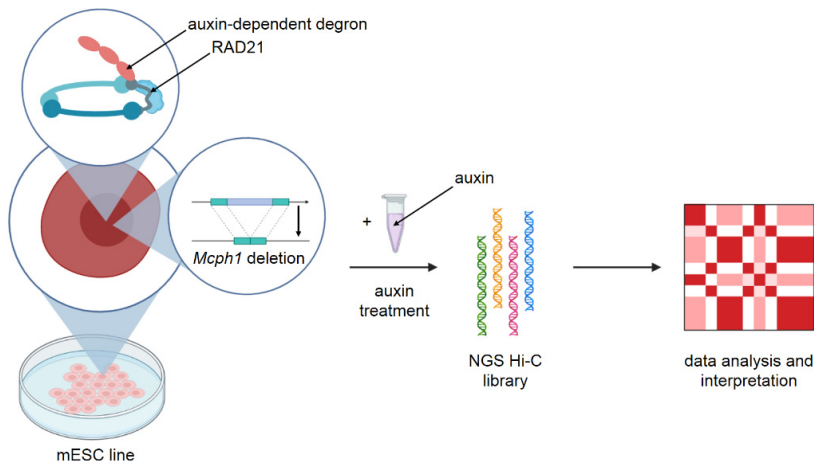
One of such regulating factors is microcephalin 1 (MCPH1) protein. It prevents condensin II from binding and extruding chromatin during interphase. Mutations in *MCPH1* gene are found in patients with primary microcephaly. Their cells possess prophase-like chromatin structure due to absence of MCPH1 protein preventing condensin II from compacting it [4]. Chromosome conformation capture (Hi-C) in these cells shown a drastic change in interphase 3D genome architecture. Very likely, this affects gene regulation, which leads to the development of a disorder.

The simplest way to investigate the gene’s function is to knock it out and then observe resulting phenotype. Unfortunately, this approach can not be used in studies of proteins of SMC complexes because knockout of their genes is lethal. RNA interference is insufficient in this case, as it does not eliminate target protein completely. Therefore, the only option left is a direct depletion of a target protein. This achieved by using degron system. It consists of two elements: first is ligand receptor, which transmits signal to the ubiquitin ligase, and the second is amino acid sequence that is recognized by ubiquitin ligase and linked to the target protein which is wanted to be destroyed [5].

Our goal is to investigate the impact which every complex – cohesin and condensin II – has in the development of MCPH-dependent microcephaly. Since they are likely



interacting, it is hard to define which of them is responsible for each feature (if both of them bind chromatin in interphase). Thence, we decided to develop a mouse embryonic stem cell (mESC) line that can be depleted of cohesin using degron system and possesses a deletion in *McpH1* gene, which allows condensin II to extrude loops in interphase.



**Fig. 1.** Graphical abstract of the work and its planned continuation

**Methods and Algorithms:** We have previously developed mESC line with an auxin-inducible degron system (with auxin as a ligand) for depletion the RAD21 protein (kleisin subunit of cohesin) and with the eGFP tag. Efficiency of the system we have proved by flow cytometry: about 95 % of the RAD21 degrade after the addition of auxin. Next, we have made a deletion in *McpH1* exon of these cells using CRISPR/Cas9 assay (guide design adopted from [4]).

**Results:** PCR genotyping and Sanger sequencing have proved the success of the modification, and the performance of degron system we verified with fluorescent microscopy and flow cytometry.

**Conclusion:** Developed cell line is suitable for further investigation of condensin II impact on 3D genome configuration and, potentially, for modelling of microcephaly on early stages of development.

After the development of a cell line, we performed a Hi-C analysis on control (auxin untreated) and treated cells and constructed a library for the next-generation sequencing. The results of sequencing are still awaited.

**Acknowledgements:** This research was supported by project 121031800061-7 (Mechanisms of genetic control of development, physiological processes, and behavior in animals).

## References

1. Gorkin D.U. et al. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*. 2014;14(6):762-775. doi: 10.1016/j.stem.2014.05.017.
2. Kim Y. et al. Shaping of the 3D genome by the ATPase machine cohesin. *Exp Mol Med*. 2020;52(12):1891-1897. doi: 10.1038/s12276-020-00526-2.
3. Ganji M. et al. Real-time imaging of DNA loop extrusion by condensin. *Science*. 2018;360(6384):102-105. doi: 10.1126/science.aar7831.
4. Houlard M. et al. MCPH1 inhibits Condensin II during interphase by regulating its SMC2-Kleisin interface. *Elife*. 2021;10:e73348. doi: 10.7554/eLife.73348.
5. Li S. et al. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat Methods*. 2019;16(9):866-869. doi: 10.1038/s41592-019-0512-x.

# Simulating of 3D genome data with predefined chromosomal rearrangements

Nuriddinov M.<sup>1\*</sup>, Mozheiko E.<sup>1</sup>, Fishman V.<sup>1,2</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* [nuriddinov@bionet.nsc.ru](mailto:nuriddinov@bionet.nsc.ru)

**Key words:** Hi-C, 3D genomic, chromosome rearrangement, simulation

**Motivation and Aim:** Many genetic based diseases including development disorders and cancer are caused by structural chromosome abnormalities that affect the genome architecture [1]. This means the detection of structural variants and aberrant positions of regulatory elements are clinically relevant. The Hi-C-like methods are shown to be efficient and precise tool to reveal both known and novel chromosomal rearrangements [2]. Using a machine-guided search to mining the Hi-C-like data seems to be useful; unfortunately, the small sizes of available experimental Hi-C datasets describing rearrangements interferes with benchmarking and an adjusting of searching algorithm parameters. To solve this problem, we developed a flexible and tunable tool to simulate the Hi-C-like contact map with the predetermined chromosomal rearrangements.

**Methods and Algorithms:** We developed a python-based software to generate Hi-C matrices of chromatin contacts expected for different types of chromosomal rearrangements. Basing on a given reference, the software estimates a dependence of Hi-C contacts from: a distance between contacted loci; a coverage of target locus by Hi-C reads; a contact preference been defined by chromatin architecture features. Accounting that, the software builds the model of contact map for rearranged genome. Additionally, the software can simulate an experimental noise by the binomial or normal distribution and differences in sequencing depth.

**Results:** We developed the software that simulates contact map for the given chromosomal rearrangement, such as copy number variations, inversion and translocations or their combination. The tool can adjust parameters to simulate different types of Hi-C experiments including capture-Hi-C. The generated models are represented as contact matrices in sparse or Juicebox-preferred .hic-format.

**Conclusion:** Described software allows to train, to tune and to test computational tools searching for chromosomal rearrangements on different types of Hi-C-like data.

**Acknowledgements:** The study is supported by Russian Science Foundation (project 21-65-00017).

## References

1. Jackson M. et al. The genetic basis of disease. *Essays Biochem.* 2018;62(5):643-723.
2. Harewood L. et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* 2017;18(1):125.

# Single-cell DNA methylation analysis of chicken diplotene oocytes

Nurislamov A.<sup>1,2\*</sup>, Lagunov T.<sup>1,2</sup>, Gridina M.<sup>2</sup>, Fishman V.<sup>1,2</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* a.nurislamov@g.nsu.ru

**Key words:** DNA methylation, epigenetics, oogenesis

*Motivation and Aim:* Recent advancements in single-cell genomic studies have opened up a variety of possibilities for investigating developmental processes. Such techniques allowed studying of DNA methylation dynamics during oogenesis in mammals [1, 2]. However, the way in which epigenetic profile changes throughout avian oogenesis is still unknown. In this study, we conducted a bisulfite sequencing for chicken oocyte nuclei in early and late diplotene stages.

*Results:* Our results showed that DNA methylation level decreases on late diplotene stage. Overall methylation levels for both stages are lower in comparison to somatic cells but higher than methylation levels in sperm. We also discovered that regions with low methylation levels in early diplotene nuclei undergo even further demethylation in the later stage and these regions are related to low-methylated regions in sperm.

*Acknowledgements:* This project is supported by RSF grant No. 20-64-46021.

## References

1. Yu B. et al. Genome-wide, single-cell DNA methylomics reveals increased non-CpG methylation during human oocyte maturation. *Stem Cell Rep.* 2017;9(1):397-407.
2. Gu C. et al. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res.* 2019;29(2):110-123.

# Whole-genome analysis of epigenetic control of human GABAergic interneuron differentiation

Pletenev I.<sup>1\*</sup>, Farhi E.<sup>2</sup>, Khrameeva E.<sup>1</sup>, Birnbaum R.<sup>2</sup>

<sup>1</sup>Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup>Ben-Gurion University of the Negev, Be'er Sheva, Israel

\* [ilya.pletenev@skoltech.ru](mailto:ilya.pletenev@skoltech.ru)

**Key words:** epigenetics, chromatin, GABAergic neurons, development

*Motivation and Aim:* GABAergic interneurons play a pivotal role in neural circuitry as they inhibit neural activity via the neurotransmitter gamma-aminobutyric acid (GABA). GABAergic interneurons undergo a complex differentiation process, which, if erroneous, may result in severe developmental abnormalities [1, 2]. They have been actively studied, yet the mechanisms underlying the regulation of their differentiation still remain unclear. Recent studies have revealed the importance of epigenetic control for correct cell differentiation [3]. An important step in analyzing the developmental epigenetic regulation is to target genomic loci with the most prominent epigenetic changes upon differentiation.

*Methods and Algorithms:* The present work encompasses the data on genome spatial organization, chromatin openness, active enhancers, and gene expression of human embryonic stem cells (hESCs) and medial ganglionic eminence (MGE) progenitor cells. We developed an algorithm that considers the combined data to pinpoint the genes potentially important for GABAergic interneuron differentiation.

*Results:* Combined epigenetic data demonstrate drastic changes in chromatin organization between hESCs and MGE progenitors. These changes include reorganization of chromatin compartments, topologically associating domains, and chromatin openness. We find novel enhancers of *FOXG1* and *SOX6*, transcription factors known to be involved in GABAergic interneuron differentiation. Analysis of epigenetic landscape suggests novel genes that might be involved in differentiation of hESCs into MGE progenitors.

*Conclusion:* The transition from hESCs to MGE progenitors is accompanied by the dramatic changes in the epigenetic landscape that potentially mediate the differentiation process. These findings shed light on epigenetic control of GABAergic interneuron differentiation and may be further used to diagnose the interneuron dysfunction.

*Acknowledgements:* The study is supported by the RSF grant 21-74-10102. This is a joint study with Dr. Ana Pombo lab at Max Delbrück Center for Molecular Medicine, Berlin, Germany.

## References

1. Lupiáñez D.G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene enhancer interactions. *Cell*. 2015;161(5):1012-1025.
2. Ling C., Rönn T. Epigenetics in human obesity and type 2 diabetes. *Cell Metab*. 2019;29(5):1028-1044.
3. Gökbuget D., Blleloch R. Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development*. 2019;146(19):dev164772.

## 3D-MAGs – spatial conformations of individual microbial genomes reconstructed from Hi-C metagenomes

Revel-Muroz A.<sup>1\*</sup>, Sobolev A.<sup>1</sup>, Solovyev M.<sup>2</sup>, Vasiliev P.<sup>1,3</sup>, Ulianov S.<sup>1,2</sup>, Ivanova V.<sup>1</sup>, Tyakht A.<sup>1</sup>

<sup>1</sup> *Institute of Gene Biology, RAS, Moscow, Russia*

<sup>2</sup> *Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *Research Centre for Medical Genetics, Moscow, Russia*

\* *revelanastasia@gmail.com*

**Key words:** Hi-C, metagenomics, spatial conformation of genome, microbiome, contact map, regulation of expression, 3D genome, chromosome interaction domain (CID)

*Motivation and Aim:* High-throughput metagenomic sequencing of a microbiome sample coupled with adequate bioinformatic analysis allows reconstructing information about multiple microbial genomes from a single sample. Termed MAGs (metagenome-assembled genomes), each of such approximations represents an unordered sets of contigs. Augmentation of metagenomics with chromosome conformation capture (Hi-C metagenomics) yields Hi-C MAGs that can outperform ordinary MAGs, particularly in terms of completeness and contamination, as well as align the contigs into a chromosome-level scaffold. By reconstructing chromosome contact maps of a MAG similarly as it is commonly done for single eukaryotes, it is possible to evaluate the 3D organization of the microorganism. However, such analysis is yet to be conducted systematically and compared with other layers of genomic information.

*Methods and Algorithms:* Here we propose the concept of a 3D-MAG – spatial structure of a single microbial genome reconstructed from Hi-C metagenome. We investigated the substantiality of the concept on novel experimental data from 3 types of microbiomes of varying diversity – fermented beverages (low diversity), gut microbiome (medium diversity) and thermophilic community from a hot spring (high diversity). Hi-C MAGs were obtained using a recently developed pipeline [1]. For each high-quality Hi-C MAG, its contact map was generated and aligned. 3D genomes were compared with published metatranscriptomic data of related species.

*Results:* 8 samples were processed and at least 3 high-quality 3D-MAGs were received in each community. Most 3D-MAGs were associated with species, whose spatial structure has not been studied yet. We discovered the second diagonal on the contact map of *Lactobacillus brevis*, which is similar to its relative *Bacillus subtilis*. In addition to bacteria, the archaeal 3D genome was also studied. We found an off-centered second diagonal and loops on a contact map of *Thermofilum uzonense*. For a number of 3D-MAGs, the relationship between gene expression and CIDs (chromatin interaction domains) boundaries was confirmed. For one of the major human gut microorganisms *Alistipes finegoldii* we compared the genome spatial structure at two time points to find differences in chromatin conformation related to the boundaries of the CIDs in the region of highly expressed genes. The change might be linked to altered gene expression of the microorganism.

*Conclusion:* Reconstruction of 3D-MAGs using Hi-C metagenomics provides information of the chromosome structure of individual species in various metagenomic

communities. Linking it to other omics-derived information about the microbe provides insights into temporal dynamics of gene regulation *in vivo*. Large scale Hi-C metagenomics profiling involving 3D-MAGs can deepen our understanding of the relation between spatial and functional organization of microorganisms

*Acknowledgements:* This work was supported by the Russian Science Foundation (No. 19-74-10092).

### *References*

1. 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.

## Higher order chromatin structure: new insight with novel correlative approach

Ryumina E.<sup>1,3\*</sup>, Bosov D.<sup>1,3</sup>, Golyshev S.<sup>1</sup>, Moiseenko A.<sup>2</sup>, Kireev I.<sup>1,2</sup>

<sup>1</sup> *Belozersky Institute of Physico-chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

\* *ryu-ekaterina@yandex.ru*

**Key words:** replication, chromatin, CLEM, TEM, STORM, electron tomography

The principles underlying spatial organization of chromatin mostly remain poorly understood. Among the methods that are frequently used in those studies microscopy occupies a prominent place. Thus, great amount of data was obtained with transmission electron microscopy. However, its main advantage – high resolution – comes at a cost: long sophisticated sample preparation requires a lot of time and inevitably casts doubts on structure nativity. The latter particularly raises concerns as chromatin is a very heterogeneous structure sensitive to the experimental conditions. Recent advances in super-resolution light microscopy allow to avoid those issues while keeping resolution relatively high. One of the most promising subdiffractional method is STORM, that can achieve resolution up to 20 nm and have both 2D and 3D modes. However, it is yet unclear to what extent STORM can be applied to studies of chromatin structure. Therefore, STORM results should be verified by method with higher resolution, such as electron microscopy.

Here we present CLEM procedure for combined STORM-TEM imaging of replicative labeled DNA which allows to obtain images of the very same structures using both fluorescent and electron microscopy. This protocol allows for high-efficiency replication labeling compatible with various 2D and 3D-electron microscopy (such as electron tomography) and fluorescent microscopy techniques. We also show that 2D-STORM can be effectively used in studies of replication domains spatial organization, whereas usage of 3D-mode seems to be problematic.

*Materials and Methods:* We adapted method of replicative labeling [1] for strong aldehyde fixation that ensures optimal preservation of chromatin near-native structure during sample preparation for electron microscopy. Our protocol includes Edu-labeling of newly synthesized DNA, label detection with Click-chemistry and biotin-streptavidin and subsequent Nanogold-Ag amplification for TEM. Click-chemistry provides simple and extremely selective labeling of replicated DNA and the use of streptavidin-Nanogold conjugates provides better penetration efficiency even into glutaraldehyde-fixed samples due to the relatively small size of the probe. This approach was applied for correlative superresolution and TEM microscopy by using a mixture of biotinylated and fluorescent (AlexaFluor-647) azides and sequentially imaging the samples by STORM microscopy and labeling the same replicative domains with streptavidin-Nanogold.

STORM images were processed with Thunder-Storm ImageJ plugin and then analyzed with SR-Tesseler (for 2D images) or Python3.7 libraries sklearn and tensorflow.

*Results:* Our modified protocol allowed us to perform electron tomography of labeled DNA with near-native structure. It enabled visualization of chromatin fibers that are

mostly 140–200 nm wide. We were also able to compare images of labeled newly synthesized DNA obtained with STORM and TEM. Those images showed striking similarity, indicating that STORM resolution is more than enough to investigate structure of replicative domains study. Apart from that STORM images can be easily analyzed with multiple tools since STORM presents tables of dots coordinates. Those dots can be clusterized (with DBSCAN for instance) and then a large set of geometrical parameters can be measured. Together with relatively short sample preparation it makes 2D-STORM a promising method in such studies.

STORM-3D turned out to be much more difficult to apply to that issue. This method is based on usage of cylindrical lense and hence astigmatism, which together with calibration data can be used to calculate the Z coordinate. However, dot distribution upon the Z axis seems to be extremely sensitive on certain calibration, which indicates low reliability of such data and makes it almost impossible to compare images obtained using different calibrations.

*Conclusions:* Our protocol allows for high contrast high-efficiency pre-embedding DNA labeling compatible with various light and electron microscopy techniques, for instance TEM, electron tomography, STORM.

Electron tomography reveals the existence of 140–200 nm fibers of 2 hours replication labeled DNA.

2D-STORM is a fast, effective and relatively simple method to examine spatial organization of chromatin.

3D-STORM requires further adjusting to perform properly and produce trustworthy results.

### *References*

1. Deng X. et al. Cytology of DNA replication reveals dynamic plasticity of large-scale chromatin fibers. *Curr Biol.* 2016;26(18):2527-2534.



# The tissue-specific relationship between 3D genome organization and gene expression at the mouse *Slc29a3/Unc5b* locus

Salnikov P.<sup>1,2\*</sup>, Yan A.<sup>1,2</sup>, Valeev E.<sup>2</sup>, Stepanchuk Y.<sup>1,2</sup>, Tihomirov S.<sup>1</sup>, Belokopytova P.<sup>1,2</sup>, Fishman V.<sup>1,2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* paul.salnikov@gmail.com

**Key words:** spatial genome organization, topologically associating domains, CTCF, gene expression

*Motivation and Aim:* Chromosomal folding in interphase nucleus is not random, and each chromosome is organized along its length into topologically associating domains (TADs), allegedly involved in gene expression regulation. Boundaries of these domains are formed by clusters of CTCF binding sites, the distribution of which is highly conserved in evolution [1]. However, full depletion of CTCF protein and disruption of TADs does not lead to massive expression disturbance [2]. There are only rare cases of significant aberrations caused by destruction of the TAD boundaries, and all such mutations affected not so much the deletion of CTCF sites as a change in the distance between regulatory elements [3–5]. Thus, there is no clear evidence of regulatory role of TADs in gene expression. We suppose tissue-specific nature of this relationship, and in this study we aimed to test this hypothesis.

*Methods and Algorithms:* We have created model animal lines using CRISPR/Cas9 mediated gene editing in zygotes. We removed four of CTCF-binding sites at the TADs border by introducing the 5-kb length deletion in the intergenic region (chr10:60755585-60761088, mm10), and two small deletions in sites located in the *Unc5b* intrones (chr10:60775582-60775773, chr10:60778683-60778844, mm10). The absence of CTCF binding was confirmed by ChIP-Seq analysis. The obtained homozygous line was used for a standard Hi-C experiment on various tissues, and was crossed with *Mus musculus castaneus* for an allele-specific expression measurement experiment. We analyzed *Unc5b*, *Slc29a3*, *Cdh23*, *Vsir*, *Psap*, *Sgpl4* gene expression in an extensive tissue set using QIAcuity Digital PCR System. The expression of *M. m. castaneus* alleles placed on homolog with a wild-type spatial organization was used as measurement controls.

*Results:* The *Slc29a3/Unc5b* locus of mouse genome contains one of the strongest insulation border between TADs, genes of the locus are housekeeping and have a pronounced regulatory circumstance. We chose this locus for several reasons, such as the wide range of tissues in which gene expression can be assessed, and the highly conserved spatial organization between species. Allele-specific digital PCR analysis shown that gene expression alterations were minor (less than 25 % for most cases), and were observed quite rarely (only three from six genes changed their expression in five from 10 analyzed tissues). However, we found two cases of oppositely directed changes (the expression of *Slc29a3* in the kidneys decreased to 25 %, while in the cerebellum it increased to 20 %; *Vsir* expression in the olfactory bulb decreased to 20 % and increased to 10 % in the bladder). We chose this cases for a more detailed study.

*Conclusion:* Our result point to the tissue-specific role of spatial genome organization. We have shown that the same deletion of CTCF sites cluster can lead to oppositely directed changes depending on the epigenetic environment and developmental trajectory. But even so, these alterations are minor and do not cause obvious prototypical manifestations that could explain the evolutionary conservatism of spatial landscape of the *Slc29a3/Unc5b* locus. At the same time, in most of analyzed cases we do not find changes of gene expression. Thus, the evolutionary significance of spatial genome organization remains unknown.

*Acknowledgements:* The study is supported by RSF Project 22-14-00247 and Ministry of Education and Science of Russian Federation, grant No. 2019-0546 (FSUS-2020-0040).

### *References*

1. Vietri Rudan M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 2015;10(8):1297-1309.
2. Hyle J. et al. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res.* 2019;47(13):6699-6713.
3. Lupiáñez D.G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell.* 2015;161(5):1012-1025.
4. Franke M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature.* 2016;538(7624):265-269.
5. Despang A. et al. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet.* 2019;51(8):1263-1271.

## Hi-C based translocations detection in human single cells

Stepanchuk Y.K.<sup>1,2\*</sup>, Gridina M.M.<sup>1,6</sup>, Valeev E.S.<sup>1,2</sup>, Saifitdinova A.F.<sup>3,4</sup>,  
Shilova N.V.<sup>5</sup>, Lebedev I.N.<sup>6</sup>, Fishman V.S.<sup>1,2,6,7</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Herzen State Pedagogical University of Russia, St. Petersburg, Russia

<sup>4</sup> International Centre for Reproductive Medicine, St. Petersburg, Russia

<sup>5</sup> Research Centre for Medical Genetics, Moscow, Russia

<sup>6</sup> Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

<sup>7</sup> Artificial Intelligence Research Institute, Moscow, Russia

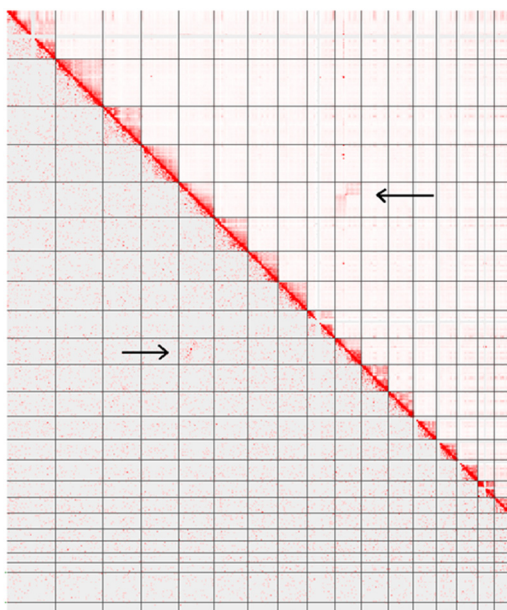
\* y.stepanchuk@g.nsu.ru

**Key words:** Hi-C, single cell Hi-C, chromosomal structural variations, chromosomal aberrations, preimplantation genetic testing

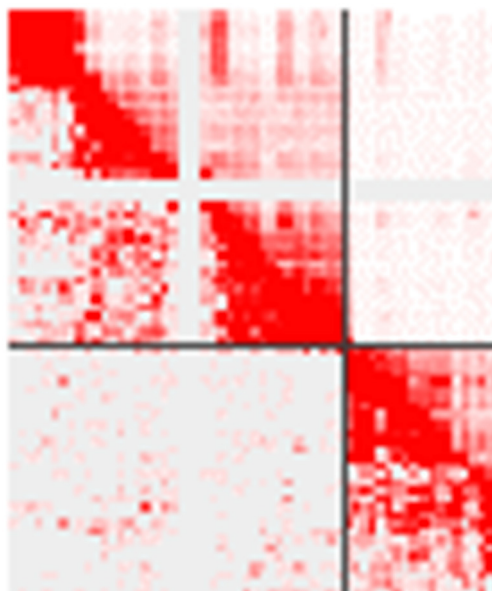
*Motivation and Aim:* More than 10 million people have already been born from *in vitro* fertilization procedure [1]. The number keeps growing as the reproductive health of the population deteriorate and increase the effectiveness of assisted reproductive technologies in overcoming fertility disorders for some categories of patients. Carriers of balanced translocations often suffer from infertility, recurrent miscarriages, and the birth of children with severe developmental disorders. In addition, some offspring with a balanced karyotype will be carriers of translocations, which will lead to the same reproductive problems in the future [2]. In this connection, the need for an effective method of analyzing the embryo is obvious. The preimplantation genetic testing detects unbalanced translocations whereas balanced translocations remain unnoticed. The aim of the study is to adapt single cell Hi-C for detection balanced and unbalanced translocations.

*Methods and Algorithms:* For low-input Hi-C, cells was collected manually with a microscope control and fixed with 2 % formaldehyde for 10 minutes. We made following modifications of the previously published protocol [3]: fixed and lysed cells were embedded in low melting agarose, chromatin ends biotinylation and subsequent target ligation products enrichment were used, and the whole genome amplification was abandoned. The libraries for NGS were prepared using the KAPA HyperPrep kit.

*Results:* We evaluated the possibility of translocations detection using low-input Hi-C. Dozens of iPSC iP10-2 cells with a translocation between chromosomes 5 and 10 (46,XX,t(5;10)) were collected and Hi-C libraries were prepared. The generated contact heatmaps displayed the reciprocal translocation between chromosomes 5 and 10 (Fig. 1), which was previously confirmed by cytogenetics – t(5;10)(q11.2;q11.2)dn and by the population Hi-C. Then we analyzed peripheral blood lymphocytes from patient P47, who, had balanced translocations between chromosomes 8, 11, 20 and 21 (46,XY,t(8;11;20;21)). Four Hi-C libraries were prepared from the P47 lymphocytes. 8 cells were taken for each experiment, is comparable to the number of cells in a biopsy for preimplantation genetic testing. Despite the shallow sequencing (less than 500 thousand reads) of the prepared libraries, we detected t(8,11) translocations (Fig. 2). If 4 replicas were combined, we were able to detect the remaining rearrangements.



**Fig. 1.** The Hi-C heatmaps of 2.5 million iP10-2 cells above the diagonal and dozens of iP10-2 cells – below



**Fig. 2.** The Hi-C heatmaps of P47 lymphocytes. Above the diagonal is population Hi-C data, below – 8 cells

**Conclusion:** Our low-input Hi-C protocol detects balanced reciprocal translocations. In addition, we believe that this method can be proposed for preimplantation genetic testing of structural chromosomal rearrangements to identify both unbalanced and balanced translocations in human embryos.

**Acknowledgements:** This research was supported by grant from Russian Science Foundation 22-14-00247.

### References

1. Fishel S. First *in vitro* fertilization baby-this is how it happened. *Fertil Steril.* 2018;110(1):5-11.
2. Verdoni A., Hu J., Surti U., Babcock M., Sheehan E., Clemens M., Drewes S., Walsh L., Clark R., Katari S., Sanfilippo J., Saller D.N., Rajkovic A., Yatsenko S.A. Reproductive outcomes in individuals with chromosomal reciprocal translocations. *Genet Med.* 2021;23(9):1753-1760.
3. Flyamer I.M., Gassler J., Imakaev M., Brandão H.B., Ulianov S.V., Abdennur N., Razin S.V., Mirny L.A., Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature.* 2017;544(7648):110-114.

# Chromatin loops are involved in spatial organization of replication in budding yeast

Ulianov K.<sup>1\*</sup>, Zhegalova I.<sup>1,2,3</sup>, Khrameeva E.<sup>1</sup>, Gelfand M.<sup>1,2</sup>

<sup>1</sup>Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup>A.A. Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia

<sup>3</sup>Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

\* Kirill.Ulianov@skoltech.ru

**Key words:** budding yeast, chromatin, chromatin loop, replication, epigenetics, Hi-C

**Motivation and Aim:** The loop extrusion model explains the formation of loop-like structures and topologically associating domains in the chromatin. According to the model, SMC complexes extrude DNA thread to form a loop, while two CTCF molecules occupy loop boundaries and constrain the extrusion [1]. However, CTCF-encoding gene appears only in Bilateria, while loop-like structures are observed in the older taxa [2]. Therefore, alternative mechanisms for the determination of the loop boundaries must exist.

**Methods and Algorithms:** We processed the multi-omic data for *Saccharomyces cerevisiae* to annotate positions of loops, estimate the expression rate of genes located in loop boundaries, and study the distribution of histone modifications and replication-related proteins around loops.

**Results:** The profiling of ChIP-seq signal revealed enrichment of Orc-6, Mcm-7, and Rif1 proteins near loop boundaries. All three are known to be responsible for the initialization and/or maintenance of the replication process [3, 4]. Besides, we found characteristic patterns around loop boundaries for histones modifications H3K4me1, H3K4me3, and H3K9ac which were reported to participate in resolving transcription-replication conflict in previous studies [5, 6]. BrdU-seq data demonstrated that newly synthesized DNA is likely placed inside loops rather than at the loop boundaries.

**Conclusion:** The concept of loop-like organization of replication factories in the nucleus was proposed previously by Guillou et al. [7]. Our findings support this idea indicating that the replication machinery might determine the positions of the loop boundaries. But more studies are needed to confirm positions of active replication origins and whether newly-appeared loops in late S phase are formed by extruded DNA undergoing the replication process.

## References

1. Busslinger G.A. et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*. 2017;544(7651):503-507.
2. Heger P. et al. The chromatin insulator CTCF and the emergence of metazoan diversity. *PNAS*. 2012;109(43):17507-17512.
3. Fragkos M. et al. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol*. 2015;16(6):360-374.
4. Hiraga S.I. et al. Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation. *EMBO Rep*. 2017;18(3):403-419.
5. Chong S.Y. et al. H3K4 methylation at active genes mitigates transcription-replication conflicts during replication stress. *Nat Commun*. 2020;11(1):809.
6. Bar-Ziv R. et al. Chromatin dynamics during DNA replication. *Gen Research*. 2016;26(9):1245-1256.
7. Guillou E. et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes Development*. 2010;24(24):2812-2822.

## Detecting three-dimensional contacts of plasmid DNA with chromatin in HEK293T cells

Yan A.<sup>1,2\*</sup>, Salnikov P.<sup>1,2</sup>, Belokopytova P.<sup>1,2</sup>, Fishman V.S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *a.yan@g.nsu.ru*

**Key words:** Chromatin compartments, 4C-experiment with plasmid

*Motivation and Aim:* Interphase chromosomes in cell nuclei are organized into functional domains that differ from each other with respect to their molecular composition. These include active and inactive chromatin domains, DNA repair compartments, nucleoli, Cajal bodies, transcription factories associated with active RNA polymerase II, etc. It was suggested that specific interactions of protein and/or RNA molecules with each other and with DNA sequences results in a biophysical process called liquid-liquid phase separation and explains formation of chromatin compartments [1]. However, the DNA motives determined compartmentalization pattern is still unknown due to the insufficient resolution of existing methods for analyzing the three-dimensional organization of the genome data. Here we propose a new method to uncover sequence determinants of chromatin compartmentalization.

*Methods and Algorithms:* We deliver the target DNA sequence as part of a plasmid to the studied cells, and using the 4C-experiment, we reveal preferences in three-dimensional contacts of the plasmid DNA sequence. We cloned the ancillary sequences required for the 4C-experiment into the pUC19 vector, transfected the resulting vector into HEK293T cells, and obtained and sequenced the 4C-libraries using NGS technologies.

*Results:* Bioinformatic analysis of the obtained data showed that using this approach, it is possible to detect spatial contacts of the plasmid with the genome with an efficiency of 7 %. Based on the analysis of the obtained data, modifications of plasmid vectors and the 4C-protocol were proposed, which allow increasing the amount of data obtained as a result of the 4C-experiment about 3 times.

*Conclusion:* as a result of our work, for the first time we have shown the possibility of using the 4C-method to study trans-contacts of plasmid DNA with chromatin in cell nucleus.

*Acknowledgements:* The study is supported by RSF Project 22-14-00247 and Ministry of Education and Science of Russian Federation, grant No. 2019-0546 (FSUS-2020-0040).

### References

1. Hildebrand E.M., Dekker J. Mechanisms and functions of chromosome compartmentalization. *Trends Biochem Sci.* 2020;45(5):385-396.

## Consecutive chromatin loops in *Dictyostelium discoideum*

Zhegalova I.<sup>1,2,3\*</sup>, Ulyanov S.<sup>4,5</sup>, Galytsyna A.<sup>1,3</sup>, Khrameeva E.<sup>1</sup>

<sup>1</sup>Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup>Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

<sup>3</sup>A.A. Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia

<sup>4</sup>Institute for Gene Biology, RAS, Moscow, Russia

<sup>5</sup>Faculty of Biology, Moscow State University, Moscow, Russia

\* ir.zhegalova@fbb.msu.ru

**Key words:** chromatin, epigenetics, chromatin loops

**Motivation and Aim:** Previous analysis of Hi-C maps [1] showed the absence of TADs at 2 kb resolution, bright points of enriched contacts are visible. The loops have been noted to cover more than half of the *Dictyostelium* genome, have a median size of 20 kb, and their positions are highly conserved, with over 90 % of anchors overlap between stages. Moreover, some loops are stand-alone, but more than half of them are consecutive; their comparative analysis is the aim of this study.

**Methods and Algorithms:** Data were obtained from Hi-C and RNA-seq experiments at 0, 2, 5, and 8 hours after fasting induction. The public ChIP-seq and ATAC-seq datasets [2] were downloaded from Sequence Read Archive (SRA) with accession number SRP269799.

**Results:** It was shown that the level of expression within the loop significantly exceeds that within the boundaries of the loops, and the loops are also characterized by a convergent arrangement of genes around the bases of the loops, thus loops in *Dictyostelium* are probably formed by transcription itself similarly to what was shown previously for mouse [3]. Further analysis revealed loop heterogeneity in terms of expression levels and active chromatin marker signal.

**Conclusion:** Together with previous results, this suggests a possible extrusion mechanism for loop formation in *Dictyostelium discoideum*, with extrusion termination dependent on transcriptional activity. However, the regulatory role of the loops is not excluded, and additional analyzes are required to test this hypothesis.

**Acknowledgements:** The reported study was funded by RFBR, project No. 20-34-90058.

### References

1. Zhegalova I. et al. Features of chromatin structure & gene expression during *D. discoideum* development. Moscow Conference on Computational Molecular Biology (MCCMB), Moscow, July 30th – August 2nd. 2021.
2. Wang S. et al. Role of epigenetics in unicellular to multicellular transition in *Dictyostelium*. *Genome Biol.* 2021;22(1):134.
3. Leidescher S., Ribisel J., Ullrich S., Feodorova Y., Hildebrand E., Galitsyna A., ..., Solovei I. Spatial organization of transcribed eukaryotic genes. *Nat Cell Biol.* 2022;24(3):327-339.

## 1.4 Section “Evolutionary genomics, bioinformatics, and molecular phylogeny”





## Detailed analysis of distribution of repetitive sequences across genomes of *Martes* (Mustelidae, Carnivora, Mammalia)

Beklemisheva V.R.\*, Lemskaya N.A., Prokopov D.Yu., Romanenko S.A., Serdyukova N.A., Utkin Ya.A., Graphodatsky A.S.

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

\* [bekl@mcb.nsc.ru](mailto:bekl@mcb.nsc.ru)

**Key words:** heterochromatin, sable, marten, telomere, nucleolar organizing region, satellite DNA, DNA content

**Motivation and Aim:** Recently more attention is paid to the significant role that repetitive elements may play in the genome functioning. Mustelidae (martens, otters, weasels, minks, badgers, wolverine and others) is a diverse family that diverged from other Carnivora about 17 million years ago [1], that is characterized by evolutionary conserved karyotypes [2, 3] and diploid number close to the ordinal ancestral ( $2n = 38$ ). The genus *Martes* (martens and sables) has a more basal position on the tree of Mustelidae. While evolutionary younger mustelids possess additional entirely heterochromatic chromosome arms, *Martes* autosomes possess moderate amounts of pericentromeric heterochromatin. Here we used an array of modern molecular cytogenetic methods and bioinformatic genome analysis to describe distribution of constitutive heterochromatin and tandem repeats (such as telomeric sequences, macrosatellites and ribosomal genes) on chromosomes of four species from the genus *Martes*: stone marten (*M. foina*,  $2n = 38$ ), sable (*M. zibellina*,  $2n = 38$ ), pine marten (*M. martes*,  $2n = 38$ ), and yellow-throated marten (*M. flavigula*,  $2n = 40$ ).

**Methods and Algorithms:** Interest in the increased amount of the constitutive heterochromatin in mustelids launched the studies of satellite repetitive sequences content and distribution on the chromosomes in the 1980s [4]. Here most abundant tandem repeats (TR) were mined using graph-based sequence clustering with TAREAN pipeline based on stone marten whole-genome sequencing data. Eleven probes were designed from high-confidence TR clusters for FISH investigation of their distribution. Constitutive heterochromatin was characterized by C-banding staining, the AT/GC-content was characterized by CDAG staining (Chromomycin A3-DAPI-after G-banding). The distribution of telomeric repeats and ribosomal genes was shown by FISH. **Results:** Overall, eight satellite probes out of 11 produced distinct unique patterns of chromosome signals. Satellite signals were detected in pericentromeric areas on  $\frac{2}{3}$  of stone marten chromosomes as well as in heterochromatic regions of sex chromosomes. Interestingly, the S46 probe had non-centromeric interstitial localization on the q-arm of X chromosome. We have identified an X-specific centromeric S40 satellite. Some of the satellites had *Martes*-wide distribution. While other satellites were stone marten specific and provided only faint signals in other species.

Constitutive heterochromatin in *Martes* was mostly centromeric on autosomes and on the X with large heterochromatin block on the Yq. We describe AT/GC-rich repetitive areas on chromosomes of four *Martes* species. The heterochromatin in *Martes* is mostly GC-rich: centromeric areas, additional heterochromatin blocks, Y-chromosome heterochromatin. Nucleolar organizing regions are represented by a single site in all four

species and are also GC-rich (as in other mammalian species). Telomeric regions were CDAG-stained as light green indicating mixed AT/GC content, while some telomeric heterochromatin blocks were GC-rich and stained bright green.

*Conclusion:* We deeply characterized repetitive fractions of the genome in four *Martes* species and revealed patterns of satellite localization and clustering. Our findings are summarized on repeat chromosomal maps.

There is apparent heterochromatin variation among species displayed in the amount and the distribution of the heterochromatin and satellite repeat representation. The array of molecular-cytogenetic methods applied here clearly displays variation of the repetitive genome component in species with conserved syntenic groups.

*Acknowledgements:* This research was funded by a grant from the Russian Science Foundation (No. 19-14-00034).

### *References*

1. Law C.J., Slater G.J., Mehta R.S. Lineage Diversity and Size Disparity in Musteloidea: Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Method. *Syst Biol.* 2018;67(1):127-144. doi: 10.1093/sysbio/syx047.
2. Atlas of mammalian chromosomes (2nd edition). Eds.: Graphodatsky A.S., Perelman P.L., O'Brien S.J. USA: Wiley-Blackwell, 2020.
3. Graphodatsky A.S. 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.
4. Lushnikova T.P. et al. EcoRI and BamHI families of repeated sequences in Mustelids. *Genetika (Rus.)*. 1989;25(8):1449-1461.

## Viromic studies of aquatic invertebrates

Butina T.V.<sup>1\*</sup>, Bukin Yu.S.<sup>1,2</sup>, Bondaryuk A.N.<sup>1,3</sup>, Nebesnykh I.A.<sup>1,4</sup>, Denikina N.N.<sup>1</sup>

<sup>1</sup> *Limnological Institute, SB RAS, Irkutsk, Russia*

<sup>2</sup> *Irkutsk State University, Irkutsk, Russia*

<sup>3</sup> *Irkutsk Anti-Plague Research Institute of Siberia and Far East, Irkutsk, Russia*

<sup>4</sup> *Irkutsk State Agrarian University named after A.A. Ezhevsky, Irkutsk, Russia*

\* *tvbutina@mail.ru*

**Key words:** viruses, viromes, metagenomic analyses, aquatic invertebrates, freshwater ecosystems

In recent decades, research in the field of aquatic virology has reached a new level due to the development of new molecular genetic approaches (high-throughput sequencing, metagenomic analysis, etc.). Despite this fact, the study of viruses of aquatic organisms is one of the little-studied areas. Until recently, researchers paid the main attention to the study of viral diseases of commercial and artificially grown (aquaculture) animals (fish, mollusks and crustaceans). Therefore, in general, the diversity of viruses of aquatic invertebrates (especially freshwater ones) remains practically unknown.

This paper presents a review of new data (both published and personal) on the diversity of viruses circulating in the organisms of marine and freshwater (including Lake Baikal) invertebrates (mollusks, etc.), which was obtained using metagenomic and/or transcriptome analysis. We discuss on various methodological approaches to the study of viromes of aquatic organisms (including those used in our studies), including methods for sampling and primary sample preparation, isolation and purification of viral particles and nucleic acids, preparation of libraries of total viral DNA/RNA and their sequencing as well as methods for bioinformatics processing of metagenomic datasets.

*Acknowledgements:* This study was funded by the Russian Science Foundation, grant No. 22-24-01120.

# Bacteriophages as vectors of gene transfer from prokaryotes to eukaryotes

Daugavet M.<sup>1\*</sup>, Sorokina N.<sup>2</sup>, Dikaya V.<sup>1</sup>, Gretsova M.<sup>3</sup>, Podgornaya O.<sup>1</sup>

<sup>1</sup> *Institute of Cytology, RAS, St. Petersburg, Russia*

<sup>2</sup> *Applied Genomics Laboratory, SCAMT Institute, ITMO, St. Petersburg, Russia*

<sup>3</sup> *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

\* *ka6tanka@yandex.ru*

**Key words:** horizontal gene transfer, eukaryotes, bacteriophages

Horizontal gene transfer (HGT) refers to the movement of genetic material between the genomes of unrelated organisms. This process is well documented for bacteria and includes conjugation, transformation and transduction. At the same time, a large amount of data on HGT cases from bacteria to eukaryote genomes has been accumulated. In our previous work, we described the ancient chimeric sequence in the genome of the ascidian *Styela rustica*, probably resulting from HGT. This sequence is coding for protein named rusticalin. According to similarity search rusticalin contain two parts: N-terminal inherited from eukaryotic ancestor and C-terminal strongly similar to bacterial sequence. Aside from this C-terminal part has reliable similarity to the enzyme of bacteriophage A500. Rusticalin homologue in the genome of another ascidian species contain specific nucleotide sequence which is very much alike bacteriophage A500 recombination site. Presence of recombination site in eukaryote genome can explain the integration of foreign genetic material and confirm bacteriophage role in this case of HGT [1].

N-terminal part of rusticalin inherited from eukaryotic ancestor is coding for special amino acid sequence called cysteine-rich repeats. This repeats are widespread among eukaryote proteins and in many cases are neighboring with conserved functional domains typical for bacteria and bacteriophages. In bacterial genomes those domains are frequently found in prophage regions showing their involvement in transduction between different bacteria. We can assume that those domains are repeatedly captured by bacteriophages during transduction process.

Eukaryotic domains with significant similarity (E value  $< 10^{-4}$ ) to prokaryotic sequences belong to 13 different protein families according to Pfam database [2]. This result led us to the assumption that cysteine-rich repeats are marking the sites in the eukaryotic genomes that are susceptible to HGT. We further hypothesize that cysteine-rich repeats in eukaryotic genomes might contain multiple recombination sites for bacteriophage integration. In order to check this hypothesis we constructed the dataset of proteins with cysteine-rich repeats and dataset of bacteriophage recombination sequences described in the literature. 43 % of all genes containing cysteine-rich repeats harbor short (10–16 bp) sequences of recombination sites. While in control dataset of genes without cysteine-rich repeats, 13 % of genes contain sequences of recombination sites (10–13 bp). We can conclude that genes containing cysteine-rich repeats are enriched with short sequences similar to bacteriophage recombination sites. This may possibly facilitate integrations of foreign genetic material in eukaryote genomes involving bacteriophages as vectors.

*Acknowledgements:* This work was supported by Russian Foundation for Basic Research (grant No. 203490077), Russian Science Foundation (grant No. 19–74-20102) and Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1075, signed 28.09.2021).

### *References*

1. Daugavet M.A., Shabelnikov S., Shumeev A., Shaposhnikova T., Adonin L.S., Podgornaya O. Features of a novel protein, rusticalin, from the ascidian *Styela rustica* reveal ancestral horizontal gene transfer event. *Mob DNA*. 2019;10:4. doi: 10.1186/s13100-019-0146-7.
2. Daugavet M.A., Shabelnikov S.V., Podgornaya O.I. Amino acid sequence associated with bacteriophage recombination site helps to reveal genes potentially acquired through horizontal gene transfer. *BMC Bioinformatics*. 2020;21(Suppl. 12):305. doi: 10.1186/s12859-020-03599-y.

# Correlation between the variability of mammalian molecular genetic sequences and the geoclimatic characteristics of their habitats

Efimov V.M.<sup>1,2,3,5\*</sup>, Polunin D.A.<sup>2</sup>, Kovaleva V.Y.<sup>3</sup>, Efimov K.V.<sup>4</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

<sup>4</sup> *Higher School of Economics, Moscow, Russia*

<sup>5</sup> *Tomsk State University, Tomsk, Russia*

\* [efimov@bionet.nsc.ru](mailto:efimov@bionet.nsc.ru)

**Key words:** *Cytb*, amino acid sequences, Glires, climatic factors, correlation

**Motivation and Aim:** The mitochondrial *Cytb* gene is currently the basis of the molecular taxonomy of organisms, as it is considered to be selectively neutral. Recently, however, this assertion has been increasingly questioned. Therefore, we performed a multivariate analysis of the covariance of amino acid sequences of the *Cytb* gene with environmental factors for a set of different rodent mammalian species. This analysis made it possible to reveal the adaptive components of the variability of the amino acid sequences of the *Cytb* gene, which significantly correlate with the geoclimatic characteristics of species habitats.

**Materials and Methods:** Amino acid sequences of the mitochondrial gene *Cytb* (378 aa) of 72 species of Old World rodent mammals Glires, which list the latitude and longitude of sampling sites, were taken from GenBank (NCBI). Based on these geographical coordinates, for each sequence, using the AQUASTAT service, 13 monthly climatic characteristics averaged over 30 years (156 variables in total), as well as the height of the terrain, were obtained.

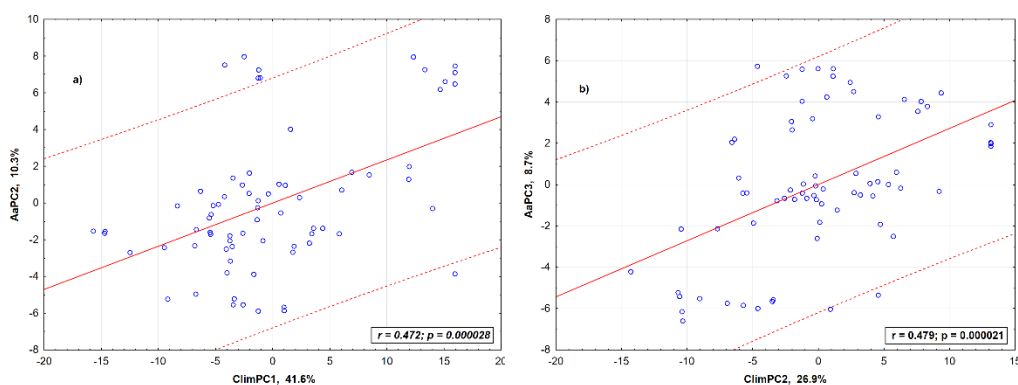
Based on these data, two matrices of interspecies Euclidean distances of size 72x72 were calculated: by primary amino acid sequences and by a set of standardized climatic variables. For amino acid sequences, the square root of p-distance was used. Principal component matrices (PC) were calculated for each distance matrix by the method of principal coordinates [1]. Between them, the matrix of Pearson's correlation coefficients is calculated (Table 1). The Figure shows scatterplots corresponding to significant correlation coefficients. The calculations were carried out using the Jacobi4 package [2].

**Results:** Based on the Table, the first amino acid PC (22.6 % of total variance) does not correlate with any climatic PC. However, the next two amino acid PCs significantly ( $p < 0.001$  according to Bonferroni criterion) correlate with the first two PCs of climatic variability. Judging by the correlations of AaPC1, it reflects the mutual substitutions of amino acids with similar physicochemical properties, which is consistent with the evolution neutrality hypothesis. In turn, ClimPC1 correlates with the temperature gradient associated with the latitude of habitats, while ClimPC2 reflects precipitation and humidification.

**Table 1.** Pearson correlation coefficient matrix ( $\times 1000$ ) between amino acid and climatic principal components

PC	AaPC1	AaPC2	AaPC3	AaPC4	AaPC5	AaPC6	AaPC7	$\lambda$ , %	S, %
ClimPC1	197	472***	66	43	-60	-16	59	41.6	41.6
ClimPC2	-72	133	479***	-163	-164	-86	-244	26.9	68.5
ClimPC3	-94	169	-47	-218	253	28	239	11.3	79.8
ClimPC4	126	43	104	-56	185	156	-274	5.2	85.0
ClimPC5	-184	-82	-33	35	-46	54	68	3.4	88.5
$\lambda$ , %	22.6	10.3	8.7	6.5	5.3	4.1	2.7		
S, %	22.6	32.9	41.6	48.1	53.4	57.5	60.2		

Note: \*\*\* means  $p < 0.001$  according to Bonferroni criterion



**Fig. 1.** Configuration of 72 species of Old World rodent mammals Glires on the plane of climatic and amino acid principal components: a) ClimPC1 – AaPC2; b) ClimPC2 – AaPC3

**Conclusion:** The results obtained support both hypotheses. The first amino acid PC, in all likelihood, corresponds to the hypothesis of neutral evolution, the next two PCs are adaptations to environmental factors. The next task is to identify the positions of substitutions and their biological meaning in both cases.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### References

- Gower J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. 1966;53(3-4):325-338.
- Polunin D. et al. JACOBI4 software for multivariate analysis of biological data. *bioRxiv*. 2019:803684.5.

## Tracing the origin and evolution of the locus containing paralogous *CENH3* genes in cereals

Elisafenko E.A.<sup>1,2,3</sup>, Evtushenko E.V.<sup>2</sup>, Vershinin A.V.<sup>2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>3</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

\* antares@bionet.nsc.ru

**Key words:** Centromere, *CENH3*, Gene duplication, Molecular evolution, Invasion of transposable elements, *Poaceae*

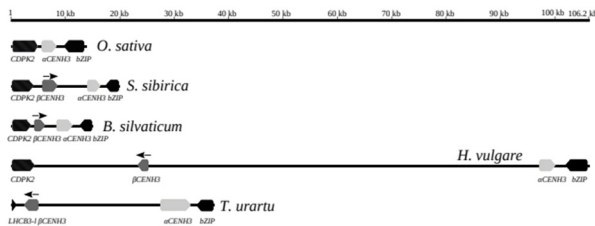
**Motivation and Aim:** The central component of centromere specification and function is the centromere-specific histone H3 (*CENH3*) [1]. The cereal family *Poaceae* is one of the largest and most diverse angiosperm families consisting of more than 12,000 species and more than 700 genera. Some cereal species (maize, rice) have one copy of the gene encoding this protein, while some (wheat, barley, rye) have two,  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes [2]. The discovery of *CENH3* duplications complicated the understanding of the evolution of the *CENH3* genes and raised the question about the mechanisms and causes of gains, losses and selective constraints. Using a homology-based approach, we trace back the mutual evolution of the structure of the *CENH3* genes and structure of the nearby regions in various tribes in the family *Poaceae*.

**Methods and Algorithms:** The sequences of  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes were obtained for 23 species and the deduced protein sequences were used for phylogenetic analysis. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model [3]. Evolutionary analyses were conducted in MEGA X [4]. To determine the exon-intron structure of the  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes, RNA-seq reads from the SRA at NCBI were mapped onto the corresponding reference genome sequences or the assembled contigs of genome sequences (if the whole-genome reference sequences were not available) using HISAT2 [5]. The composition of repeated DNA sequences in the *CENH3* locus in different species was determined using the *Viridiplantae RepeatMasker* program and database [6]. The genes neighboring to *CENH3* genes were identified by similarity with the sequences of *Triticum urartu* and *Oryza sativa* using TBLASTX at NCBI or, locally, BLASTN.

**Results:** We have established that the syntenic group or the *CENH3* locus with the *CENH3* gene and the boundaries defined by the *CDPK2* and *bZIP* genes first appeared about 50 Mya in a common ancestor of the Bambusoideae subfamilies, Oryzoideae and Pooideae. This locus came to Pooideae with one copy of *CENH3* in the most ancient tribes Nardeae and Meliceae (Fig. 1A). The  $\beta$ *CENH3* gene as a part of the locus appeared in the tribes Stipeae and Brachypodieae around 35–40 Mya. The duplication was accompanied by splicing abnormalities, leading to changes in the exon-intron structure and long deletions in the N-terminal domain of the protein  $\beta$ *CENH3*. Purifying selection acts mostly on  $\alpha$ *CENH3*s, while  $\beta$ *CENH3*s and especially their NTT domains form more heterogeneous structures, in which clade-specific amino acid motifs are present. In barley species, the  $\beta$ *CENH3* gene assumed an inverted orientation relative to  $\alpha$ *CENH3* and the left-border gene, *CDPK2*, was substituted with *LHCB-1* (Fig. 1). A head-to-head



orientation of the paralogs is a distinctive feature of all Triticeae species studied. The  $\beta$ CENH3 inversion that we found in the CENH3 locus of the barley species is, in our experience, a Triticeae-only feature and occurred in their common ancestor approximately 15–16 Mya. Invasion by mobile elements and concomitant rearrangements in the structure of the CENH3 locus took place in an independent way during the formation of each species' genome. As a result of these processes, we observe significant variations in the size of the locus, the set of TE families and the amount of changes in their structure (Fig. 1). As the evolution and domestication of plant species went on, the locus was growing in size due to an increasing distance between  $\alpha$ CENH3 and  $\beta$ CENH3 because of a massive insertion of the main LTR-containing retrotransposon superfamilies, *gypsy* and *copia*, without any evolutionary preference on either of them. A comparison of the molecular structure of the locus in the A, B and D subgenomes of the hexaploid wheat *Triticum aestivum* showed that invasion by mobile elements and concomitant rearrangements took place in an independent way even in evolutionarily close species.



**Figure 1. The structure and evolutionary changes of the CENH3 locus.** Syntenic genes comprising the CENH3 locus. The species with important events, such as the formation of a syntenic group (rice), the emergence of the beta paralog (*Stipa*, *Brachypodium*), the inversion of the beta paralog (barley) and replacement of the 5' marker gene (*T. urartu*) in their evolutionary history are given as examples.

**Conclusion:** The CENH3 duplication in cereals was accompanied by changes in the exon-intron structure of the  $\beta$ CENH3 paralog. The observed general tendency towards the expansion of the CENH3 locus reveals an amazing diversity of ways in which different species implement the scenario described above.

**Acknowledgements:** The study is supported by grant of Russian Foundation for Basic Research (project No. 20-04-00699).

## References

1. Postberg J., Forcob S., Chang W.-J., Lipps H.J. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol.* 2010;10:259.
2. Evtushenko E.V., Elisafenko E.A., Gatzkaya S.S., Lipikhina Y.A., Houben A., Vershinin A.V. Conserved molecular structure of the centromeric histone CENH3 in *Secale* and its phylogenetic relationships. *Sci Rep.* 2017;7:17628. doi: 10.1038/s41598-017-17932-8.
3. Jones D.T., Taylor W.R., Thornton J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics.* 1992;8:275-282. doi: 10.1093/bioinformatics/8.3.275.
4. Kuma S., Stecher G., Li M., Knyaz C., Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35:1547-1549.
5. Kim D., Langmead B., Salzberg S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357-360. doi: 10.1038/nmeth.3317.
6. Smit A., Hubley R., Green P. RepeatMasker Open-4.0.6 2013-2015. <http://www.repeatmasker.org> 2018.

## Genome assembly of a new *Wolbachia pipientis* strain: a promising source for studying *Drosophila melanogaster* endosymbiosis

Klimenko A.I.<sup>1,2\*</sup>, Korenskaia A.E.<sup>2,3</sup>, Shishkina O.D.<sup>1,3</sup>, Andreenkova O.V.<sup>1</sup>, Shatskaya N.V.<sup>1</sup>, Vasiliev G.V.<sup>1</sup>, Gruntenko N.E.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [klimenko@bionet.nsc.ru](mailto:klimenko@bionet.nsc.ru)

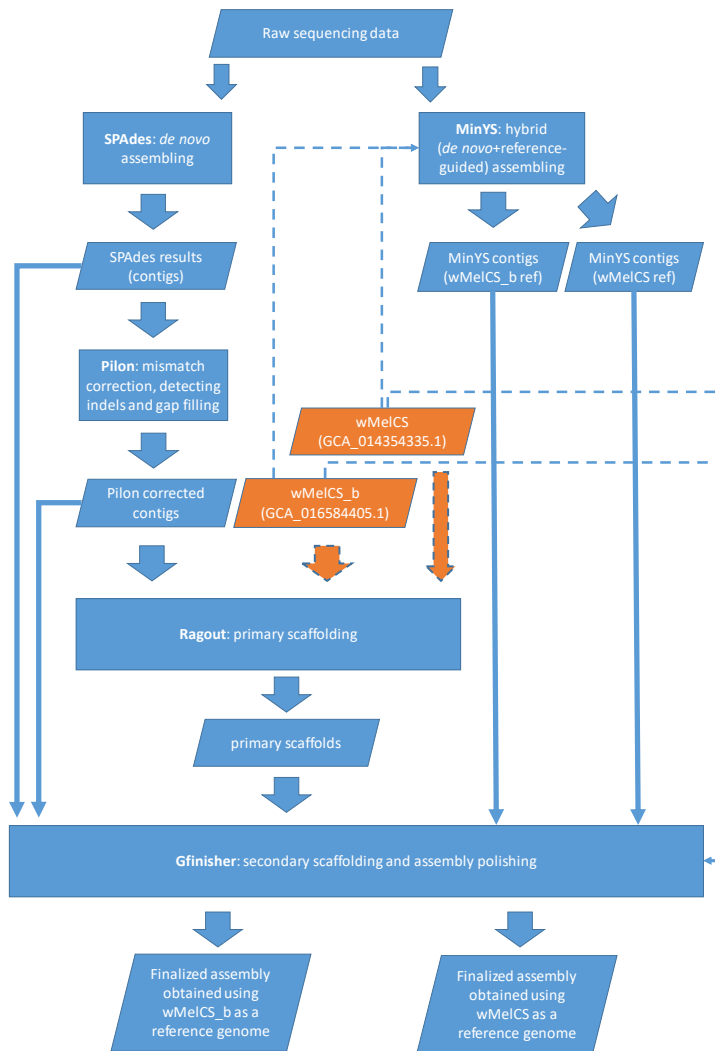
**Key words:** wolbachia pipientis, genomics, microbiology, drosophila melanogaster endosymbionts

*Motivation and Aim:* One of the promising directions in searching the ways to control the number of both harmful and beneficial insects is the study of the possibilities of controlling the insect microbiome, which plays a very important role in the regulation of their vital activity. In this regard, the study of the intracellular symbiotic bacterium of the genus *Wolbachia*, found in more than half of existing insect species, and capable of controlling the reproductive function and physiology of the host insect, may be of particular interest. We had previously found a unique strain *Wolbachia pipientis* (wMelPlus), which provides an increase in stress resistance and fertility of its host, the model species *Drosophila melanogaster* [1].

*Methods and Materials:* Distinct samples containing *Wolbachia* cells of wMelPlus and wMelCS strains were sequenced using Illumina MiSeq platform and characterized using the following pipeline of bioinformatic analysis. We used SPAdes *de novo* assembler [4] and MinYS hybrid assembler [3] for a primary assembling of contigs. Subsequent scaffolding and polishing of the genome assembly have been performed using Pilon, Ragout and Gfinisher [2] employing two independent reference genomes of a wMelCS strain (see Figure 1), which is closely related to the strain of our interest. We assessed the quality of the interim and resulting assemblies using QUAST. The obtained genomes have been annotated using Prokka [5] annotation pipeline and compared with the already published genomes of *W. pipientis*.

*Results:* The genome assembly of a new *W. pipientis* strain has been obtained. Applying the methods of assembly correction and polishing allowed to significantly reduce the number of fragments, approach the expected GC-content, increase the number of identified genes and achieve the expected genome size. Using wMelCS\_b (GCA\_016584405.1) as a reference genome has demonstrated the best assembly quality for both wMelCS and wMelPlus strains, genome fraction ~96 %. However, the genome is still fragmented and there are some misassemblies to be resolved in future using long read sequencing and assembly. Nonetheless, the obtained genome assembly of wMelPlus *Wolbachia pipientis* strain exhibits enough quality for subsequent comparative genomics analysis and can be regarded as a promising source for studying different aspects of its endosymbiosis with *Drosophila melanogaster*.

*Acknowledgements:* The study is supported by RSF grant No. 21-14-00090.



**Fig. 1.** The bioinformatic genome assembly, correction and polishing pipeline used in the current work. Solid lines illustrate the work- and data-flow. Dashed lines depict using as a reference genome

## References

1. Burdina E.V., Bykov R.A., Menshanov P.N., Ilinsky Y.Y., Gruntenko N.E. Unique Wolbachia strain wMelPlus increases heat stress resistance in *Drosophila melanogaster*. *Arch Insect Biochem Physiol.* 2021;106(4):e21776. doi: 10.1002/arch.21776.
2. Guizelini D., Raittz R.T., Cruz L.M., Souza E.M., Steffens M.B.R., Pedrosa F.O. GFinisher: A new strategy to refine and finish bacterial genome assemblies. *Sci Rep.* 2016;6:34963. doi: 10.1038/srep34963.
3. Guyomar C., Delage W., Legeai F., Mougel C., Simon J.-C., Lemaitre C. MinYS: mine your symbiont by targeted genome assembly in symbiotic communities. *NAR Genom Bioinform.* 2020;2(3):lqaa047. doi: 10.1093/nargab/lqaa047.
4. Prjibelski A., Antipov D., Meleshko D., Lapidus A., Korobeynikov A. Using SPAdes De Novo Assembler. *Curr Protoc Bioinform.* 2020;70(1):e102. doi: 10.1002/cpbi.102.
5. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068-2069. doi: 10.1093/bioinformatics/btu153.

## Different approaches to chromosome rearrangement detection in the model cell line

Koksharova G.<sup>1,2\*</sup>, Khabarova A.<sup>1</sup>, Salnikov P.<sup>1,2</sup>, Fishman V.<sup>1,2</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\**g.koksharova@g.nsu.ru*

**Key words:** chromosome rearrangements, recombination, chromosome conformation capture, epigenetics

*Motivation and Aim:* It is widely accepted that chromosome rearrangement can contribute to pathology and disease formation. However, the molecular underpinnings of the mechanism are not fully understood. Recent studies suggest that 3D genome organization is one of the factors at play [1], so to search for epigenetic consequences of chromosome rearrangements we have created a model cell line that allows to generate different combinations of rearrangements between 18 localized sites, with most of the rearrangements being in hemizygous state. For efficient use of this model system there is a strong demand for methods, allowing robust and scalable detection of the induced chromosome rearrangements. To achieve this, we devised two different approaches.

*Methods and Algorithms:* To induce a variety of chromosome rearrangements we genetically modified a near-haploid HAP1 cell line, introducing multiple LoxP-sites into the genome. We induced Cre-recombinase expression in the obtained cells for up to 20 weeks and collected them at different time points. Then we applied an inversePCR-based screening method that allows us to localize both ends of the rearrangement site [2] and thus detect rearrangements between different sites. Coverage was used to estimate the frequency of each rearrangement. We also modified this rearrangement detection strategy by performing digestion and ligation steps in fixed chromatin conditions similar to DNase I Hi-C protocol [3].

*Results:* Both developed methods allowed us to detect rearrangements, with the second one being less biased and more consistent in replicas. According to our results, Cre-mediated recombination occurs much more often in *cis* than in *trans*, although several interchromosomal rearrangements were detected.

*Conclusion:* With a viable method to detect chromosome rearrangements in population screening we are able to choose optimal parameters of Cre-recombinase expression to generate a panel of subclones carrying different combinations of rearrangements. This work makes a step towards dissection of epigenetic changes accompanying different chromosome rearrangements.

*Acknowledgements:* The study is supported by RSF Project 22-14-00247 and Ministry of Education and Science of Russian Federation, grant No. 2019-0546 (FSUS-2020-0040).

### References

1. Redin C., Brand H., Collins R.L. et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet.* 2017;49(1):36-45.
2. Salnikov P.A., Khabarova A.A., Koksharova G.S., Mungalov R.V., Belokopytova P.S., Pristyazhnik I.E., Nurislamov A.R., Somatich P., Gridina M.M., Fishman V.S. Here and there: the double-side transgene localization. *Vavilov J Genet Breed.* 2021;25(6):607-612.
3. Gridina M., Mozheiko E., Valeev E. et al. A cookbook for DNase Hi-C. *Epigenetics Chromatin.* 2021;14(1):15.

## Теория *carcino-evo-devo* – единая теория биологического развития

Kozlov A.P.

*Институт общей генетики им. Н.И. Вавилова РАН, Москва*

*Биомедицинский центр и Санкт-Петербургский политехнический университет Петра Великого,  
Санкт-Петербург, Россия*

Гипотеза возможной эволюционной роли наследуемых опухолей (основная гипотеза) была впервые опубликована нами в 1979 году [Kozlov, 1979] и с тех пор развивалась в серии публикаций. Основная гипотеза заключается в том, что эволюционная роль наследуемых опухолей состоит в предоставлении эволюционирующим организмам дополнительных клеточных масс для экспрессии эволюционно новых генов, возникающих в ДНК зародышевых клеток, и эволюционно новых сочетаний генов, что приводит к возникновению эволюционно новых типов клеток, тканей и органов [Козлов, 1983, 1987, 1988, 2008; Kozlov, 1979, 1996, 2010, 2014].

После выхода в свет нашей монографии *Evolution by Tumor Neofunctionalization* [Kozlov, 2014], впоследствии переведённой на русский [Козлов, 2016] и китайский [Kozlov, 2019] языки, основная гипотеза стала принимать очертания теории. В монографии, кроме обсуждения предпосылок и формулировки основной гипотезы, были также суммированы экспериментальные данные, полученные в лаборатории автора и другими авторами, подтверждающие нетривиальные предсказания основной гипотезы.

В работе, опубликованной в отечественном журнале *Acta Naturae* в 2019 году, был сделан следующий важный шаг в развитии теории: были проанализированы взаимоотношения новой теории с уже существующими биологическими теориями и предпринята попытка объяснения целого ряда необъяснённых биологических явлений с позиций новой теории. Был сделан вывод о непротиворечивости и взаимной дополнительности новой теории по отношению к существующим теориям. Теории было дано название *carcino-evo-devo*, которое происходит из понятий *carcinoembryonic* и *evo-devo*, уже имеющих распространение в научной литературе и восходящих к работам выдающихся отечественных учёных Г.И. Абелева и А.Н. Северцова.

Серия работ, опубликованных в 2022 году [Kozlov, 2022a, 2022b, 2022c], завершает теорию *carcino-evo-devo* в её современном состоянии. В этих статьях разработано учение об опухолеподобных органах и их значении в эволюции [Kozlov, 2022a, b], а также представление о роли биокомпьютерных процессов в эволюции сложности и месте опухолей в этих процессах [Kozlov, 2022c].

Несколько нетривиальных предсказаний теории *carcino-evo-devo* были подтверждены в нашей лаборатории.

В 1987 году, когда было известно всего несколько онкогенов, нами была предсказана необходимость соответствия между числом онкогенов и числом типов дифференцированных клеток, которых тогда было описано около двухсот [Козлов, 1987; Kozlov, 1996]. Действительно, к настоящему времени описано более двухсот онкогенов, что примерно соответствует числу морфологически

описанных типов клеток у человека [Makashov et al., 2019]. Более того, с помощью компьютерной геномики нами было подтверждено предсказание о параллельной эволюции отдельных классов генов, связанных с опухолегенезом и дифференцировкой – онкогенов, генов-супрессоров опухолевого роста, и дифференцировочных генов [Makashov et al., 2019]. Эти примеры ярко демонстрируют предсказательную мощь нашей теории.

Предсказание об экспрессии в опухолях эволюционно новых генов было подтверждено во многих наших публикациях. Нами был описан новый класс генов – эволюционно новые гены, экспрессирующиеся преимущественно в опухолях (TSEEN genes – tumor specifically expressed, evolutionarily novel genes). Мы описали как единичные TSEEN гены (например, ген *PBVO1*, характеризующийся повышенной экспрессией в опухолях молочной железы и простаты, возникший *de novo* у человека), так и целые семейства TSEEN генов [reviewed in Kozlov, 2016].

Возможность участия опухолей в образовании эволюционно новых органов была подтверждена на модели «шапочек» золотых рыбок. «Шапочки» золотых рыбок обладают способностью к неограниченному росту, что является опухолевым признаком, а симметрия, определённая локализация и связь с определёнными стадиями развития являются признаками нормальных органов. Нами было гистологически доказано, что «шапочки» являются доброкачественными опухолями [Козлов и др., 2012]. Это первый описанный в мировой литературе пример искусственного отбора опухолей на новую функцию в организме. С использованием модели трансгенных индуцибельных опухолей рыб нами было показано приобретение прогрессивных функций, не встречающихся у рыб, человеческими ортологами TSEEN генов рыб [Matyunina et al., 2019]. Эти данные являются прямым подтверждением основной гипотезы.

В 2022 году вышла серия из двух наших статей, в которых мы рассмотрели ещё одно нетривиальное предсказание нашей теории: эволюционно новые органы, если они действительно происходили из наследуемых опухолей или опухолеподобных структур, должны рекапитулировать некоторые признаки опухолей в своём развитии [Kozlov, 2022 a, b]. В первой из статей были рассмотрены многочисленные свидетельства, что такие эволюционно молодые органы, как плацента, молочная железа, простата и человеческий мозг, действительно имеют много признаков опухолей, включая инвазивность на определённых стадиях развития и более высокую заболеваемость раком. Мы предложили называть эволюционно молодые органы, имеющие много признаков опухолей, опухолеподобными органами [Kozlov, 2022a].

В следующей работе из этой серии были рассмотрены опухолевые признаки жировых тканей млекопитающих, в совокупности составляющих, согласно современным воззрениям, эволюционно новый жировой орган млекопитающих. Оказалось, что для жирового органа млекопитающих характерно большинство т.н. «основных признаков опухолей» (“hallmarks of cancer”), включая способность к безграничной экспансии и способность инфильтрировать нормальные органы и разрушать их. Таким образом, ожирение представляет собой опухолеподобный процесс [Kozlov, 2022b]. В этой статье нами была описана генная сеть, участвующая в развитии жирового органа млекопитающих и состоящая из ортологов TSEEN генов рыб, описанных нами в предыдущей статье [Matyunina et al., 2019]. Особенностью генов, участвующих в этой сети, является то, что каждый

из них является, в зависимости от контекста, онкогеном или геном-супрессором опухолевого роста. Данные, опубликованные в этой серии работ, позволяют предположить, что при происхождении морфологических новшеств у млекопитающих опухолевые или опухолеподобные процессы могли играть особенно важную роль. Таким образом, учение об опухолеподобных органах является дальнейшим развитием теории *carcino-evo-devo*.

В следующей статье, основываясь на теории пространства возможностей Карла Поппера и представлениях о компьютерных процессах, протекающих в природе, мы предлагаем концепцию биологического пространства возможностей, в котором эволюция осуществляется с помощью биокомпьютерных процессов с участием опухолей в качестве поисковых систем [Kozlov, 2022c].

Таким образом, новая биологическая теория *carcino-evo-devo* объединяет в едином рассмотрении три основных вида биологического развития: эволюционное, индивидуальное и неопластическое развитие, и поэтому имеет потенциал объединения существующих биологических теорий. Теория *carcino-evo-devo* состоит из нескольких разделов, описывающих различные аспекты коэволюции нормального и неопластического развития. Она обладает предсказательной силой и объясняет целый ряд необъяснённых биологических явлений.

## Search and analysis of prophages related to phage AP-16-3 in genomes of *Sinorhizobium* spp.

Kozlova A.P.\*, Muntyan V.S., Vladimirova M.E., Roumiantseva M.L.

Federal State Budget Scientific Institution All-Russia Research Institute for Agricultural Microbiology (FSBSI ARRIAM), Pushkin, St. Petersburg, Russia

\* alexandrak95@mail.ru

**Key words:** bacteriophages, rhizobiophages, prophages, nodule bacteria

**Motivation and Aim:** Bacteriophages can cause lytic infection of bacteria-host which is accompanied by metabolism reprogramming of the host and its destruction as a result. Alternatively, bacteriophages can initiate lysogenic infection of bacteria cell and integrate into chromosome as prophages and then they are transmitted to subsequent generations. Rhizobiophage 16-3 [NCBI RefSeq NC\_011103] was identified in the genome of the *Rhizobium* (*Sinorhizobium*) *meliloti* Rm41 in the 1960s. This is one of the first well studied bacteriophages of nodule bacteria of *S. meliloti* species. In this paper data on the AP-16-3 phage isolated by us in mountainous region of Dagestan is presented. The genome of this phage showed high similarity to genome of the phage 16-3. So, it was a point of interest to evaluate an abundance of phage sequences similar to the AP-16-3 phage in strains from the genus *Sinorhizobium*.

**Methods and Algorithms:** The AP-16-3 phage genome was sequenced using MiSeq, Illumina, assembled (SPAdes, Flye, Racon and Medaka modules, Pilon), and annotated by Prokka. The BLASTn and BLASTp tools were used for genome analysis. Prophage sequences were searched in genomes of 53 strains of *S. meliloti*, of 9 strains of *S. medicae*, and of 7 strains of *S. fredii* (NCBI database) using the PHASTER web server. The phylogenetic tree was constructed using the OPTSIL algorithm of the VICTOR web tool.

**Results:** The object of the study was the soil bacteriophage AP-16-3 showed a high similarity with the temperate rhizophage 16-3. The phage was isolated from a soil sample collected at the northern mountainous region of Dagestan, Caucasus. The AP-16-3 phage genome was 61 kb and 94 ORFs were identified, among which were those whose predicted products are necessary for phage infection cycle. Sequences similar to Rhizobium phage AP-16-3 were detected in genomes of 12 strains of *S. meliloti* and in single strains of *S. medicae* and *S. fredii*. Nucleotide sequence analysis showed that 6 out of 8 prophages were intact, while other two were incomplete or defective prophages. The size of the intact prophage reached 63 kb, while the size of defective and incomplete prophages were 33.1 and 21.7 kb respectively. Prophage sequence identity with AP-16-3 ranged from 79.74 to 92.45 with up to 43 % coverage. The number of ORFs encoding protein sequences in intact prophages were up to 144, while in defective and incomplete prophages were 29 and 27 respectively. It was revealed that all prophages contained ORFs encoding tail proteins but prophages were distinct by ORFs determining synthesis of structural elements of virion. A phylogenetic analysis of the identified prophage sequences was carried out. The application of the OPTSIL algorithm showed that a set of identified phage sequences that were similar to the AP-16-3 phage nevertheless belonged to at least 7 genera of viruses (bootstrap 57 %), presumably of the Siphoviridae



family. Thus, at least 23 % of strains of *S. meliloti* harbored sequences homologous to Rhizobium phage AP-16-3 but these sequences were distinct and assigned to different phage genera.

*Conclusion:* The obtained data support the high prevalence and wide range of lytic activity of Rhizobium phage 16-3 and its related bacteriophages.

*Acknowledgements:* This paper was produced with support from the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075–15-2022–320, dated 20 April 2022, on providing grants in the form of subsidies from the federal budget of the Russian Federation. The grant was provided for state support for the creation and development of a World-class Scientific Center, “Agrotechnologies for the Future”.

## Development and application of a computational pipeline to analyze genes encoding phospholipase domains in flatworms

Lachynova M.E.<sup>1,2\*</sup>, Turnaev I.I.<sup>1</sup>, Afonnikov D.A.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* m.bocharnikova@g.nsu.ru

**Key words:** phospholipase A2, domains, flatworms, molecular evolution, bioinformatics pipeline

*Motivation and Aim:* Phospholipases are a group of enzymes (class of hydrolases) that catalyze the process of phospholipid hydrolysis [1]. This work will focus on A2 phospholipases (PLA2), which detach the SN-1 and SN-2 acyl chain. They are known to be venom toxins in snakes [2]. PLA2 are classified into several types, such as secreted, cytosolic, and calcium-independent [2]. Secreted PLA2 (sPLA2) are the best studied and has quite a wide range of functions. In humans sPLA2 are expressed and released by human inflammatory cells including macrophages, monocytes, T cells, mast cells, and neutrophils, and increased concentrations of different isoforms of sPLA2 have been detected in the blood of patients with inflammatory and autoimmune diseases [2]. In contrast to snake or human phospholipases, PLA2 homologs in other taxons are poorly studied. The analysis is complicated by the fact that PLA2 from different groups have a similar structure, despite the variation in domain composition.

In this work, we search for A2 phospholipases in flatworms, a taxon which comprises of an enormous diversity of species occurring in seas, lakes and land masses. As parasites flatworms pose a threat to public health, especially in Siberia and South-East Asia. Diseases caused by parasitic flatworms can cause serious consequences such as cholelithiasis or cancer [3].

The aim of this work is to create a computational pipeline for bioinformatics search and analysis of PLA2 genes in genomes, their functional annotation and study on its basis the features of structure, function and evolution.

*Methods and Algorithms:* The bioinformatics pipeline inputs annotated phospholipase A2 of vertebrates, and flatworm genomic data. The output of the pipeline is the list of PLA2-domain containing protein sequences, their classification in orthologous groups, phylogeny and domain structure.

*Results:* PLA2 containing proteins in flatworms were identified and classified according their orthologous groups of 11 families (G1, G2, G3, G4, G5, G6, G7, G10, G12, G15, and G16), which have common ancestors with the molluscan genes, were first identified. We also examined the primary structure and domain composition of the PLA2 proteins. The analysis of gene duplication events showed that the phospholipases of groups G4, G6, G15 are presented in free-living and parasitic flatworms, and the phospholipases of groups G1, G2, G5, G7, G10, G12, G16 are presented mainly in free-living organisms. Group G3 phospholipases among parasitic flatworms are found only in trematodes, include an N-terminal domain with unknown function and are separated into two groups based on the presence of the WxGxG motif in the calcium-binding loop.

*Conclusion:* From this work follows that flatworms have several groups of PLA2. PLA2 of flatworms and PLA2 of mollusks were diverged from one common ancestor.

Trematodes has an atypical primary structure of PLA2 G3, whether it is associated with carcinogenesis yet to find out.

*Acknowledgements:* The work was supported by the budget project No. FWNR-2022-0020.

### *References*

1. Knorre D.G., Myzina S.D. Biological Chemistry. M.: High School, 1998.
2. Dennis E.A. et al. Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. *Chem Rev.* 2011;111(10):6130-6185.
3. Report G. On the State of Sanitary and Epidemiological Welfare of the Population in the Russian Federation in 2014: State Report. T. 206. M.: Rospotrebnadzor, 2015.

## Evaluation of the bacterial ribosomal protein S1 gene (*rpsA*) from the position of structural repetition

Machulin A.<sup>1\*</sup>, Deryusheva E.<sup>2</sup>, Galzitskaya O.<sup>3,4\*</sup>

<sup>1</sup> *Skryabin Institute of Biochemistry and Physiology of Microorganisms, Federal Research Center "Pushchino Scientific Center for Biological Research of the RAS", Pushchino, Russia*

<sup>2</sup> *Institute for Biological Instrumentation, Federal Research Center "Pushchino Scientific Center for Biological Research of the RAS", Pushchino, Russia*

<sup>3</sup> *Institute of Protein Research, RAS, Pushchino, Russia*

<sup>4</sup> *Institute of Theoretical and Experimental Biophysics, RAS, Pushchino, Russia*

\* *and.machul@gmail.com, ogalzit@vega.protres.ru*

**Key words:** protein S1, *rpsA* gene, structural repetition

**Motivation and Aim:** The family of bacterial ribosomal S1 proteins (*rpsA*) with repetitive S1 domains differs in a canonical sense from other similar proteins [1]. The S1 protein is essential for cell viability because it interacts with both mRNA and proteins. Protein domain repeats are known to arise due to tandem duplications of internal genes. Analysis of the variability of base composition and tandem gene duplications in *rpsA* genes using a genomics-based approach can provide insight into the distinctive features and possible function and evolution of structural repetition [2].

**Methods and Algorithms:** To study the mechanisms of repeat expansion, we studied 1324 *rpsA* sequences of S1 proteins with different number of S1 domains. Search, collection and analysis of the data were performed using a script implemented in PyCharm v.2018 software development environment, based on Python 3.3 language (<https://www.python.org>). The Multiple Sequence Alignment was implemented by the Clustal Omega service (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). For codon nucleotide sequence alignment was done by MEGA Software using Clustal (codon) (<https://www.megasoftware.net/>). To evaluate sequence similarity MEGA pairwise distance analysis was used.

**Results:** Analysis of *rpsA* showed that the gene regions encoding individual S1 domains have no a strictly repetitive structure between groups containing different number of domains. The part of *rpsA* encoding the central domains in multidomain S1 proteins is more conserved than the terminal domains. The maximum value of *rpsA* identity for full-length proteins was found for S1 proteins containing six structural domains (58 %). The four- and six-domain S1 proteins predominates. The regions of *rpsA* genes encoding adjacent domains are more identical. The regions of the gene encoding residues that form the RNA-binding site remain conserved.

**Conclusion:** All the results obtained characterize the structural features of *rpsA* that determine the functioning of the S1 proteins. The portions of the *rpsA* gene encoding central domains in multidomain ribosomal S1 proteins are more conserved than the terminal domains what correlates with the assumption that duplication is found mostly in the central region of a protein chain between other repeats. Also, analysis of datasets in [3] would suggest that simultaneous duplication of several domains underlies the formation of repetitive regions, while duplication of a single domain is rare. This fact may explain the predominance of four- and six-domain ribosomal S1 proteins. At the

same time, the parts of *rpsA* genes encoding adjacent domains are more identical what correlates with the data suggesting that duplication occurs mainly in the middle of the protein chain between other repeats.

*Acknowledgements:* The study is supported by the Russian Science Foundation, grant No. 18-14-00321 (Galzitskaya O.).

### *References*

1. Deryusheva E.I. et al. Sequence and evolutionary analysis of bacterial ribosomal S1 proteins. *Proteins*. 2021;89:1111-1124.
2. Deryusheva E.I. et al. Structural, functional, and evolutionary characteristics of proteins with repeats. *Mol Biol*. 2021;55:748-775.
3. Björklund A.K. et al. Expansion of protein domain repeats. *PLoS Comput Biol*. 2006;2:e114.

## Study of distribution and evolution of restriction-modification systems with restriction endonuclease of the AlwI family

Makarikova O.<sup>1\*</sup>, Popova A.<sup>2</sup>, Alexeevski A.<sup>3,4</sup>

<sup>1</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Moscow Polytechnical University, Moscow, Russia

<sup>3</sup> A.N. Belozersky Research Institute, Lomonosov Moscow State University, Moscow, Russia

<sup>4</sup> FSI FSC SRISA RAS, Moscow, Russia

\* makarikovaolga82@gmail.com

**Key words:** restriction-modification, Pfam domains, phylogeny

*Motivation and Aim:* Restriction-modification (R-M) is the defense system of prokaryotes against phage and plasmid infection. The system has two enzymatic activities: restriction endonuclease (RE) and methyltransferase (MTase). The former hydrolyzes invaded DNA within or near their recognition site whereas the latter prevents self degradation of the cell genome by methylation of the recognition site. The system type II REs of the AlwI family represented in Pfam by the RE\_AlwI domain are promised for bioengineering usage as nickases [1]. The aim of this work is to study the distribution and evolution (horizontal gene transfer) of R-M systems containing proteins of the family.

*Methods and Algorithms:* Pfam HMM-profile RE\_AlwI was marked according to the structural domains (DNA-recognition, linker, catalytic) of the solved 3D structure of N.BspD6I nickase [1] (pdb code 2EWF) by aligning the sequence of it with the profile. New HMM-profile cat\_AlwI for complete catalytic domain was created using HMMER3 on the base of a validated and non-redundant sample of 87 sequences of the catalytic domain only from REBASE (database containing information on all R-M systems) proteins with the RE\_AlwI domain.

HMM-profiles RE\_AlwI and cat\_AlwI were compared by searching all proteins of 1089 complete prokaryotic proteomes from Uniprot database, containing at least one protein with the RE\_AlwI Pfam domain, according database reference field in the entry annotation.

We have classified R-M systems as follows. A class is defined as a group of R-M systems with the same sets of catalytic domains of REs and MTases. All 818 R-M systems in REBASE having RE with the cat\_AlwI domain were separated and classified. For the most represented class with one RE and two MTases of the MethyltransfD12 Pfam family genome coordinates of the genes of RE proteins found by the cat\_AlwI profile in the abovementioned 1089 proteomes and MTases found in the same 1089 proteomes by the MethyltransfD12 profile were obtained. Proteins located in the same locus and the most likely associated with the same R-M system were selected.

Phylogenetic trees of 79 RE and 158 unique MTase sequences were built for further evolution analysis.

*Results:* The RE\_AlwI domain profile was shown to cover only a small part of the catalytic domain. Therefore we have constructed the new cat\_AlwI domain based only on the catalytic one. Hits of the cat\_AlwI HMM-profile were found in 1005 REs with RE\_AlwI domains according to the Uniprot entry annotation. We have shown that

84 sequences from 1089 complete prokaryotic proteomes containing RE\_AlwI but not cat\_AlwI domain are found by RE\_AlwI profile due to the similarity of the DNA-binding and the linker domains rather than the catalytic one. Therefore the cat\_AlwI profile proved to better represent the AlwI family and it has been used for further investigations.

Intriguingly R-M systems having RE with the cat\_AlwI domain make up 10 classes while their corresponding methyltransferases belong to 3 different families. The class with one RE and two MTases of the MethyltransfD12 Pfam family turned out to be the most represented.

For class including 24 R-M systems with RE of AlwI family and two MTases of N6\_N4\_mtase family, phylogenetic tree of 48 MTases was shown to divide into two clades of equal size, so that one MTase belongs to one clade and the second from the same system to another. The uncommon separation of MTases tree turned out to occur in the most represented class for 158 MTase sequences as well. We are excited to see the differences in the MTase sequences that may be associated with non palindromic recognition sites which requires further investigation.

*Conclusions:* The new profile based on the catalytic domain of N.BspD6I nickase was shown to better represent the AlwI RE family. The results obtained in the phylogenetic analysis suggest that two MTases belonging to the one R-M system undergo divergent evolution. Differences observed in the MTase sequences may have occurred in response to emergence of non palindromic recognition sites.

*Acknowledgements:* The study is supported by the Russian Science Foundation grant No. 21-14-00135.

### *References*

1. Zheleznaya L.A. et al. Nicking endonucleases. *Biochemistry (Moscow)*. 2009;74(13):1457-1466.

## Textual and structural analysis of RNA polymerase II core promoter of evolutionary diverse organisms

Melikhova A.V.<sup>1</sup>, Anashkina A.A.<sup>2\*</sup>, Il'icheva I.A.<sup>2</sup>

<sup>1</sup> *I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

\* *anastasia.a.anashkina@gmail.com*

**Key words:** RNA polymerase II promotor, sequence specificity DNA structure

**Motivation and Aim:** The general structure and action of all eukaryotic and archaeal RNA polymerases machinery have an astonishing similarity despite the diversity of core promoter sequences in different species. The goal of our work is to find common characteristics of the DNA region that define it as a promoter for the RNA polymerase II (Pol II). We have analyzed the profiles of textual and a large number of physical and structural characteristics, namely mono- and dinucleotides distributions, frequencies of occurrence of tetranucleotides TATA and AAAA, stacking energy, roll, stiffness of the duplex structure to roll alteration, slide, stiffness of the duplex structure to slide alteration, mobility to bend towards major groove, ultrasound cleavage intensities, and DNase I cleavage intensities. The indexes of these characteristics in each position of core promoter sequence were averaged over representative sets of promoters for each of the evolutionary diverse species from animals, plants and unicellular fungi and a unicellular protozoan (*H. sapiens*, *M. mulatta*, *M. musculus*, *R. norvegicus*, *G. gallus*, *C. familiaris*, *D. melanogaster*, *A. mellifera*, *D. rerio*, *C. elegans*, *A. thaliana*, *Z. mays*, *S. cerevisiae*, *S. pombe*, *P. falciparum*).

**Methods and Algorithms:** The sets of nucleotide sequences from the core promoter regions of different organisms, aligned at the TSS, were retrieved from the EPD New section of the Eukaryotic Promoter Database (EPD) (<http://epd.vital-it.ch>). For our analysis we used numerical parameterization at the level of ten different base-pair steps, collected at the database DiProDB (<http://diprodb.fli-leibniz.de>), as well as the relative intensities of cleavage of the central phosphodiester bond in 16 dinucleotides and 256 tetranucleotides [1], and the new data on the propensities of hexanucleotides to the DNase I cleavage [2]. Methods are described in detail in the article [3]. We have written programs in Python 3.10 to analyze sequences of RNA polymerase II promoters. Logos were made at <http://weblogo.threeplusone.com>.

**Results:** The core promoters (80 b.p. at positions from -50 to +30) of all analyzed organisms, representing the kingdoms of animals, plants, fungi, and protozoa have a common structural organization. Regardless of the AT content in genomic DNA, they have two singular regions: a hexanucleotide with coordinates -2 -- +4 (INR), surrounding the transcription start site (TSS), and an octanucleotide, separated from TSS at a distance of about 28 b.p., located upstream. In the TSS position (-1, +1), the occurrence of PyPu/PyPu is exceptionally high, with a noticeable predominance of the d(CA/TG) dinucleotide. The minor groove of TATA-box is broadened and conformational motion is decreased in that region. Special characteristics of conformational behavior are revealed in mammalian, best of all in *H. sapiens*, at the



region, which connects the end of TATA-box and the transcription start site (TSS). The intensities of conformational motions in the complementary strands are periodically changed in that region in opposite phases.

*Conclusion:* Comparison of profiles of textual, structural, structural-dynamic and physico-chemical characteristics of the nucleotide sequences of promoters of fifteen organisms representing animals, plants, unicellular fungi, as well as unicellular protozoan (*P. falciparum*), allows us to conclude that there exist a general scheme of the structural organization of the core promoter regions for all organisms, which does not depend on the level of their evolutionary development and the AT content in their genomes. Sequences in singular regions (TATA box and TSS) are not random. Their selection was based on the conformational characteristics of dinucleotides and their nearest neighbors. The number of sequence variants, that represent the TBP binding region (TATA-box) depends on the percentage of AT content in the genome. Conformational characteristics of the octanucleotide in this position can affect the kinetics of DNA-protein complex formation in the promoter of a particular gene. Obtained results may be useful in genetic engineering for artificial modulation of the promoter strength.

*Acknowledgements:* This research was funded by Russian Science Foundation, grant No. 21-14-00346.

### *References*

1. Grokhovsky S.L. et al. Sequence-specific ultrasonic cleavage of DNA. *Biophys J.* 2011;100(1):117-25.
2. Lazarovici A. et al. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc Natl Acad Sci USA.* 2013;110:6376-6381.
3. Il'icheva I.A. et al. Structural features of DNA that determine RNA polymerase II core promoter. *BMC Genom.* 2016;17(1):973.

## Essential and advantageous genes detection in prokaryotic genomes

Moshensky D.\*, Alexeevski A.

*A.N. Belozersky Institute of Physico-Chemical Biology MSU, Moscow, Russia*

\*moshenskydenis@gmail.com

**Key words:** essential gene, genes annotation, comparative genomics, Ka/Ks, orthologs alignment

Among all proteins, there are those whose loss of functionality leads to a significant adaptation decrease. Such proteins are claimed as essential proteins. Essential proteins genes detection is an important challenge for the fight against pathogens and for biotechnologies development. In the first case, essential proteins could be the targets for new antibiotics. In the second, essential proteins should be remained while an industrial strains modification or should be replaced with proteins with the same functions, resistant to required conditions. Essential proteins detection is conducted with both experimental and computational methods. Experimental methods have a number of drawbacks: extreme difficulty, strong dependency on the quality of a genome annotation, weak reproducibility. Computational methods use comparative genomics approaches, genes expression analysis, proteins interactions and other approaches. Most of them produce accurate results only for well-studied organisms because these methods require qualitative annotation not only for analyzed genome but also for related organisms. Besides, generally, these methods are based on additional experimental data like RNA-seq data or transposon mutagenesis data. We have developed a program EAGLE (Essential and Advantageous Genes Location Explorer, <https://github.com/loven-doo/EAGLE>) for essential genes detection in bacterial genomes. This program predicts essential and advantageous (other non-essential functional genes) genes for an input genome sequence using only genome sequences for related bacteria that are organized in the database called EAGLEdb. EAGLEdb consists of basic taxons (taxons of same level or a set of same level taxons). EAGLE detects all ORFs in input genome and calculates 11 features for each ORF. Among these features are the difference between reference phylogenetic tree and phylogenetic tree of an ORF orthologs from the basic taxon genomes, standard deviation from the uniformity for conservative columns in ORF orthologs alignment, Ka/Ks ratio [1–3] the length of the ORF, representation in the basic taxon genomes, two features describing stop-codons distribution, four features based on distances between sequences in ORF orthologs alignment.

We have applied our program to the genomes from DEG database [4]. Functional genes in complete genome of an organism can be detected with up to 90 % accuracy using only relative genomes without any annotation.

Acknowledgements: This work is supported by RSF grant No. 21-14-00135.

### References

1. Yang Z., Bielawski J.P. *Trends Ecol Evol.* 2000;15(12):496-503.
2. Yang Z., Nielsen R. *Mol Biol Evol.* 2000;17:32-43.
3. Wang D., Zhang Y., Zhang Z., Zhu J., Yu J. *Genomics Proteomics Bioinform.* 2010;8(1):77-80.
4. Zhang R. *Nucleic Acids Res.* 2004;32:D271-D272.

## Analysis of environment adaptation features of the bacteria of *Flavobacterium* genus by comparative genomics

Petrushin I.<sup>1,2</sup>, Gutnik D.<sup>2</sup>, Pushmina E.<sup>2\*</sup>, Belikov S.<sup>1</sup>, Chernogor L.<sup>1</sup>

<sup>1</sup> Limnological Institute, SB RAS, Irkutsk, Russia

<sup>2</sup> Irkutsk State University, Irkutsk, Russia

\* nodreamistoobig@yandex.ru

**Key words:** environment adaptation, *Flavobacterium*, polysaccharide utilization loci, comparative genomics

The bacteria in natural ecological niches evolve in a highly competitive environment. Struggling for nutrient resources and defending against pathogens bacteria adapt to the environment or host. As a rule, the more stable habitat conditions are the smaller set of genes the bacterium needs. Representatives of the *Flavobacterium* found in aquatic, bottom sediment, terrestrial, polar and glacial niches, some of them are host associated and known as animal pathogens and plant endophytes. There are some strains discovered in the engineered environment. Like many of the Bacteroidetes they are capable of decomposing a wide range of polysaccharides. For that their genome contains polysaccharide utilization loci (PULs). A number of researchers claim that genome organization and size depend on niche (aquatic or terrestrial one) (Kolton et al., 2013; Wan, 2019). Moreover, PULs composition differs for “aquatic” and “terrestrial” strains. In present work we consider to narrow the habitats for: soil (rhizosphere), host-associated, polar (glacier) and aquatic (including sediments). Besides the metabolism of polysaccharides, we are interested in the impact of other bacterial life support systems. Comparing such data of the amount of the genes involved in biochemical processes, it becomes possible to define the degree of similarity of selected strains. In a previous work a strain of *Flavobacterium* sp. SLB02 was isolated from diseased freshwater sponges of the species *Lubomirskia baicalensis* (Petrushin et al., 2020). The genome of that bacterium has a size that is characteristic of “terrestrial” strains isolated from an organism living in fresh water. The goal of this work is analysis of environment adaptation signs of the representatives of the family *Flavobacterium* and, in particular, the strain of *Flavobacterium* sp. SLB02 by methods of comparative genomics.

**Methods and Algorithms.** There are more than 600 available genomes of the genus *Flavobacterium* in the NCBI GenBank database. For the analysis we took a list of strains described in (Kolton et al., 2013; Wan, 2019) and extended the list by representatives from the different niches. Genes of polysaccharide utilization (glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, carbohydrate-binding modules) were detected by dbCAN2 software (Zhang, 2018). Functional annotation of genes was made by prokaryotic genome annotation service RAST SEED. The whole genomes phylogeny was constructed using the PhyloPhlAn software based on 400 conservative protein sequences (Fig. 1). The closest strain is *Flavobacterium* sp. Leaf82, isolated from leaves of *A. thaliana*. Visualization of the strains composition based on a set of features in multidimensional space was performed by different methods: the principal component (PCA), the multidimensional scaling (MDS) and t-distributed stochastic neighbor embedding (t-SNE).



**Fig. 1.** Phylogenetic tree of the strain SLB02 with closely-related species of *Flavobacterium*. The multilocus tree was built based on a maximum-likelihood method of approximately 400 universal marker genes by PhyloPhlAn with the maximum-likelihood method

**Results and Discussion:** We performed the analysis of several groups of genes related to environment adaptation of representatives of the *Flavobacterium* genus. A diagram that allows us to estimate the similarity of strains by a set of characteristics was constructed with the help of multidimensional analysis. The most of the analyzed strains belongs to one of two clusters we determined as aquatic/sediments and soil ones. Some of host-associated strains are close to the aquatic ones, but *Flavobacterium* sp. SLB02 is close to soil, rhizosphere and endophyte samples. According to the previous studies, host-associated bacteria have several features: genome reduction, diverse CRISPR-Cas system, secretion systems and higher GC-content than for free-living species. The genome of *Flavobacterium* sp. SLB02 has reduced CRISPR-Cas system (only two spacers), has size typical to terrestrial or soil strains and has no any complete secretion systems. To study to the possible origin of *Flavobacterium* sp. SLB02 we extracted the sequences of 16S regions from the metagenomic sequencing data from water samples of Lake Baikal and performed the phylogenetic analysis, but no close related sequences were found. The applied technique can be used to choose the strains with similar genomic features, associated with habitat conditions. It can help to overcome the difficulty of defining the habitats of the bacteria with unknown origin. The aim of further work is a detailed analysis of the genomic features of adaptation to the niches or host (for the host-associated strains).

## References

1. Kolton M., Sela N., Elad Y., Cytryn E. Comparative Genomic Analysis Indicates that Niche Adaptation of Terrestrial Flavobacteria Is Strongly Linked to Plant Glycan Metabolism. *PLoS One*. 2013;8(9):e76704.
2. Petrushin I., Belikov S., Chernogor L. Cooperative Interaction of Janthinobacterium sp. SLB01 and Flavobacterium sp. SLB02 in the Diseased Sponge *Lubomirskia* baicalensis. *Int J Mol Sci*. 2020;21:8128.
3. Wan X. Comparative genome analyses reveal the genomic traits and host plant adaptations of flavobacterium akiainvivens IK-1t. *Int J Mol Sci*. 2019;20(19):4910.
4. Zhang H., Yohe T., Huang L., Entwistle S., Wu P., Yang Z., Busk P., Xu W., Yin W. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95-W101.
5. Brettin T., Davis J.J., Disz T. et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep*. 2015;5:8365.

## Chromosome evolution in Ruminantia

Proskuryakova A.A.

*Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*  
*andrena@mcb.nsc.ru*

In the order of Cetartiodactyla the suborder Ruminantia is represented by mammals that typically possess horns, reduced incisors, and a multi-chambered stomach. It includes a lot of domestic livestock and rare hoofed animals. Within the ruminants there are two infraorders: Pecora and the Tragulina. Pecora includes pronghorns, bovinds, deers, giraffes and musk deers, while Tragulina represented nowadays entirely by the chevrotains.

Genome evolution in Ruminantia is characterized by two features: the wide among mammals spans of diploid chromosome numbers ranging from 7 to 70 and non-standard X-chromosome morphology. We have comprehensively studied genomes of representatives from each family in Ruminantia by a wide range of complex approaches: classical and banding cytogenetics, BAC clone's and comparative chromosome mapping, genome sequencing. Major trends in genome evolution of each ruminant family were identified. By dense BAC mapping of the X chromosome across Ruminantia we have identified three major conserved synteny blocks and reconstructed the structure of putative ancestral cetartiodactyl X chromosome. We have demonstrated that intrachromosomal rearrangements such as inversions and centromere repositions are main drivers of remarkably sped up cetartiodactyl's chromosome X evolution. Localization of ribosomal DNA on a range of species allowed us to illustrate dynamics of the distribution of these repeats on ruminant chromosomes. Comprehensive mapping data allowed improving Ruminantian and Pecoran ancestral karyotype reconstructions: to identify the order of conservative segments and positions of evolutionary breakpoints. Overall, we have characterized major trends in the genome evolution of Cetartiodactyla that can now be deeply studied by genome sequence analysis.

*Acknowledgements:* The work was supported by the Russian Science Foundation (RSF, 19-14-00034).

## Noncanonical prokaryotic X family DNA polymerases

Prostova M.<sup>1</sup>, Shilkin E.<sup>1</sup>, Kulikova A.<sup>1</sup>, Makarova A.<sup>1</sup>, Ryazansky S.<sup>1,2</sup>,  
Kulbachinskiy A.<sup>1</sup>

<sup>1</sup>*Institute of Molecular Genetics, National Research Centre “Kurchatov Institute”, Moscow, Russia*

<sup>2</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

**Key words:** specialized DNA polymerases, prokaryotic PolX, exonuclease, PHP domain, nonhomologous end-joining

DNA polymerases of the X family (PolX) play essential functions in eukaryotic DNA repair. PolX polymerases are also found in prokaryotes but their diversity and functions remain largely unstudied. We have performed a comprehensive bioinformatic analysis of prokaryotic PolXs and found that they are widespread in bacteria and are also present in several archaeal lineages. Most polymerases share a common architecture, with the central polymerase part consisting of the thumb, palm and fingers domains, preceded by an N-terminal dRP-lyase domain and followed by a C-terminal exonuclease PHP domain. Unexpectedly, we have revealed a previously unrecognized group of noncanonical PolXs that seem to be lacking DNA polymerase activity. They contain substitutions of key catalytic residues and deletions in their polymerase and dNTP binding sites in the palm and fingers domains, but contain a functional exonuclease domain. Biochemical experiments with representative noncanonical PolXs from the *Deinococcus* genus confirmed that they are indeed inactive as DNA polymerases but are highly efficient 3'-5' exonucleases. The biological function of altered polymerases is possibly connected with double-strand break DNA repair, since both canonical and altered PolXs are often encoded together with the components of the non-homologous end joining pathway, suggesting an evolutionary conservation of this PolX function. This is a remarkable example of polymerases that have lost their main polymerase activity, but may be involved in the nuclease processing of DNA repair intermediates.

# The cytogenetic map of the Nile crocodile (*Crocodylus niloticus*, Crocodylidae, Reptilia) with *in situ* localization of major tandemly arranged repetitive DNAs

Romanenko S.A.<sup>1\*</sup>, Proskuryakova A.A.<sup>1</sup>, Prokopov D.Yu.<sup>1</sup>, Davletshina G.I.<sup>1,2</sup>, Pereira J.C.<sup>3,4</sup>, Ferguson-Smith M.A.<sup>4</sup>, Trifonov V.A.<sup>1,5</sup>

<sup>1</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Animal and Veterinary Research Centre, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

<sup>4</sup> Cambridge Resource Center for Comparative Genomics, Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

<sup>5</sup> Novosibirsk State University, Novosibirsk, Russia

\* [rosa@mcb.nsc.ru](mailto:rosa@mcb.nsc.ru)

**Key words:** banding, chromosome, genomics, karyotype, repeats, reptile, whole-genome sequencing

**Motivation and Aim:** Among other reptiles, Crocodylians demonstrate small diploid chromosome numbers (30–42), little interspecific variation of the chromosome arm number, and the absence of dot-shaped microchromosomes [1]. The Nile crocodile (*Crocodylus niloticus*,  $2n = 32$ ) karyotype described in 1967 [2] is composed of 20 meta-, six submeta-, and six telocentric chromosome and has a pair of NOR-bearing chromosomes. Since then, no study of the karyotype of this species has been carried out. **Methods and Algorithms:** We applied cytogenetic approaches (CBG- [3], GTG- [4], and CDAG-banding [5]) for detailed *C. niloticus* karyotype description. Using low-coverage whole-genome sequencing data and graph-based clustering approach [5] we identified nine prevailing tandemly arranged repetitive elements in the genome of *C. niloticus* and together with 18S/28S-rDNA [6] and telomeric DNA probe [7] applied them as probes for molecular cytogenetic studies.

**Results:** CBG-banding revealed blocks of pericentromeric heterochromatin on all pairs of chromosomes. Less intensely stained interstitial heterochromatic blocks were found in all chromosomes, except for pairs 4, 7, 9 and 15. Intensive CMA<sub>3</sub>-positive pericentromeric blocks were seen in all *C. niloticus* chromosomes except pair 2. Chromosomes 6 and 11 carry large bright CMA<sub>3</sub>-positive blocks in p-arms. Weak DAPI-positive blocks were revealed in pericentromeric regions of chromosomes 1 and 2. The 18S/28S-rDNA probe gave the only interstitial signal on the pair 11. Telomeric repeat (TTTAGG)<sub>n</sub> marked distal parts of all chromosomal arms and did not demonstrate any interstitial signals. Eight tandemly arranged repeats demonstrated a chromosome-specific cluster organization and were localized mainly in the pericentromeric regions of chromosomes. Cluster 48 contains a dispersed repeat, which forms fairly clear blocks during hybridization in distal region of q-arm of chromosome 1 and in centromeric regions of chromosomes 6, 9, and 11.

**Conclusion:** An initial physical chromosome map of the Nile crocodile genome based on molecular markers will facilitate further genomic studies in reptile.

**Acknowledgements:** The work was supported by RFBR grant No. 19-54-26017.

## References

1. Srikulnath K. et al. Role of chromosome changes in crocodylus evolution and diversity. *Genomics Inform.* 2015;13(4):102-111.
2. Cohen M.M., Clark H.F. The somatic chromosomes of five crocodylian species. *Cytogenetics.* 1967;6(3):193-203.
3. Sumner A.T. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 1972;75:304-306.
4. Seabright M. A rapid banding technique for human chromosomes. *Lancet.* 1971;2:971-972.
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. *Chromosome Res.* 2018;26:307-315.
6. 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;246:519-527.
7. Ijdo J.W. et al. Improved telomere detection using a telomere repeat probe (TTAGGG)<sub>n</sub> generated by PCR. *Nucleic Acids Res.* 1991;19(17):4780.
8. Novák P. et al. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics.* 2010;11:378.



## Genetic variety of Baikal endemic amphipods (Crustacea: Amphipoda) *Eulimnogammarus verrucosus* (Gerstf., 1858) in the Angara river

Saranchina A.<sup>1\*</sup>, Drozdova P.<sup>1,2</sup>, Rzhechitskiy Y.<sup>1</sup>, Mutin A.<sup>1</sup>, Timofeyev M.<sup>1,2</sup>

<sup>1</sup> Institute of Biology, Irkutsk State University, Irkutsk, Russia

<sup>2</sup> Baikal Research Centre, Irkutsk, Russia

\* alexandra147801@gmail.com

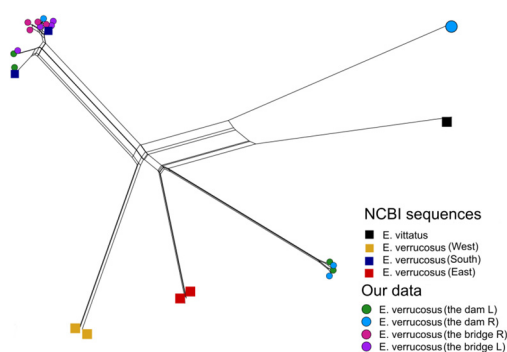
**Key words:** Amphipods, Baikal, Angara, phylogenetics, COI

Lake Baikal is a speciation biodiversity hotspot for several taxa including amphipods. Over 350 amphipod species and subspecies have been described to this moment, and this number keeps growing. *Eulimnogammarus verrucosus* is widespread along the Baikal shoreline and a popular model object. However, the separation of the species into three lineages (western, southern, and eastern) was only discovered 150 years after the species description. This diversity was determined based on the fragment of the cytochrome C oxidase (COI) gene [1].

Angara is the single outflow of Baikal and the apparent border between the western and southern *E. verrucosus*. They inhabit the head of the river at about 1-km distance. Further along Angara, there are two bridges (can promote migration) and one dam (a barrier for expansion). Study of *E. verrucosus* genetic origin in this area will uncover if there is hybridization between the lineages.

Four animals similar to *E. verrucosus* were sampled from each of four spots: on both coasts before the dam and after the second bridge (the most distant bridge from the river head). The COI sequences from these animals were sequenced and aligned with the UGENE [2] program. Then, a genetic network, was constructed in the SplitsTree [3].

All the samples from both coasts after the second bridge and two samples from the left coast before the dam belonged to the southern haplogroup. Two animals from the left coast and tree animals from the right coast before the dam constituted a separate group. No animals of the western haplogroup were found (Fig. 1).



**Fig. 1.** Phylogenetic network based on DNA sequence comparisons of COI fragment from *E. verrucosus*

These data allow us to conclude that the southern lineage of *E. verrucosus* inhabits both coasts of the Angara river and that there is yet another lineage of this species separated from Baikal ones.

*Acknowledgements:* This study was supported by the Russian Science Foundation No. 22-14-00128.

### *References*

1. Gurkov A. et al. Indication of ongoing amphipod speciation in Lake Baikal by genetic structures within endemic species. *BMC Evol Biol.* 2019;19:138.
2. <http://ugene.net/ru/>
3. <https://uni-tuebingen.de/>

## Cryptic diversity in the *Dendrobaena schmidti* complex (Lumbricidae, Annelida)

Shekhovtsov S.V.<sup>1,2\*</sup>, Rapoport I.B.<sup>3</sup>

<sup>1</sup> Institute of the Biological Problems of the North, FEB RAS, Magadan, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Tembotov Institute of Ecology of Mountain Territories, RAS, Nalchik, Russia

\* shekhovtsov@bionet.nsc.ru

**Key words:** earthworms, cryptic taxa, *Dendrobaena schmidti*, *cox1*, ITS2

**Motivation and Aim:** Even the most well-studied earthworm species were found to contain high cryptic genetic diversity [1], which is due to the paucity of diagnostic morphological characters and considerable intraspecific variation. *Dendrobaena schmidti* (Michaelsen, 1907) is a Caucasian endemic with many described subspecies and high morphological diversity. Here we studied genetic diversity of a sample of *D. schmidti* from several populations from the Western Caucasus with pronounced variation in size and color.

**Methods and Algorithms:** *D. schmidti* individuals were collected in seven locations from the Western Caucasus. Fragments of the *cox1* gene and the ITS2 nuclear ribosomal spacer were amplified using universal primers; phylogenetic trees were constructed in MrBayes v.3.2.6. We also performed a GBS analysis of seven specimens of *D. schmidti* of different size morphs; the obtained reads were assembled using Stacks v.2.55 and analyzed in Structure v.2.3.4.

**Results:** For both genes studied, we found that our sample was split into two groups [2]. The first group included somewhat bigger (3–7.5 cm) individuals that were only slightly pigmented or totally unpigmented (when fixed by ethanol). The second group contained small (1.7–3.5 cm) specimens with dark purple pigmentation. In one of the studied locations these two groups were found in sympatry. However, there were no absolute differences either in general appearance (pigmented/unpigmented, small/big) or among diagnostic characters. Although the two groups differed in size (the majority of individuals from the first group were 5–6 cm long, and of the second one, 2–3 cm), the studied samples overlapped to a certain degree. Pigmentation, despite apparent differences, was also unreliable, since it was heavily affected by fixation of the specimens. GBS analysis suggested that the studied sample was split into two to four groups, and notably one specimen belonging to the small form was found to contain similar number of alleles of the small and large forms.

**Conclusion:** Based on the obtained data we can conclude that *D. schmidti* consists of at least two species that have identical states of diagnostic characters, but differ in general appearance. However, first generation hybrids between these cryptic species may form.

**Acknowledgements:** Supported by the Russian Foundation for Basic Research (RFBR) grant No. 20-54-56030\_Iran\_t.

### References

1. Shekhovtsov S.V. et al. Species delimitation of the *Eisenia nordenskioldi* complex (Oligochaeta, Lumbricidae) using transcriptomic data. *Front Genet.* 2020;11:1508.
2. Shekhovtsov S.V. Morphotypes and genetic diversity of *Dendrobaena schmidti* (Lumbricidae, Annelida). *Vavilov J Genet Breed.* 2020;24(1):48-54.

## The recombination between 3-domain Cry-toxin genes as a mechanism to extend the list of affected hosts

Shikov A.E.<sup>1,2\*</sup>, Malovichko Y.V.<sup>1,2</sup>, Nizhnikov A.A.<sup>1,2</sup>, Antonets K.S.<sup>1,2</sup>

<sup>1</sup> All-Russia Research Institute for Agricultural Microbiology (ARRIAM), Pushkin, St. Petersburg, Russia

<sup>2</sup> St. Petersburg State University, St. Petersburg, Russia

\* a.shikov@arriam.ru

**Keywords:** Cry toxins, *Bacillus thuringiensis*, recombination

**Motivation and Aim:** The application of *Bacillus thuringiensis* (*Bt*) both in form of spores and toxins produced by the bacterium provides a wide range of activities against different pests, including insects and nematodes. Of these pesticidal moieties, crystalliferous Cry toxins, in particular, three-domain Cry toxins represent the most effective insecticides with high host specificity. The exact mechanism driving the evolution of these toxins remains unknown, however, a hypothesis that the emergence of new Cry toxins occurs as a result of domain exchange was proposed. In this work we for the first time performed the large-scale analysis of the recombination events to reveal the role of recombination-driven domain exchange in 3-D Cry toxins' evolution.

**Methods and Algorithms:** We gathered data from several sources, namely, Genbank, IPG (identical protein database), *Bt* assemblies, and the *Bt*-nomenclature, and retrieved sequences of 3-D cry toxins using our tool developed earlier, CryProcessor [1]. We then reconstructed phylogenies based on the sequences of each domain and the whole toxin. The topologies of the phylogenies were compared afterward. Next, we used RDP (Recombination Detection Protocol) tool to reveal recombination events and filtered raw data based on the similarity between sequences and their position on phylogenetic trees. To collect the data on the specificity of the toxins we used literature and public databases and compared specificity between parents and children in recombination events. Finally, we used the ETE3 utility to reveal how recombination affects evolutionary selection.

**Results:** After filtration of the toxin sequences, 370 clusters of Cry toxin sequences were obtained. The topologies of the trees differed significantly, and it was especially evident for the third domain. We then revealed more than 70 recombination events. Importantly, the events affected each domain, however, the third domain was exchanged more frequently. Fisher test showed that recombination events lead to significant ( $p < 0.01$ ) changes in host specificity both in terms of orders and species. The analysis of evolutionary selection revealed that recombination exchanges are accompanied by the intensive selection, both positive and negative.

**Conclusion:** The results clearly support the hypothesis of the role of domain exchanges between toxins in their evolution. Not only are the recombination events widespread, but they also lead to alterations in host specificity and are subjected to evolutionary selection as well. So, they might be considered as large evolutionary events, leading to extending of the list of possible susceptible hosts.

**Acknowledgements:** This study was supported by the Russian Science Foundation (grant No. 20-76-10044).

### References

1. Shikov A.E. et al. No more tears: Mining sequencing data for novel *Bt* Cry toxins with CryProcessor. *Toxins*. 2020;12(3):204.

## Evolution of protein regulators of circadian rhythm in mammals

Shilovsky G.A.<sup>1,2</sup>, Zverkov O.A.<sup>1</sup>, Rubanov L.I.<sup>1</sup>, Seliverstov A.V.<sup>1</sup>, Lyubetsky V.A.<sup>1\*</sup>

<sup>1</sup> *A.A. Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia*

<sup>2</sup> *Faculty of Biology, Moscow State University, Moscow, Russia*

\* *lyubetsk@iitp.ru*

The Cry protein (cryptochrome) is an important target of anti-aging medicine mediating the effect of chronobiotics on lifespan, physical activity, and circadian rhythm. Chronobiotics are agents that can cause phase adjustment of the body clock. F-box and leucine-rich repeat protein 21 (Fbxl21), predominantly expressed in the hypothalamus, is associated with circadian rhythms by regulating the stability of the Cry protein, which binds to the CBD domain in Fbxl21. The same domain is present in Fbxl3 (an Fbxl21 paralog also involved in circadian rhythm regulation). The domain loss renders Fbxl21 and Fbxl3 nonfunctional. Both Cry and Fbxl3 are highly conserved in mammals. Thus, our findings of significant variation of the CBD domain in Fbxl21 certain species as well as of pseudogenization or loss of the Fbxl21 gene were unexpected. We used original programs to systemically screen such events for Fbxl21 and other protein factors of circadian rhythm regulation in mammals. Fbxl21 protein was missing in representatives of Metatheria and Prototheria as well as in certain placentals including human, gorilla, rabbit, hare, armadillo, etc. The Fbxl21 protein is shortened and lacks the domain in bison; stop codon is found within the domain in seal and elephant, and unique cysteine is replaced with tyrosine at position 364 in manatee. Interestingly, an experimental mutation at the same domain position in Fbxl3 modifies the period of biorhythms. Fbxl21 started to accumulate substantial modifications in the Homininae ancestor after the split of gibbons and great apes. Numerous independent Fbxl21 losses or pseudogenizations were demonstrated; for instance, the pseudogenization is observed only in the human and gorilla while other great apes have the functional gene. Among other things, the data obtained indicate the prevalence of positive selection on Fbxl21 in these taxa while its paralog Fbxl3 is under stabilizing selection in mammals.

# PhyloBench, a benchmark for evaluation of phylogenetic programs

Sigorskiĭ A.<sup>1\*</sup>, Efremov A.<sup>1</sup>, Penzar D.<sup>1</sup>, Karyagina A.<sup>2,3,4</sup>, Spirin S.<sup>2,5,6\*</sup>

<sup>1</sup> Faculty of Bioengineering and Bioinformatics of Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Belozersky Institute of Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia

<sup>4</sup> All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

<sup>5</sup> Scientific Research Institute for System Analysis, RAS, Moscow, Russia

<sup>6</sup> National Research University Higher School of Economics, Moscow, Russia

\* sas@belozersky.msu.ru

**Key words:** phylogenetic inference, benchmark

*Motivation and Aim:* There are a lot of algorithms and programs for phylogenetic inference based on a multiple sequence alignment. Many programs allow users to choose a number of parameters and options. Different programs often produce different results, the same holds for different options of a same program. Most published comparisons of phylogenetic methods either are based on simulated alignments or use some calculated features, as log likelihood, for the evaluation. The aim of the present work is creation of a benchmark that allows such comparison on large sets of natural orthologous protein sequences using species trees as reference trees.

*Results:* PhyloBench consists of protein sequence alignments and of reference trees for these alignments. We used sequences of evolutionary protein domains extracted from the Pfam [1] database. To select the sequences, we first fixed a taxonomic group of species and selected 60 species from this group. Then using home-made scripts we formed as many as possible orthologous groups of domains from proteins of those 60 species. The species tree was constructed starting with NCBI Taxonomy, unresolved nodes of the tree were resolved using branches of the trees inferred from all obtained orthologous groups. This procedure was repeated 12 times, for different taxonomic groups covering all major taxa of cellular organisms. As a result, the benchmark consists of 12 taxonomic sets, for each of them there are sets of alignments of 15, 30, and 45 sequences randomly selected from 60-sequence orthologous groups. Also PhyloBench includes three “Combined” sets containing 15-, 30-, and 45-sequence alignments selected from taxonomic sets and representing three domains of life (Archaea, Bacteria and Eukaryota) in equal proportions. For testing the benchmark we used comparison of inferences made with real sequence alignments of domains and those made with artificially damaged alignments. For comparison of inferred trees with reference trees we used a number of tree comparison measures and chose the measure that allows us to obtain the maximum statistical significance during the test. As a result, the Robinson–Foulds distance [2] is proved to be the best tree comparison measure. We demonstrate a statistically significant difference between results obtained from real and damaged alignments thereby confirming applicability of our benchmark. Using PhyloBench, we performed a number of comparisons of phylogenetic methods and their parameters. In particular, we confirmed recent results that alignment filtering does not

improve the accuracy of phylogenetic inference [3] and that distance methods, such as minimum evolution, are superior to maximum likelihood and maximum parsimony [4].

*Availability:* The benchmark, i.e. alignments and reference trees, is available online <http://mouse.belozersky.msu.ru/phylobench/pb.html> together with the service that allows to compare trees inferred by any user's method with trees inferred by three widely used methods: maximum parsimony, maximum likelihood and BioNJ.

*Conclusion:* PhyloBench allows to evaluate a quality of any tool that infers phylogeny from a protein sequence alignment.

*Acknowledgements:* The work was supported by the Russian Science Foundation grant No. 21-14-00135.

### *References*

1. Mistry J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419.
2. Robinson D.R., Foulds L.R. Comparison of phylogenetic trees. *Math Biosci.* 1981;53(1-2):131-147.
3. Tan G. et al. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol.* 2015;64(5):778-791.
4. Gonnet G.H. Surprising results on phylogenetic tree building methods based on molecular sequences. *BMC Bioinformatics.* 2012;13:148.

## Search for coevolution in protein interaction network of translation termination factors

Tarasov O.V.\*, Danilov L.G., Zhouravleva G.A.

*St. Petersburg State University, Department of Genetics and Biotechnology, St. Petersburg, Russia*

\* [o.tarasov@spbu.ru](mailto:o.tarasov@spbu.ru)

**Key words:** protein coevolution, translation

*Motivation and Aim:* In eukaryotes, two proteins named eRF1 and eRF3 (abbreviation for Eucaryotic Release Factor) play major role in translation termination process. Aside from providing release of a nascent peptide from a ribosome, eRFs couple translation termination with other molecular pathways and therefore have a vast network of interactions with other proteins. Although these proteins are highly conservative, their sequences may vary even at low taxonomic level, thus rising a question whether such alterations reflect any degree of coevolution.

*Methods and Algorithms:* To look for traces of coevolution, we selected a set of proteins known to interact tightly with either eRF1 or eRF3 in the *Saccharomyces cerevisiae* yeast. These proteins are the following: Sup45 (eRF1) versus Sup35, Upf1, Rli1, Dbp5, Tif35, and Tif34; Sup35 (eRF3) versus Sup45, Upf1, Upf2, Upf3, Pab1 (yeast gene names are indicated). We focused at low taxonomic level using primarily different *Saccharomyces* yeast species. The protein sequences were obtained from Genbank and from “The 1002 Yeast Genomes Project” site (<http://1002genomes.u-strasbg.fr/files/>) [1]. We aligned sequences using MUSCLE and tested the resulting alignments using CAPS2 software (<http://caps.tcd.ie/>) [2].

*Results:* The proportion of variable positions varied from 1 to 10 % for different proteins within one species and from 5 to 15 % at interspecific level, with most substitutions occurring in a small number or just in a single sample. Such level of conservativeness imply that most sites within one protein or a pair of proteins might be not coevolving. Results of CAPS2 analysis supported this assumption: we found no significant correlation between amino acid substitutions in any sites. However, we are proceeding the analysis using a larger taxonomic sample.

*Acknowledgements:* The study is supported by the RSF grant No. 18-14-00050.

### References

1. Peter J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 2018;556:339-344.
2. Fares M.A., McNally D. CAPS: coevolution analysis using protein sequences. *Bioinformatics*. 2006;22(22):2821-2822.



## Plasmid genome dynamics of sequenced strains of *Xanthomonas campestris* pv. *campestris*

Tesic S.<sup>1\*</sup>, Kyrova E.<sup>2</sup>, Ignatov A.<sup>1</sup>

<sup>1</sup> RUDN University, Moscow, Russia

<sup>2</sup> All-Russian Institute of Plant Protection, St. Petersburg, Russia

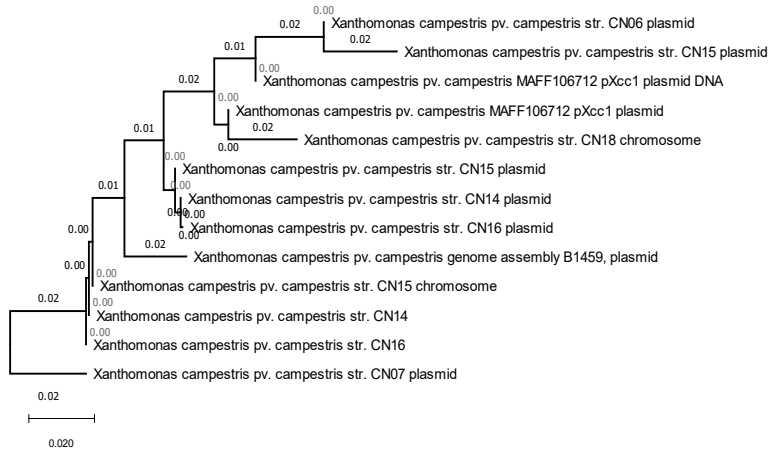
redagro9@gmail.com

**Keywords:** *Xanthomonas campestris* pv. *campestris*, plasmids, plasmid-homologues genes

*Xanthomonas campestris* pv. *campestris* (*Xcc*), the causal agent of black rot of crucifers, remains a threat to all Brassica crops around the world for almost 140 years [3, 7, 9]. Bacteria of the genus *Xanthomonas* cause diseases on over 500 plant species [6]. Recent genomic and genetic data improved our understanding of *Xcc* taxonomy and virulence mechanisms. Genome sequences of 71 *X. campestris* strains have been published by December 2020 [2]. Each has ~65 % GC content and consists of a 4.9–5.2 Mb chromosome. Some strains have with up to 3 plasmids. A survey of 46 *Xcc* isolates from a global collection that included long-read sequencing showed that 29 strains encode 26 unique transcription activator-like effector (TALE) proteins [1], a class of type III effector (T3E) proteins translocated to host plant cells via type III secretion system (T3SS). *Xcc* strains have been found to contain between zero and four TALE genes found on the chromosome or on plasmids [1]. While TALEs play important role in many *Xanthomonas* pathogens, more research is needed to determine whether they act as virulence factors for *Xcc*. *Xcc* TALE genes are often found on plasmids and in association with mobile genetic elements. Plasmids in *Xanthomonas* species were associated with bacteria's adaptation to antibiotics, heavy metals, and to resistant host plants via horizontal transfer of virulence genes, including T3SS effector protein genes. When 117 strains representing 26 pathovars were examined for plasmid content and RFLP of plasmid DNAs, all tested strains of 10 pathovars contained plasmids, while all tested strains of 13 pathovars contained no detectable plasmids. Three pathovars varied in plasmid content. The apparent stability of the plasmids provides a natural genetic marker that can be strain-specific and useful in epidemiological investigations [5].

It was known that most of epidemic *Xcc* isolates do not have plasmids, and plasmids are not stably maintained in *Xcc* cells after artificial conjugation [9]. NCBI genebank has 125 genomes of *Xcc*, and only 18 (14 %) genomes includes plasmid sequences (<https://www.ncbi.nlm.nih.gov/genome/browse/#/plasmids/151/>). Size of the plasmids vary from 867 bp in *Xcc* CFBP 11247 to 147,727 bp in CFBP 4955. The smallest know *Xcc* plasmid (Genebank acc. NZ\_CM002634.1) has transposase gene (family IS3) only. The largest known plasmid in *Xcc* CFBP 11247 (NZ\_CM002552.1) is highly similar to plasmids in *Xcc* CN05 and CN13 (99.7–99.9 for 83–87 % of DNA length) and in *X. euvesicatoria* LMG930 (plasmid pLMG930.2), *X. euvesicatoria* 85-10 (plasmid p\_XCV\_1), and *X. campestris* pv. *vesicatoria* plasmid pXCV183 with 99.68 % identity for 83 % of sequence length. Remarkably, the putative type 4 secretion systems encoded on this plasmid is similar to the Icm/Dot systems of the human pathogens *Legionella pneumophila* and *Coxiella burnetii*. DotG/IcmE/VirB10 family protein on *Xcc* CFBP 11247 plasmid has homologues in a few other *Xcc* strains (Cf4B1, Cf3C, and

GBBC 1372) only, but commonly present in most of *X. euvesicatoria* strains. The type IV secretion system (T4SS) is known to play an important role in the conjugative transfer of genetic material, and also provides fitness to pathogenic bacteria for their enhanced survival [4]. Another group of large *Xcc* plasmids represented by MAFF106712 pXcc1 has T3E gene Xcc1\_44040 (TAL group) and was found in many *Xcc* strains either as plasmid or chromosome part (Fig. 1).



**Fig. 1.** Phylogenetic relations of sequences similar to complete genome of MAFF106712 pXcc1 plasmid in other *Xcc* strains

High level of inter-strain variation mediated by plasmids harboring virulence-related genes, including the T4SS and T3SS effectors highlights the importance of strain-specific genome dynamics mediated by such elements. Majority of epidemic *Xcc* strains sequenced to the date do not have plasmids but carry plasmid-homologues genes in their chromosomes. This suggests the importance of plasmid integration into bacterial chromosome via unknown so far mechanism in providing *Xcc* strains high virulence towards commercial brassicas.

## References

1. Denancé N., Szurek B., Doyle E.L., Lauber E., Fontaine-Bodin L., Carrère S., Guy E., Hajri A., Cerutti A., Boureau T., Poussier S., Arlat M., Bogdanove A.J., Noël L.D. Two ancestral genes shaped the *Xanthomonas campestris* TAL effector gene repertoire. *New Phytol.* 2018;219:391-407.
2. Dubrow Z.E., Bogdanove A.J. Genomic insights advance the fight against black rot of crucifers. *J Gen Plant Pathol.* 2021;87:127-136. doi: 10.1007/s10327-021-00987-x.
3. Garman H. A bacterial disease of cabbage. *Agric Sci.* 1892;6(7):309-312.
4. Kaur A., Bansal K., Kumar S., Sonti R.V., Patil P.B. Complete genome dynamics of a dominant-lineage strain of *Xanthomonas oryzae* pv. *oryzae* harbouring a novel plasmid encoding a type IV secretion system. *Access Microbiol.* 2019;1(9):e000063.
5. Lazo G.R., Gabriel D.W. Conservation of plasmid DNA sequences and pathovar identification of strains of *Xanthomonas campestris*. *Phytopathology.* 1987;77(3):448-453.
6. Leyns F., De Cleene M., Swings J.G., De Ley J. The host range of the genus *Xanthomonas*. *Bot Rev.* 1984;50(3):308-356.
7. Pammel L.H. Bacteriosis of rutabaga (*Bacillus campestris* n sp.). *Am Mon Microsc J.* 1895;16:145-151.
8. Poplawsky A.R., Urban S.C., Chun W. Biological role of xanthomonadin pigments in *Xanthomonas campestris* pv. *campestris*. *Appl Environ Microbiol.* 2000;66(12):5123-5127.
9. Smith E.F. The black rot of the cabbage. Washington, D.C.: U.S. Government Printing Office, 1898.

# Composition of sex chromosomes of veiled chameleon (*Chamaeleo calypttratus*, Iguania, Squamata) reveals new insights into sex chromosome evolution of iguanian lizards

Tishakova K.<sup>1,2\*</sup>, Prokopov D.<sup>2</sup>, Davletshina G.<sup>2,3</sup>, O'Brien P.<sup>4</sup>, Ferguson-Smith M.<sup>4</sup>, Giovannotti M.<sup>5</sup>, Trifonov V.<sup>1,2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup> Cambridge Resource Centre for Comparative Genomics, Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

<sup>5</sup> Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy

\* k.tishakova@g.nsu.ru

**Key words:** sex chromosomes evolution, squamate reptile, homologous synteny blocks

**Motivation and Aim:** The genetic sex determination is widespread among vertebrate animals and has evolved multiple times in different clades. It is assumed that some homologous synteny blocks (HSBs) were involved in the formation of sex chromosome more often than others due to their genetic content. Squamate reptiles represent interesting model organisms to study sex chromosome evolution, because they demonstrate a striking diversity of sex determination systems. In the present work, we investigated the chromosome composition of the veiled chameleon (*Chamaeleo calypttratus*, CCA henceforth) to identify sex chromosomes and study their genetic content.

**Methods and Algorithms:** We obtained flow-sorted DNA-libraries of chromosomes for *C. calypttratus* and sequenced them using Illumina MiSeq platform. The reads were aligned to the genomes of reference species, such as the green anolis, *Anolis carolinensis* (ACA) and the plateau fence lizard, *Sceloporus thristichus* (STR). A pair of sex chromosomes were identified using PCR-assisted mapping with male-specific RAD-seq markers.

**Results:** We identified CCA5 as a pair of sex chromosomes. NGS data analysis demonstrated homology between CCA5 and ACA5+ACAX. ACA5 is described in the sex chromosome of pygopodid geckos and skinks, whereas ACAX form the sex chromosomes of most Pleurodont iguanas, soft-shelled turtles and some geckos. These HSBs contain at least three genes, PITX2, SOX5 and SOX10, which may be promising candidates for the role of master sex-determination gene.

**Conclusion:** Thus, the sex chromosomes of *C. calypttratus* contain the homologous synteny blocks, that are involved in the formation of the sex chromosomes of many reptiles.

**Acknowledgements:** The study is supported by the research grant No. 19-54-26017 from the Russian Foundation for Basic Research, the research grant No. 2019-0546 (FSUS-2020-0040) from the Ministry of Science and Higher Education (Russia) via the Novosibirsk State University.

# The use of next-generation sequencing for genetic characterization of tick-borne encephalitis virus variants from Western Siberia

Tkachev S.E.<sup>1,2\*</sup>, Shaikhutdinov N.M.<sup>1,3</sup>, Shigapova L.Kh.<sup>1</sup>, Chicherina G.S.<sup>4</sup>, Yakimenko V.V.<sup>5</sup>, Shagimardanova E.I.<sup>1</sup>

<sup>1</sup> *Research Center "Regulatory Genomics" of Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>2</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Skolkovo Institute of Science and Technology, Moscow, Russia*

<sup>4</sup> *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

<sup>5</sup> *Omsk Research Institute of Natural Focal Infections, Omsk, Russia*

\* *sergey.e.tkachev@gmail.com*

**Key words:** tick-borne encephalitis virus, Western Siberia, subtypes, lineages

*Introduction:* Tick-borne encephalitis virus (TBEV), a member of the *Flaviviridae* family, is the causative agent of a severe disease of the human central nervous system. Currently, in accordance with classification TBEV is divided into three subtypes: Far Eastern, Siberian (TBEV-Sib) and European [1]. Until recently, three genetic lineages were described for TBEV-Sib: Zausaev, Vasilchenko, and Baltic; subsequent studies showed that there are at least two more lineages, Obskaya and Bosnia [2, 3].

In the majority of TBEV collections genetic studies, sequencing of only genome fragments is used, which limits the use of the data obtained to assess genetic diversity, the distribution of certain virus variants, phylogeography, as well as the patterns that determine the evolution of their genomes, and only complete genome sequences provide the necessary information. The use of "classical" approaches based on Sanger sequencing for this is inefficient and time consuming. The solution to this problem could be the use of high-throughput sequencing (next-generation sequencing, NGS) to analyze large sets of TBEV strains samples from virus collections.

Currently, only about 250 TBEV complete genome sequences are presented in the GenBank databases (<https://www.ncbi.nlm.nih.gov/nucleotide/>) (and only about 30 of them for TBEV-Sib), so the task of replenishing data on the genomes of various TBEV genetic variants is a crucial task.

*Aim of the study:* The aim of this work was to develop approaches to high-throughput sequencing of TBEV strains, obtain TBEV complete genome sequences of strains from different regions of Western Siberia, and perform molecular genetic analysis of the obtained data in comparison with the available TBEV sequences from the GenBank database.

*Materials and Methods:* For the study, 85 TBEV strains previously isolated by bioassay method from various natural samples (different ticks and small mammals species) in various regions of Western Siberia (Novosibirsk oblast, Omsk oblast, Khanty-Mansiysk Autonomous District, the Altai Republic of the Russian Federation, as well as Kazakhstan) were taken. From the isolated genomic RNAs, genomic libraries were constructed and sequenced using Miseq (Illumina). After assessing the quality of the

data and trimming, genomes were assembled either *de novo* (with sufficiently high coverage) or using the data mapping method with published reference TBEV genomes (with low coverage). For further analysis, TBEV complete genome sequences from the GenBank database corresponding to various subtypes and genetic lineages of the virus were used.

*Results:* Analysis of determined complete genome sequences of TBEV strains isolated in various regions of Western Siberia revealed predominantly TBEV-Sib strains, with single findings of TBEV of the Far Eastern subtype in the Altai Republic and the Novosibirsk Region, and of the European subtype in the Omsk Region. Zausaev lineage was found to be the most represented among the studied strains, and the Vasilchenko lineage was represented to a lesser extent, which correlates with our earlier data. Also, strains of the Baltic lineage were discovered and characterized in the Novosibirsk and Omsk oblasts, as well as, for the first time, in Kazakhstan. No representatives of the Obskaya lineage previously described in the indicated territory were found, which, as previously shown, is most genetically different from other TBEV-Sib lineages. Analysis of the chronogram demonstrated that within the TBEV-Sib, the Obskaya lineage diverged from the common ancestor the earliest, after that the Bosnia lineage was separated, then the Baltic lineage, and the Zausaev and Vasilchenko lineages diverged most recently.

*Acknowledgements:* The study was part of Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by the Pfizer Quality Improvement project No. 65238411 "Optimization of methods for mass complete genome sequencing of tick-borne encephalitis virus strains".

### *References*

1. Virus taxonomy: classification and nomenclature of viruses. In: King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J. (Eds.). Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press, 2012.
2. Tkachev S.E. et al. New genetic lineage within the Siberian subtype of tick-borne encephalitis virus found in Western Siberia, Russia. *Infect Genet Evol.* 2017;56:36-43. doi: 10.1016/j.meegid.2017.10.020.
3. Tkachev S.E. et al. Genetic diversity and geographical distribution of the Siberian subtype of the tick-borne encephalitis virus. *Ticks Tick Borne Dis.* 2020;11(2):101327. doi: 10.1016/j.ttbdis.2019.101327.4.

## Evolution of the YUCCA protein family: phylogenetic relationships between the YUCCA and Type IIb FMO families

Turnaev I.I.<sup>1,2\*</sup>, Gunbin K.V.<sup>1</sup>, Suslov V.V.<sup>1</sup>, Akberdin I.R.<sup>1,3,4</sup>, Kolchanov N.A.<sup>1,2,4</sup>, Afonnikov D.A.<sup>1,2,4</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>4</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [turn@bionet.nsc.ru](mailto:turn@bionet.nsc.ru)

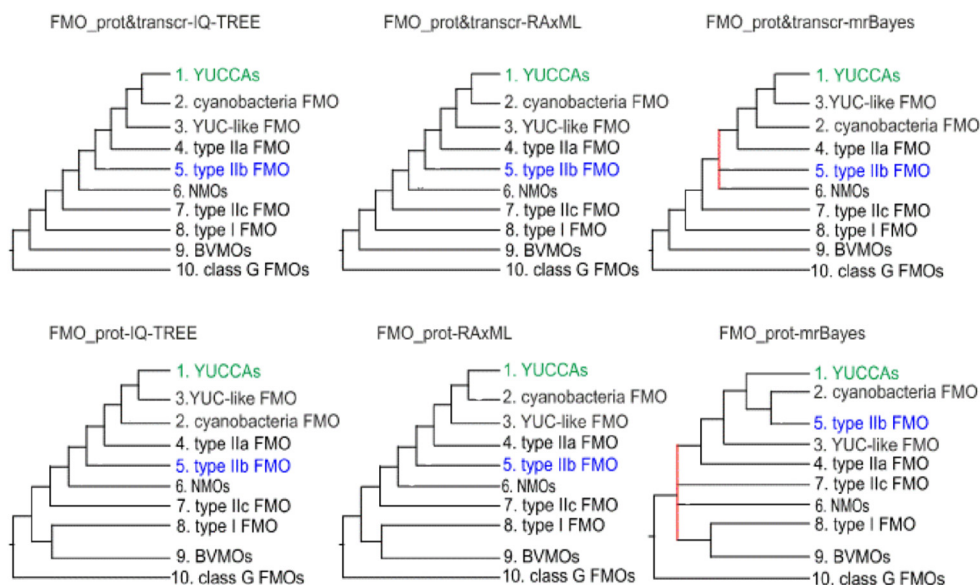
**Key words:** Auxin biosynthesis, YUCCA, flavin-dependent monooxygenase, phylogeny, charophytes, horizontal gene transfer

**Motivation and Aim:** YUCCA is the second of two enzymes in the auxin biosynthesis pathway (TAA/YUCCA) in terrestrial plants. The origin of the YUCCA family is the subject of intense debate: YUCCAs appeared as a result of horizontal gene transfer (HGT) from bacteria to the most recent common ancestor (MRCA) of land plants [2] or to MRSA of charophyte algae and land plants [3]. This necessitated an in-depth comparative analysis of the YUCCA proteins and their closest protein families.

**Methods and Algorithms:** Two samples of YUCCA homologues were created. The first sample includes only protein sequences. Protein sequences homologous to YUCCA were recognized in db PlazaPlant 2.5, NCBI, and the *Picea abies* genome project (Spruce Genome Project) using BlasP. Proteins from Riebel et al., [1], namely the seven amino acid sequences FMO-A to FMO-G of *Rhodococcus jostii* RHA1 bacteria and FMO-X of *Stenotrophomonas maltophilia* bacteria were added. The second sample includes protein sequences from the first sample supplemented with transcriptome sequences. The transcriptome sequences of YUCCA homologues were recognized among the sequences of the 1000 Plant Genomes Project and Green Algal Tree of Life Project using the BlasP program. Multiple sequence alignment was performed using the Promals program. IQ-TREE 1.6.12, RAxML 8.1.24, and mrBayes 3.2.7 were used to reconstruct the phylogenetic tree.

**Results:** Phylogenetic trees for an extended sample including both protein and transcriptome sequences of YUCCA homologues were reconstructed by IQ-TREE (Fig. 1. FMO\_prot&transcr-IQ-TREE tree cladogram), RAxML (Fig. 1. FMO\_prot&transcr-RAxML) and mrBayes (Fig. 1. FMO\_prot&transcr-RaxML). Phylogenetic trees from a sample of YUCCA homolog protein sequences were reconstructed using IQ-TREE (Fig. 1. FMO\_prot-IQ-TREE), RAxML (Fig. 1. FMO\_prot-RAxML) and mrBayes (see Fig. 1, FMO\_prot-RaxML) programs (see Fig. 1). Class G flavoprotein monooxygenase proteins were used as an outgroup for all trees. In all 6 phylogenetic trees (see Fig. 1), 10 groups of protein or mRNA sequences (excluding the outgroup class G FMOs, a total of 9 groups) belonging to three subclasses of class B FMO are well distinguished. 1) FMO subclass, includes groups: YUCCAs, only plant sequences are represented; YUC-like FMOs (only bacteria); cyanobacterial FMOs (cyanobacteria); FMO type IIa and FMO type IIc (both bacteria only); FMO type IIb includes bacteria, plants, fungi and protists; FMO type I (bacteria, protists, plants and

animals sequences). 2) Subclass NMO contains one NMO group (bacteria and fungi). 3) Subclass, BVMOs (fungi, protists and bacteria).



**Fig. 1.** Simplified cladograms for class B flavoprotein monooxygenases proteins. The type IIb FMO and YUCCAs families are highlighted in blue and green text respectively. Red line highlighted polyfurcations

The type IIb FMO clade (highlighted in blue text in Fig. 1) also contains sequences of plant organisms: four sequences of the lycophyte *Selaginella moellendorffii* and a YUCCA homologue from the charophyte alga *Klebsormidium nitens*, GAQ82387.1.

The type IIb FMO clade on all trees (see Fig. 1) except the FMO\_prot-mrBayes tree is not sister to the YUCCA clade (highlighted in green text in fig. 1). The type IIb FMO clade in the FMO\_prot-mrBayes tree is not uniquely defined as sister to the YUCCA clade. In addition, given the tetrafurcation in the FMO\_prot-mrBayes tree, its phylogeny is much less reliable than that of the other 5 trees to Fig. 1. Additional analysis of the domain structure, active sites, and distribution in taxa also showed a significant difference between the clade type IIb FMO and YUCCA proteins. Thus, our results support the hypothesis of Yue et al. [2] on the origin of YUCCA as a result of horizontal transfer of the ancestral FMO gene from bacteria to MRSA of all land plants and does not support the hypothesis of Wang et al. [2] about an earlier origin of YUCCA in charophytic algae.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### References

1. Riebel A., de Gonzalo G., Fraaije M.W. Expanding the biocatalytic toolbox of flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. *J Mol Catal B Enzym.* 2013;88:20-25.
2. Yue J., Hu X., Huang J. Origin of plant auxin biosynthesis. *Trends Plant Sci.* 2014;19:764-770.
3. Wang C., Liu Y., Li S.-H., Guan-Zhu Han G.-Z. Origin of plant auxin biosynthesis in charophyte algae. *Trends Plant Sci.* 2014;19:741-743.

## Classification of prokaryotic DNA methyltransferases by sequence similarity of the catalytic domain

Volobueva M.<sup>1,2\*</sup>, Rusinov I.<sup>1</sup>, Karyagina A.<sup>1,3,4</sup>, Spirin S.<sup>1,2,5</sup>, Alexeevski A.<sup>1,2,5</sup>

<sup>1</sup> *Belozersky Institute of Physical and Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *Faculty of Bioengineering and Bioinformatics, Lomonosov MSU, Moscow, Russia*

<sup>3</sup> *Gamaleya National Research Center for Epidemiology and Microbiology, Moscow, Russia*

<sup>4</sup> *Institute of Agricultural Biotechnology, RAS, Moscow, Russia*

<sup>5</sup> *Scientific Research Institute for System Analysis, RAS, Moscow, Russia*

\* *mashila6799@gmail.com*

**Key words:** DNA methyltransferase, classification, restriction-modification system

*Motivation and Aim:* DNA methylation is widely used by prokaryotic organisms in functioning of restriction-modification systems, and also for genome repair, regulation of gene expression and many other purposes. DNA methyltransferases (MTases) have been classified by methylation type (m<sup>5</sup>C, m<sup>6</sup>N, and m<sup>4</sup>C) [1], and the order of conservative motifs in amino acid (aa) sequences [2]. In REBASE (<http://rebase.neb.com>) MTases are also classified by the R-M system types (Type I, Type II, Type III, and their subtypes). Universal aa sequence similarity-based classification of DNA MTases is still missing, but such classification could be useful for computational detection and annotation of MTase genes in newly sequenced genomes and metagenomes. The main aim of this work is the classification of all known prokaryotic MTases by sequence similarity of the catalytic domain.

*Methods and Algorithms:* All currently available 230873 annotated MTase sequences were obtained from REBASE (ver. 108), which are components of restrictionmodification systems, as well as solitary MTases. We use HMM profiles from Pfam (<https://pfam.xfam.org>, ver. 34.0) and SUPFAM (<https://supfam.org/SUPERFAMILY>, ver. 1.75) to determine domain architectures of the MTases with *hmmsearch* (e-value < 0.01). HMM profiles that correspond to catalytic MT domains were selected.

*Results:* We propose a new classification of MTases based on HMM profiles from the Pfam and SUPFAM databases. The classification consists of two levels. The first level is the classification according to the sequence similarity and proposed homology of catalytic domains of these enzymes. At the second level, a class of the first level is divided into groups corresponding to domain architectures, i.e. sequences of all Pfam domains in sequence. Thus, one group includes proteins that have similar sequences along their entire length. In total, we selected eight HMM profiles from the Pfam database and 12 HMM profiles from the SUPFAM database that locate the catalytic domain in the MTase sequence. On the basis of known 3D structures of MTase classes' representatives, we identified in HMM profiles regions that correspond to SAM-binding and catalytic subdomains and motifs. We compared and contrasted the HMM profiles of Pfam and SUPFAM catalytic domains (Table 1).



**Table 1.** Comparison of Pfam (rows) and SUPFAM (columns) profiles. The cells list the numbers of proteins with overlapping Pfam and SUPFAM profiles. The high enough values are marked in bold. Only sequences having both Pfam and SUPFAM hits are included

Family abbreviation	SF-DAM		SF-DCM			SF-N6_DNA			SF-TaqI	SF-TypeII		
Model	48856	52484	45633	46303	46923	51816	52618	54378	53087	36976	37952	45988
DNA_methylase	1	287	<b>4096</b>	<b>23667</b>	<b>13198</b>	2	21	214	37	0	3	64
Dam	0	3	5	60	8	0	0	0	0	1	0	0
Eco57I	0	0	1	112	111	4	49	<b>6086</b>	285	0	0	0
EcoRI_methylase	0	0	0	1	3	0	0	0	0	0	0	0
MT-A70	0	0	0	5	0	0	1	0	4	0	0	0
MethyltransfD12	<b>3533</b>	<b>31140</b>	49	181	57	0	3	0	0	0	0	14
N6_Mtase	0	0	2	16	9	<b>42947</b>	<b>208</b>	<b>7790</b>	52	0	0	0
N6_N4_Mtase	7	2	3	36	29	0	0	1	0	<b>3699</b>	<b>1807</b>	<b>39563</b>

Based on a comparison of overlaps of matches of HMM profiles containing full SAMbinding and catalytic subdomains, all 20 profiles were divided into the following groups: i) DNA\_methylase, SF-DCM; ii) MethyltransfD12, SF-DAM; iii) N6\_N4\_Mtase, SFTypeII; iv) N6\_Mtase, Eco57I, SF-N6\_DNA, SF-TaqI; v) Dam; vi) MT-A70; vii) EcoRI\_methylase. At least one of the above profiles is found in 90 % of MTases from REBASE.

*Conclusion:* The MTases were classified according to sequence similarity. The developed classification significantly fills the gaps of the previous classifications.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant No. 21-14-00135).

### References

1. Wilson G.G., Murray N.E. Restriction and modification systems. *Annu Rev Genet.* 1991;25:585-627.
2. Bujnicki J.M. Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol Biol.* 2002;2:3.

## Universal microsatellite markers for Baikal endemic sponge family Lubomirskiidae

Yakhnenko A.<sup>1,2\*</sup>, Itskovich V.<sup>1</sup>

<sup>1</sup>Limnological Institute, SB RAS, Irkutsk, Russia

<sup>2</sup>Joint Institute for Nuclear Research, Dubna, Russia

\*yakhnenkoas@gmail.com

**Key words:** Baikal sponge, microsatellite markers, population genetics

**Motivation and Aim:** Lake Baikal is a unique natural ecosystem with a high level of endemism. Baikal endemic sponges are representatives of a bunch of closely related species of the family Lubomirskiidae. Due to the close relationship between species, some molecular genetic markers are universal for most species of this family, for example classic barcoding marker COI. Despite great progress in studying the evolution and phylogeny of the Baikal endemic sponges, the population structures of the species have not been studied at all. The study of the population structure is also of interest in connection with the events of mass mortality and disease of sponges that have been identified in the last decade on Lake Baikal. The most suitable molecular genetic markers for studying the population structure of sponges are microsatellites. There are no developed sets of microsatellite markers for Baikal endemic sponges. For other freshwater sponges, only a few studies of the population structure for members of the Spongillidae family have been published.

**Methods and Algorithms:** In this work, we tested the universality of a set of 28 microsatellites developed based on the complete genome of *Ephydatia muelleri* and the draft genome of *Lubomirskia baikalensis* among representatives of *Ephydatia muelleri*, *Lubomirskia abietina*, *Baikalospongia bacillifera*, *Swartschewskia papyracea* and *Rezinkovia echinata* species.

For all species at least three samples were tested. DNA from sponge samples was isolated using the CTAB method. PCR was performed in a Peltier Thermal Cycler using a ScreenMix-HS kit. For *E.muelleri* microsatellite markers specificity and the exact size of PCR products were estimated using fragment analysis on an ABI 3130xl Genetic Analyzer and analyzed with the GenMarker 3.01 software. For other species marker specificity was tested by PCR products visualization in a 2 % agarose gel with Syber Green dye.

**Results:** For *E.muelleri* 17 loci specifically amplified and gave clear bands in agarose gel electrophoresis. Of 17 loci 11 were selected based on the results of fragment analysis. There were eight dinucleotide microsatellites, two trinucleotide microsatellites and one interrupted repeat. All microsatellite sequences can be found in *Ephydatia muelleri* genome data [1] in the following positions: scaffold\_0019 – 1357021-1357338, 1810514-1810717, 1386901-1387195, 6087901-6087619, 1043658-1043935; scaffold\_0016 – 868299-867998; scaffold\_0020 – 6431917-6432205; scaffold\_0012 – 8003328-8003132; scaffold\_0023 – 541061-540749, 2188522-2188841; scaffold\_0018 – 7273471-7273864. For *Lubomirskia abietina*, *Baikalospongia bacillifera*, *Swartschewskia papyracea* and *Rezinkovia echinata* species same 17 loci from 28 tested were successfully amplified and gave clear bands on gel electrophoresis. All samples of

Lubomirskiidae species will be further tested for these 17 loci by fragment analysis for exact microsatellite marker sets determination.

*Conclusion:* A set of 11 microsatellite markers for *E.muelleri* was successfully developed. Specific amplification of 17 out of 28 microsatellite loci for representatives of every genus from Lubomirskiidae family gives strong evidence for high cross-species specificity of microsatellites in freshwater sponges. For microsatellite marker sets development for all studied species more analyses will be carried out.

*Acknowledgements:* The reported study was funded by RFBR and the Government of the Irkutsk Region, project No. 20-44-383010.

### *References*

1. Kenny N.J. et al. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nat Commun.* 2020;11(1):3676.

## A low molecular weight protein from Baikal amphipod hemocytes as a potential antimicrobial peptide

Zolotovskaya E.\*, Drozdova P.

Irkutsk State University, Irkutsk, Russia

\* zolotovskayaelenad@gmail.com

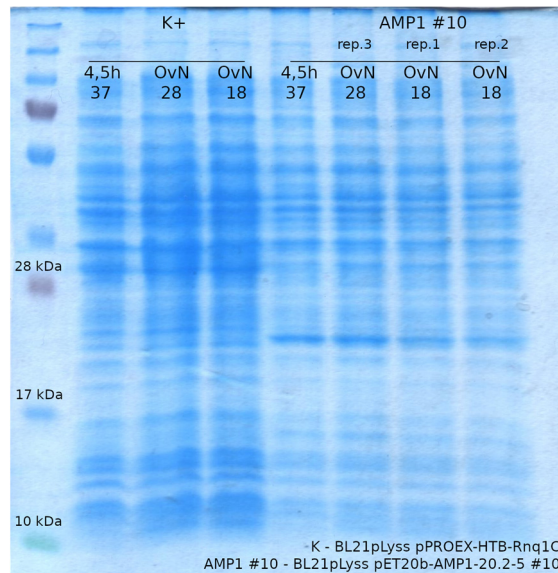
**Key words:** immune system, antimicrobial peptides, amphipods, Lake Baikal

*Motivation and Aim:* Protection against foreign bodies, provided by the immune system, is important for all organisms. One of the key immune components are antimicrobial peptides (AMPs), which are expressed by immune cells, hemocytes, into hemolymph [3]. AMPs bind to microbial membranes, causing cell lysis. Over 350 species of endemic crustaceans, amphipods, inhabiting diverse ecological niches of Lake Baikal, could have acquired unique immune components during their evolution. This work is devoted to the study of a potential antimicrobial peptide from the Baikal amphipod *Eulimnogammarus verrucosus* hemocytes.

*Methods and Algorithms:* The amphipods were caught in the littoral of Lake Baikal near the bay in Listvyanka. Hemolymph was collected from animals, followed by purification of hemocytes [5] and protein isolation [1]. Hemocyte proteome were analyzed using liquid chromatography/tandem mass spectrometry. Data analysis was carried out using SearchGUI v3.3.17 and Peptide Shaker v1.16.44 software. Protein identification was conducted using *E. verrucosus* transcriptome assembly GHHK01 [2]. The search for proteins with antibacterial properties was performed using NCBI Conserved Domain Search web tool. The results of the mass spectrometric analysis are published in Zolotovskaya et al., 2021 [4].

*Results:* We found a low molecular weight protein, which we designated AMP1. It contains an anti-lipopolysaccharide domain, interacting with the membranes of Gram-negative bacteria. To study the AMP1 properties, we designed the pET20b-AMP1-6His vector. pET20b allows for efficient gene of interest heterologous expression and also contains six histidine sequence for proteins purification by affinity chromatography. The obtained vector was used to transform *Escherichia coli* BL21 pLysS strain. To completely block AMP1 expression, cells were grown in solid LB medium with and without 1 % glucose. On the medium with glucose, the number of bacterial colonies was noticeably fewer than on medium without glucose. Protein expression in BL21 pLysS was induced by adding isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Cultivation of bacterial cells in liquid medium was carried out with and without glucose. Cells were grown at the temperature of 37, 28 or 18 °C for 24 hours, for determining the optimal protein expression conditions. During this cultivation we observed low speed of bacterial culture growth. The AMP1 expression was shown in culture without glucose, whereas there was no AMP1 expression in the culture with glucose. The highest amount of protein was formed during 24 hours at the temperature of 28 °C (Fig. 1). Low bacterial growth in liquid medium and small colony size on solid medium indicates pET20b-AMP1-6His toxicity and indirectly demonstrates antibacterial properties of the AMP1 protein.

**Conclusion:** We analyzed the first hemocyte proteome of an amphipod and found a protein that is a potential antimicrobial peptide.



**Fig. 1.** Protein expression in *E. coli*. K+ – pPROEX-HTb-Rnq1C, used as a positive control; AMP1 #10 – pET20b-AMP1-6His

### References

1. Bedulina D.S. et al. Preliminary analysis of hemocyanins in hemolymph plasma of Baikal endemic amphipods. *J Stress Physiol Biochem.* 2016;12(1):74-86.
2. Drozdova P. et al. Comparison between transcriptomic responses to short-term stress exposures of a common Holarctic and endemic Lake Baikal amphipods. *BMC Genomics.* 2019;20(1):712.
3. Rowley A.F. The immune system of Crustaceans. Encyclopedia of immunobiology. Elsevier Academic Press, 2016;1:437-453.
4. Zolotovskaya E. et al. Hemocyte proteome of the Lake Baikal endemic *Eulimnogammarus verrucosus* (Crustacea: Amphipoda) sheds light on the immune-related proteins. *Biological Communications.* 2021;66(4):290-301.
5. Shchapova E. et al. Application of PEG-Covered non-biodegradable polyelectrolyte microcapsules in the crustacean circulatory system on the example of the amphipod *Eulimnogammarus verrucosus*. *Polymers.* 2019;11(8):1246.

# 2 Symposium “Systems computational biology”



- 2.1 Section “Reconstruction, computational analysis and modeling of biological systems” [190](#)
- 2.2 Section “Modeling of population and ecological systems and processes” [237](#)

# Reconstruction and analysis of the gene network of the external pathway of apoptosis in viral hepatitis C

Adamovskaya A.<sup>1,2,3\*</sup>, Demenkov P.S.<sup>1,2</sup>, Ivanisenko V.A.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *adamovskaya@g.nsu.ru*

**Key words:** apoptosis, extrinsic pathway, hepatitis C virus, ANDSystem, gene network

Viral hepatitis C is one of the most common and dangerous infectious diseases in the world, caused by the hepatitis C virus (HCV) and affecting primarily liver cells. The infection can develop in the body for years, becoming chronic, and almost never manifest itself. Chronic hepatitis C is one of the main causes of liver cirrhosis and hepatocellular carcinoma. Existing drugs are expensive and not enough effective due to the frequent occurrence of drug resistance of the hepatitis C virus due to mutations in its genome. Penetrating into the body, HCV seeks to exercise control over the biological processes occurring in the host cells in order to increase its survival and replication efficiency. Several studies have shown that HCV has a regulatory effect on apoptosis, which plays an important role in the regulation of physiological and pathological processes.

The study of molecular-genetics mechanisms of regulation of biological processes by viral proteins can help in the search for new targets for drug exposure. The creation of antiviral drugs aimed at cellular targets can solve the problem of drug resistance in pathogens. Understanding the mechanisms of regulation of apoptosis by HCV proteins can help in the search for targets and the development of pharmacological drugs for the treatment of diseases such as hepatocellular carcinoma, neurodegenerative disorders, etc., in which the process of programmed cell death is disrupted.

The aim of the work is to reconstruct and analyze the molecular-genetics mechanisms of regulation of the external pathway of apoptosis in viral hepatitis C.

In this work, using the ANDSystem [1], the gene network for regulating the external pathway of apoptosis in viral hepatitis C and frame models were reconstructed. We described the potential regulation of the apoptosis process by viral proteins through the FADD, CASP8, and CFLAR proteins. Further, a comprehensive analysis of the associative gene network was performed.

*Acknowledgements:* Work 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.

## References

1. 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(2).

# A novel and robust molecular switch actuating the quantitative model of eukaryotic Cdk control

Barberis M.

*Systems Biology, School of Biosciences and Medicine and Centre for Mathematical and Computational Biology, CMCB, University of Surrey (UOS), Guildford, United Kingdom*

*Swammerdam Institute for Life Sciences, University of Amsterdam (UvA), Amsterdam, Netherlands*

\* [matteo@barberislab.com](mailto:matteo@barberislab.com); [m.barberis@surrey.ac.uk](mailto:m.barberis@surrey.ac.uk)

**Key words:** kinetic modeling, network motifs, cell cycle dynamics, transcriptional regulation, budding yeast

*Motivation and Aim:* The eukaryotic cell cycle is driven by waves of cyclin-dependent kinase (cyclin/Cdk) activities that rise and fall with a timely pattern called “waves of cyclins”. This pattern guarantees coordination and alternation of DNA synthesis with cell division, and its failure results in an abnormal cell proliferation. Although details about transcription of cyclins – the regulatory subunits of Cdk – are available [1], a lack of understanding exists about network designs responsible for timely waves of cyclins.

*Methods and Algorithms:* We have recently identified a transcriptional mechanism that regulates the timing of waves of mitotic cyclin (Clb) in budding yeast. This mechanism involves the evolutionarily conserved Forkhead (Fkh) transcription factors which synchronize Clb appearance. We have also shown that cyclin/Cdk-mediated positive feedback loops (PFLs) and Clb3-centered regulations sustain cyclin/Cdk autonomous oscillations. Here, through a computational and experimental strategy, we reveal a novel and robust molecular design that ensures cell cycle time keeping through interlocking transcription with cyclin/Cdk dynamics in budding yeast. Deterministic and stochastic analyses of thousand structures of a minimal kinetic model of the mitotic cyclin (Clb)/Cdk1 network are verified against *in vivo* quantitative data of Clb dynamics, for their ability to generate Clb waves.

*Results:* A novel molecular switch is unraveled that identifies a transcriptional cascade through Fkh. The Fkh-mediated cascade among Clb cyclins and Clb/Cdk1-mediated PFLs are pivotal for synchronizing Clb/Cdk1 waves. Furthermore, our model predicts a definite Fkh activation pattern underlying this design, with a progressive Clb/Cdk1-mediated Fkh phosphorylation. Experimental validation confirms the prediction, highlighting the Clb/Cdk–Fkh axis being pivotal for a timely pattern of waves of cyclins.

*Conclusion:* This work rationalizes the quantitative model of Cdk control for budding yeast, identifying regulatory motifs in place that keep a well-timed cell cycle. Altogether, our effort reveals a conserved, functional design principle in cell cycle control [3].

*Acknowledgements:* The study is supported by the UvA-Systems Biology Research Priority Area Grant and by the Systems Biology Grant of the University of Surrey.

## References

1. Haase S.B., Wittenberg C. Topology and control of the cell-cycle-regulated transcriptional circuitry. *Genetics*. 2014;196(1):65-90.
2. Barberis M. Cyclin/Forkhead-mediated coordination of cyclin waves: an autonomous oscillator rationalizing the quantitative model of Cdk control for budding yeast. *NPJ Syst Biol Appl*. 2021;7(1):48.
3. Barberis M. Quantitative model of eukaryotic Cdk control through the Forkhead CONTROLLER. *NPJ Syst Biol Appl*. 2021;7(1):28.



## **MiRNA-dependent regulation of ERBB signaling pathway genes in patients with NASH**

Bograya M.\*, Vulf M., Shunkina D., Karpeeva E., Komar A., Litvinova L.

*Center for Immunology and Cellular Biotechnology, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

\* *mbograya@mail.ru*

**Key words:** miRNA, NAFLD, GSEA, ERBB signaling

NAFLD occupies a leading position among internal diseases, and its prevalence among adult population of Russia is 37 %. The development of NAFLD is associated with obesity, type 2 diabetes, dyslipidemia, and metabolic syndrome. Recently, there is increasing evidence for the role of miRNA in the development of metabolic disorders.

In our study, we first comprehensively analyzed the miRNA profile in NAFLD. Screening of miRNAs in the blood serum of the studied patients revealed that miRNAs are differentially expressed at different stages of NAFLD. MiRWalk algorithms were used to identify genes potentially deregulated by differentially expressed miRNAs and to functionally enrich (GSEA) the obtained gene sets using the KEGG and REACTOME databases. One of the pathways yielded was the ERBB signaling pathway, which leads to the mobilization of various metabolic pathways. A total of 58 ERBB pathway genes were predicted to be targets of miRNAs that are elevated in NASH patients. Of particular interest are ERBB4 and its ligand NRG4, which are most likely targeted by miRNA-15b, miRNA-222-3p, miRNA-423-5p, and mir-23a-5p. Investigating the mechanisms of the little-studied miRNA-dependent regulation of the ERBB signaling in the liver is important for finding new participants in the pathogenesis of liver diseases and/or their diagnostic markers.

*Acknowledgements:* This research was funded by State Assignment Grant No. FZWM-2020-0010 and Grant of the President of the Russian Federation No. MK-2072.2022.3.

## Using a combination of mathematical multiphase poroelasticity model of parenchyma and experimental approaches to study hydrocephalus

Cherevko A.<sup>1\*</sup>, Valova G.<sup>1</sup>, Khe A.<sup>1</sup>, Akulov A.<sup>2</sup>, Romashchenko A.<sup>2</sup>, Bogomyakova O.<sup>3</sup>, Tulupov A.<sup>3</sup>

<sup>1</sup> Lavrentyev Institute of Hydrodynamics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> International Tomography Center, SB RAS, Novosibirsk, Russia

\* cherevko@mail.ru

**Key words:** cerebrospinal fluid, hydrocephalus, mathematical models

*Motivation and Aim:* There are a large number of pathological conditions of the central nervous system characterized by impaired movement of intracerebral fluids [1]. An important example is hydrocephalus, a pathology in which the brain ventricles increase, which leads to displacement and compression of brain tissue. This condition is well described in terms of clinical manifestations, but its causes and development are poorly understood.

*Methods and Algorithms:* The report considers a complex mathematical model of cerebral cerebrospinal fluid and hemodynamics of a person based on a model of multiphase poroelasticity for brain matter. The interaction of cerebral fluids is given by a set of four numerical coefficients. The displacement of the ventricular wall of the brain and the magnitude of pressure on it are studied [2, 4]. A multiple linear regression with interaction is constructed that allows us to quantify the effect of these coefficients on the average ventricular wall displacement. This mathematical model was tested on laboratory animals. This is necessary in order to further use experimental approaches and increase the number of real parameters involved in modeling.

*Results:* It is shown that the considered model allows to describe both the healthy state of organism and the state of organism with hydrocephaly and the transition between them that takes place when the model parameters change [3]. Based on the MRI data of real patients, the patterns of behavior of these values are determined depending on the parameters of the model in hydrocephalus. The prevailing influence of an arterial-liquor component was observed [4]. Using MRI data and a mathematical model in laboratory mice, intertrain differences in the tendency to hydrocephalus were found.

*Conclusion:* The effect of interaction parameters on mean ventricular wall displacement and periventricular pressure is described qualitatively. Based on the regression model analysis the prevailing influence of the capillary-CSF component was found. The analysis reveals the relationship between the interaction coefficients and the pathological conditions. The data used, obtained on laboratory animals, made it possible to construct an animal version of the described mathematical model.

*Acknowledgements:* The study was carried out at the expense of the grant of the Russian Science Foundation No. 22-11-00264, <https://rscf.ru/project/22-11-00264/>.

## References

1. Yankova G., Bogomyakova O., Tulupov A. The glymphatic system and meningeal lymphatics of the brain: new understanding of brain clearance. *Rev Neurosci.* 2021;32(7):693-705.
2. Yankova G.S., Cherevko A.A., Khe A.K., Bogomyakova O.B., Tulupov A.A. Study of hydrocephalus development using poroelastic models. *Appl Mech Tech Phys.* 2020;61(1):17-29.
3. Yankova G.S., Cherevko A.A., Khe A.K., Bogomyakova O.B., Tulupov A.A. Mathematical modeling of normotensive hydrocephalus at different levels of detailed brain geometry. *Appl Mech Tech Phys.* 2021;62(4):148-157.
4. Valova G., Bogomyakova O., Tulupov A., Cherevko A. Influence of interaction of cerebral fluids on ventricular deformation: A mathematical approach. *PLoS One.* 2022;17(2):e0264395.

## Development of an algorithm for facts extraction from the texts of biomedical articles

Derevianchenko S.<sup>1\*</sup>, Demenkov P.<sup>2</sup>

<sup>1</sup> Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* [sod97@yandex.ru](mailto:sod97@yandex.ru)

**Key words:** facts extraction, text mining, natural language processing

Nowadays a huge amount of new text information appears every day. This information published by billions of different sources. Today one of the most important tasks of information processing is to extract the important and useful part of it from the entire amount. A popular and useful approach is facts extraction from texts.

A fact in the text is structured information about some objects found in this text and their relationship to each other. Thus, facts extraction from texts is the task of searching for structured information in unstructured text data. Facts extraction it's one of Text Mining tasks. In this work is presented a new algorithm for solving the problem of facts extraction from texts. Main characteristics of algorithm are:

- Unsupervised learning (the algorithm is trained on unlabeled data corpora).
- Easy adaptation to new topics (the quality and efficiency of the algorithm are evaluated on texts from biomedical topics. However, it's assumed that the quality remains the same using the algorithm in a new domain).
- Low requirements for computing resources (the ability to train and inference the algorithm on medium-power computers, for example, conventional PCs).

The main idea of the algorithm is the interaction of Latent Dirichlet allocation model and parsing of sentences' dependency trees. Using of these components allows the algorithm to meet all the requirements stated above.

The presented algorithm is characterized by high quality of facts extraction and a low percentage of errors. Its evaluation metrics on the test dataset:

Type of facts	Percent of extracted facts, %	Percent of sentences with incorrect extracted facts, %
Expression regulation	67.1	4.3
Expression upregulation	73.8	5.4
Upregulation	68.4	9.6
Regulation	60.3	4.2
Transport regulation	65.2	4.5

Presented algorithm preserves its quality in other domains. This is its metrics on the Web news dataset:

Type of facts	Percent of extracted facts, %	Percent of sentences with incorrect extracted facts, %
News	62	1
Sport news	64	2

*Acknowledgements:* Work was supported by the budget project No. FWNR-2022-0020.

## Computer modeling of glioblastoma gene network and hub genes analysis

Dergilev A.I.<sup>1,2</sup>, Gubanov N.V.<sup>1</sup>, Voropaeva E.N.<sup>1</sup>, Mozyleva Y.A.<sup>3</sup>, Korobeynikova A.V.<sup>3</sup>, Shakirova R.R.<sup>3</sup>, Orlova N.G.<sup>4,5</sup>, Orlov Y.L.<sup>2,3,6\*</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>4</sup> *Financial University under the Government of the Russian Federation, Moscow, Russia*

<sup>5</sup> *Moscow State Technical University of Civil Aviation, Moscow, Russia*

<sup>6</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

\* [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** plant biology, genomics, transcription factor binding, ChIP-seq, databases, *A. thaliana*

*Motivation and Aim:* Glioblastoma is the most aggressive type of brain tumors resistant to a number of antitumor drugs. The problem of therapy and drug treatment course is complicated by extremely high heterogeneity in the benign cell populations, the random arrangement of tumor cells, and polymorphism of their nuclei. The pathogenesis of gliomas needs to be studied using modern technologies to find new therapy targets. For an object such as gliomas, it is necessary to conduct new studies based on modern cellular technologies, genome-wide technologies for high-throughput sequencing, and the integration of available information from international databases and genomic projects [2].

*Methods and Algorithms:* There are open international databases on gene expression including glioblastoma (microarrays and sequencing data of GEO NCBI, <http://www.ncbi.nlm.nih.gov/geo/profile/>), expression in various types of tumor cells (The Cancer Gene Atlas, [cancergenome.nih.gov](http://cancergenome.nih.gov/)), gene expression for brain compartments (Allen Brain Atlas), protein interactions databases such as HPRD (<http://hprd.org/>), KEGG biochemical reactions (<http://www.genome.jp/kegg/>), Interactome (<http://interactome.org/>), sequenced tumor genomes, including gliomas and glioblastomas (<https://cghub.ucsc.edu/>). The Allen Institute has developed the Ivy Glioblastoma Atlas Project database (<http://glioblastoma.alleninstitute.org/>) according to patients with glioma. We demonstrate that these publicly available online bioinformatics tools can give helpful information for annotation of gene list for glioblastoma, reconstruction of gene network and comparative analysis of the related diseases. This approach for gene network analysis of list of gene names using online bioinformatics tools presents application of bioinformatics methods to annotation of complex human diseases. We used DAVID (Database for Annotation, Visualization and Integrated Discovery) tool (<https://david.ncifcrf.gov/summary.jsp>) and the PANTHER (Protein ANalysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) resource for gene ontology analysis. Then we applied gene network reconstruction tools. Prioritization of genes was performed using the resource ToppGene: Candidate gene prioritization (<https://toppgene.cchmc.org/>).

*Results*, the most significant GO categories for glioblastoma genes are protein binding, positive regulation of protein phosphorylation, membrane location (according to cellular compartments – GOTERM\_CC classification), and kinase activity. In general, according to the categories from the table, membrane proteins and kinases play an important role, apparently associated with signal transduction into the cell. UP\_KEYWORD (terms from UniProt (<https://www.uniprot.org/>)) shows associations with tumors and kinase activity. The categories of processes related to the development of the nervous system are also presented. The hub genes for glioblastoma were identified by the Cytoscape tool (<https://cytoscape.org/>), and pathway enrichment analysis of the genes was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) following [4].

Yang and Yang (2021) recently analyzed glioma genes by GSEA, and the significantly enriched KEGG pathways involved in synapse signaling and oxytocin signaling pathways [5]. The core molecules of GBM and the DrugBank database were assessed to identify 10 drugs included tetrachlorodecaoxide related to cancer and neuropsychiatric diseases. The study of the structure of the gene network shows a high connectivity of genes and their products. Glioma progression is strongly connected with different types of epigenetic phenomena, such as histone modifications, DNA methylation, chromatin remodeling, and aberrant microRNA

*Conclusion*: Overall, online bioinformatics tools allow to model main parameters of gene network related to given disease [3]. Zhu and colleagues (Zhu et al., 2021) provided a framework of workflow for potential therapeutic drug discovery and predicted 10 potential drugs for glioblastoma therapy. Integration of the results obtained by online bioinformatics tools and genetics data leads to novel knowledge in the study of complex oncological diseases. Note recent thematic journal issues on the chronic disease markers (<https://www.frontiersin.org/research-topics/21036/high-throughput-sequencing-based-investigation-of-chronic-disease-markers-and-mechanisms>) in *Frontiers in Genetics* and related works highlighting importance of new drug targets search [1].

*Acknowledgements*: The authors are grateful to N.Y. Oparina for science discussion. This publication has been supported by the RUDN University Scientific Projects Grant System, project No. R.3-2022-ins).

## References

1. Anashkina A.A., Leberfarb E.Y., Orlov Y.L. Recent trends in cancer genomics and bioinformatics tools development. *Int J Mol Sci.* 2021;22:12146. doi: 10.3390/ijms22212146.
2. Gubanova N.V., Orlova N.G., Dergilev A.I., Oparina N.Y., Orlov Y.L. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools. *J Integr Bioinform.* 2021;18:20210031. doi: 10.1515/jib-2021-0031.
3. Orlov Y.L., Galieva A.G., Orlova N.G., Ivanova E.N., Mozyleva Y.A., Anashkina A.A. Reconstruction of gene network associated with Parkinson disease for gene targets search. *Biomeditsinskaya Khimiya.* 2021;67(3):222-230. doi: 10.18097/PBMC20216703222.
4. Li C., Pu B., Gu L., Zhang M., Shen H., Yuan Y., Liao L. Identification of key modules and hub genes in glioblastoma multiforme based on co-expression network analysis. *FEBS Open Bio.* 2021;11(3):833-850. doi: 10.1002/2211-5463.13078.
5. Yang J., Yang Q. Identification of Core Genes and Screening of Potential Targets in Glioblastoma Multiforme by Integrated Bioinformatic Analysis. *Front Oncol.* 2021;10:615976. doi: 10.3389/fonc.2020.615976.

## Computer analysis of genes associated with Kaposi's sarcoma

Elistratova M.G.<sup>1\*</sup>, Shakirova R.R.<sup>1</sup>, Orlov Y.L.<sup>1, 2, 3</sup>

<sup>1</sup> *I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\**mar.elist99@gmail.com*

**Key words:** biomedicine, Kaposi's sarcoma, disease susceptibility, bioinformatics, ontology, e-Health

*Motivation and Aim:* Kaposi's sarcoma is a cancer that develops from cells that line the lymph or blood vessels [1]. It usually appears as tumors on the skin or mucosal surfaces (inside the mouth), but these tumors can also develop in other parts of the body, such as the lymph nodes (bean-sized clusters of immune cells throughout the body), the lungs, or the digestive tract. It is important to describe new target genes for this disease. We aimed reconstruct gene network for future development of novel therapeutics or repurposing of existing treatment strategies, particularly in identifying new biomarkers as hub genes in such network.

*Methods and Algorithms:* We used open bioinformatics tools and databases following the approaches for gene network reconstruction and analysis presented in [2, 4]. The task is to build (collect) a list of genes associated with the development of Kaposi's sarcoma, analyze the categories of gene ontologies for such a list, and reconstruct the gene network. First, we constructed the list of genes associated with a hereditary predisposition to Kaposi's sarcoma. The Internet-resource OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) was used to analyze the genes. A search for the keyword Kaposi sarcoma (Kaposi's sarcoma) gave 273 genes - ACTFS, ACTG2, ADSL, ALKBH1, etc. We used also GeneCards.org platform for the annotation of sarcoma genes. Next step is estimation of enriched gene ontologies categories in the gene list and gene network reconstruction. The PANTHER resources (Protein Analysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) were used to analyze the categories of gene ontologies.

*Results:* The most significant categories for Kaposi's sarcoma genes were regulation of metabolic processes of the connection of nitrogen and macromolecules, regulation of the cellular metabolic process, regulation of cell proliferation, regulation of the process of biosynthesis and the metabolic process of RNA. Next, the categories of gene ontologies for molecular functions were calculated. The most significant categories are heterocyclic binding, organic cyclic binding, protein binding, and to a lesser extent specific DNA binding. Further, similarly, with the help of PANTHER, an ontology table for cellular components was built. The most significant categories for Kaposi's sarcoma genes are the nucleoplasm, nucleus, nuclear chromosome, chromatin, intracellular non-membrane organelles (ribosomes). Thus, PANTHER for Kaposi's sarcoma genes confirm the categories of gene ontologies of protein binding, nuclei, non-membrane organelles, metabolic processes, and RNA biosynthesis. The most significant categories for this list of genes include the regulation of metabolic processes, cell proliferation, and transcriptional disruption in cancer.

GeneMANIA (<https://genemania.org/>) and STRING-DB (<https://string-db.org/>) resources were used to reconstruct the gene network of interactions between Kaposi's sarcoma genes. In the center of the network are Kaposi's sarcoma genes that have a large number of connections with other elements – MYC, TP53, CDKN1A. From STRING-DB it can be found that the network is quite sparse, some objects (proteins) do not contact with others. A central, strongly connected cluster of genes stands out. In total, 214 genes from the list were recognized by STRING-DB tool. The network has a non-randomly large number of connections (at level  $p < 1.0e-16$ ), the average degree of network node (protein) connectivity is 12.7, and the clustering coefficient is 0.501. There are several clusters of the network – the largest one includes the genes HARS, KRAS, NRAS, PTPN11. There are several smaller, but also connected in a common network of clusters, the largest number of links in the TP53 gene (a known oncogene). In general, analysis of the structure of the gene network for SC genes shows the existence of a dense, connected, rather large cluster of genes (network nodes) that includes known oncogenes, such as TP53.

*Conclusion:* Kaposi's sarcoma is a dangerous disease, but with a fairly high survival rate [5]. Treatment for Kaposi's sarcoma can be local or systemic. Local therapy includes applications of 30 % prospidin ointment, radiation methods, cryotherapy, injections of chemotherapeutic drugs into the tumor, applications with dinitrochlorobenzene, injections of interferon-alfa into the tumor, and some other methods [7]. Radiation therapy, local chemotherapy, surgical method, cryotherapy with liquid nitrogen, photodynamic therapy, systemic polychemotherapy or palliative monochemotherapy approaches also used [6]. The study of the structure of the gene network shows a high connectivity of genes and their products. An analysis of the literature (PubMed) showed a continuing increase in publications on this topic – a total of 11,800 publications to date. There are new approaches based on circular RNA studies [3].

*Acknowledgements:* This publication has been supported by the Sechenov University and Potanin Foundation grant for masters' teachers (TK22-000797).

## References

1. Ablashi D.V., Chatlyne L.G., Whitman J.E. Jr, Cesarman E. Spectrum of Kaposi's sarcoma-associated herpesvirus, or human herpesvirus 8, diseases. *Clin Microbiol Rev.* 2002;15(3):439-64. doi: 10.1128/CMR.15.3.439-464.2002.
2. Gubanova N.V., Tsygankova A.R., Zavjalov E.L., Romashchenko A.V., Orlov Y.L. Biodistribution of 10B in Glioma Orthotopic Xenograft Mouse Model after Injection of L-para-Boronophenylalanine and Sodium Borocaptate. *Biomedicines.* 2021;9(7):722. doi: 10.3390/biomedicines9070722.
3. Lakiotaki E., Kanakoglou D.S., Pampalou A., Karatrasoglou E.A., Piperi C., Korkolopoulou P. Dissecting the Role of Circular RNAs in Sarcomas with Emphasis on Osteosarcomas. *Biomedicines.* 2021;9(11):1642. doi: 10.3390/biomedicines9111642.
4. Orlov Y.L., Galieva A.G., Orlova N.G., Ivanova E.N., Mozyleva Y.A., Anashkina A.A. Reconstruction of gene network associated with Parkinson disease for gene targets search. *Biomeditsinskaya Khimiya.* 2021;67(3):222-230. doi: 10.18097/PBMC20216703222.
5. Requena C., Alsina M., Morgado-Carrasco D., Cruz J., Sanmartín O., Serra-Guillén C., Llombart B. Kaposi Sarcoma and Cutaneous Angiosarcoma: Guidelines for Diagnosis and Treatment. *Actas Dermosifiliogr (Engl Ed).* 2018;109(10):878-887. doi: 10.1016/j.ad.2018.06.013.
6. Serquina A.K.P., Tagawa T., Oh D., Mahesh G., Ziegelbauer J.M. 25-Hydroxycholesterol Inhibits Kaposi's Sarcoma Herpesvirus and Epstein-Barr Virus Infections and Activates Inflammatory Cytokine Responses. *mBio.* 2021;12(6):e0290721. doi: 10.1128/mBio.02907-21.
7. Trostrup H., Bigdeli A.K., Krogerus C., Kneser U., Schmidt G., Schmidt V.J. A Multidisciplinary approach to complex dermal sarcomas ensures an optimal clinical outcome. *Cancers (Basel).* 2022;14(7):1693. doi: 10.3390/cancers14071693.



## **Morphogenesis of *Drosophila melanogaster* mechanoreceptors: system regulation**

Furman D.P., Bukharina T.A.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

*Novosibirsk State University, Novosibirsk, Russia*

*furman@bionet.nsc.ru; bukharina@bionet.nsc.ru*

**Key words:** drosophila, mechanoreceptors, gene networks, central regulatory circuit

The drosophila mechanoreceptors, residing on the head and body of imago according to a strictly ordered pattern, are the result of determined conversion of the ectodermal cells of imaginal discs into progenitor neural cells. The morphogenesis of mechanoreceptors comprises three successive stages: (1) separation of the proneural clusters out of the ectodermal cell population; (2) separation of a single sensory organ precursor (SOP) cell in every proneural cluster; and (3) its asymmetric division with further specialization of the daughter cells into the components of definitive mechanoreceptor. The listed events of the morphogenesis of both an individual mechanoreceptor and the overall bristle pattern are controlled by a hierarchically organized molecular genetic system. This system limits the number of neurally predetermined cells, first to several tens of cells in the proneural clusters and then to a single SOP cell within a cluster. Our reconstruction of the system suggests that its elements group into three modules that correspond to the stages of mechanoreceptor morphogenesis – the gene networks named “Neurogenesis:prepattern”, “Neurogenesis:determination”, and “Neurogenesis: asymmetric division”. These gene networks form a hierarchically organized molecular-genetic system that ensures the morphogenesis of an individual mechanoreceptor and the bristle pattern as a whole. The connecting link of the networks is the *AS-C* proneural genes complex and ASC proteins. The activity of the complex acquires a special role at the stage of separation of SOP cells. This stage is controlled by the “Neurogenesis:determination” gene network and its operation is coordinated by the central regulatory circuit (CRC).

*Acknowledgements:* The study is supported by FWNR-2022-0020.

# Petri net modeling and simulation of biological networks

Hofestädt R.

AG Bioinformatics and Medical Informatics, University Bielefeld, Bielefeld, Germany

\* ralf.hofestaedt@uni-bielefeld.de

**Key words:** Petri net, biological networks, modeling, simulation

*Motivation and Aim:* Modeling and simulation of metabolic networks is a classical topic of Bioinformatics and Chemical Informatics. However, first publications can be found published in the field of Biophysics/Biomathematics more than 60 years ago. Considering the classical papers we can identify two fundamental classes of models. Discrete models like automata or formal languages and analytical models like complex differential equations. At the end of the last century more and more discrete models were published for modeling and simulation of metabolic networks. The main reason for discrete models was (is) the gap of relevant molecular data. Based on new omics technologies this situation changed during the last years. Therefore, “Quantitative Biology” came up as a new research topic. Therefore, flexible methods for modeling biological networks are useful. Today it seems that the Petri net approach is the “best method” for modeling and simulation of complex biological networks.



**Fig. 1.** The tool OMPetri is available via <https://agbi.techfak.uni-bielefeld.de/OMPetri>

*Results:* This presentation will discuss the key features of Petri nets, the advantages of Petri nets for modeling and simulation of biological networks, the relevant simulation tools and the OMPetri net tool, which allows the user friendly access to this new and powerful Petri net simulation tool. OMPetri based on the OpenModelica environment and the library PNlib which is specified entirely and transparently.

*Conclusion:* The new Petri net simulation tool is available via: <https://agbi.techfak.uni-bielefeld.de/OMPetri>. This tool is useful for modeling and simulation of biological networks and can be characterized by the following features:

- discrete and continuous places,
- discrete, continuous and stochastic transitions,
- extended arcs,
- arc weight functions,
- capacities,
- enabling function for places / arc to discrete transitions,
- delay for discrete transitions,
- maximum speed function for continuous transition,
- random delay parameter for stochastic transitions,
- condition function for each transition and
- valid marking.

### *References*

1. Brinkrolf C., Ochel L., Hofestädt R. VANESA: An open-source hybrid functional Petri net modeling and simulation environment in systems biology. *Biosystems*. 2021;210:104531. doi: 10.1016/j.biosystems.2021.104531.

# Functional and evolutionary characteristics of the gene network controlling appetite in mice: lessons from knockout or knockdown animals

Ignatieva E.V.<sup>1,2\*</sup>, Mustafin Z.S.<sup>1,2</sup>, Lashin S.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *eignat@bionet.nsc.ru*

**Key words:** gene network, appetite disorders, phylostratigraphic age index, divergence index, knockout, knockdown

*Motivation and Aim:* Appetite is a drive to consume food. It is one of the most primitive instincts promoting survival and 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]. In efforts to understand the molecular-genetic basis of appetite abnormalities and to collect data on potential targets for new therapies, we (1) have compiled a list of genes regulating food intake identified in experiments on knockout or knockdown animals; (2) have constructed gene network involving physical interactions between these genes or encoded proteins; (3) analyzed functional and evolutionary characteristics of genes in the network

*Methods and Algorithms:* Data on genes regulating appetite were collected with the use of text-mining-based approach, performing queries to PubMed using such terms as “knockout” or “knockdown” and terms designating appetite manifestations or disorders. If the genes have been identified in rats, zebrafish, etc., then an appropriate mouse ortholog has been identified too. To reconstruct the gene network comprising genes/proteins involved in physical interaction with each other we have used computer systems STRING [2], GeneMania [3] and ANDVisio [4]. Evolutionary characteristics of genes (phylostratigraphic age index (PAI) and divergence index (DI)) have been calculated as it is described in [5] taking into account sequences of orthologous genes that are 50 % or more identical to one under consideration.

*Results:* Data on 96 genes regulating appetite has been collected. Knockout or knockdown of these genes lead to hypophagia or hyperphagia in experimental animals (mainly in mice). The largest part of genes encoded cell surface receptors and signal molecules (hormones, neuropeptides), and about a quarter of all genes are genes encoding such regulatory proteins as transcription factors (TFs) and enzymes regulating protein activities. It has been found that the total number of anorexigenic genes significantly exceeds the total number of orexigenic genes. Vice versa, the number of genes encoding orexigenic TFs significantly exceeds the number of genes encoding anorexigenic TFs.

Using PAI and DI we have compared evolutionary characteristics of mouse appetite-controlling genes with the same characteristics defined for all protein-coding mouse genes. A number of distinctive features have been found: (1) the set of mouse appetite-

controlling genes is enriched with genes whose evolutionary age corresponds to the appearance of vertebrates; (2) this set of genes contains an increased proportion of genes with low DI ( $< 0.2$ ). It has also been found that the distribution of PAI values for orexigenic genes has less variance than the distribution of PAI values of all protein-coding mouse genes, i.e., young genes (aged at time of *Mammalia* origin and later) are absent in the set of orexigenic genes.

Gene network comprising genes/proteins involved in physical interaction with each other has been constructed. The highest number of interactions have *Stat3*, *Foxo1*, *Pomc*, and *Sirt1*. All genes that make up this network have low DI (below 0.45), indicating that these genes are under high selection pressure.

*Conclusion:* Reviewing PubMed publications that present the results of knockout or knockdown experiments, we have collected data on 96 mouse genes regulating appetite. Analysis of the gene network involving these genes or encoded proteins has revealed a complex nature of the molecular-genetic mechanisms controlling appetite. We identified genes with the highest functional load in the gene network and revealed features of evolutionary characteristics of the appetite-controlling genes. Taking into account these findings may be useful in the search for new pharmacological targets for the treatment of appetite disorders.

*Acknowledgements:* The study is supported by the State Budgeted Project FWNR-2022-0020.

### References

1. Mitchell J.E., Crow S. Medical complications of anorexia nervosa and bulimia nervosa. *Curr Opin Psychiatry*. 2006;19(4):438-43.
2. Szklarczyk 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.
3. Warde-Farley D. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(Suppl\_2):W214-W220.
4. Ivanisenko V.A. et al. A new version of the ANDSys 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.
5. Mustafin Z.S. et al. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov J Genet Breed*. 2021;25(1):46-56.

## **ANDSystem: cognitive computer system for automated gene networks reconstruction and analysis**

Ivanisenko V.A.<sup>1,2,3\*</sup>, Demenkov P.S.<sup>1,2</sup>, Ivanisenko T.V.<sup>1,2</sup>, Ivanisenko N.V.<sup>1,2</sup>,  
Antropova E.A.<sup>1</sup>, Adamovskaya A.V.<sup>1,3</sup>, Revva P.M.<sup>1,3</sup>, Lavrik I.N.<sup>1,4</sup>,  
Kolchanov N.A.<sup>1,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup> *Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University Magdeburg, Magdeburg, Germany*

\* [salix@bionet.nsc.ru](mailto:salix@bionet.nsc.ru)

The main source of data for the reconstruction of gene networks are scientific publications. Among the existing software and information systems that use methods of automatic knowledge extraction from scientific texts, cognitive computer systems are of particular interest. A distinctive feature of these systems is the full cycle of knowledge engineering, which includes the stage of identifying entities in the text, establishing relationships between them, integrating the acquired knowledge and presenting it in the form of semantic networks or a knowledge graph. In addition, a characteristic feature of such systems is the availability of a wide range of tools designed for data analysis and interpretation, including gene prioritization, identification of overrepresented biological processes or diseases, etc. Examples of cognitive systems in the field of biomedicine are Watson [1], Ingenuity Pathway Analysis (IPA) [2], STRING [3], Pathway Studio [4]. We have developed a cognitive software and information system ANDSystem [5-7]. The ANDSystem considers 12 types of objects, including organisms, cells/tissues, diseases, genes, metabolites, drugs, etc. For these objects, knowledge about physical interactions, catalytic reactions, regulatory relationships, such as regulation of expression, regulation of activity, regulation of stability, etc. is extracted from texts. There are more than 20 types of interactions in total.

ANDSystem has been used to solve a wide range of problems in the reconstruction of the molecular mechanisms of diseases, the interpretation of omics data, and other problems in the field of biomedicine. In particular, an urgent problem is the study of 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 the ANDSystem, the associative gene networks describing the potential regulation of the apoptosis process by HCV viral proteins were reconstructed. Another direction of research was related to the study of the molecular mechanisms of pathological processes in hepatocellular carcinoma. Induction of apoptosis by the cell response to mechanical stress caused by tumor tissue compaction is important antitumor mechanism [8, 9]. The analysis of gene networks made it possible to identify potential signaling pathways linking genes involved in the cell response to mechanical stress with key genes of the external pathway of apoptosis. Potential molecular mechanisms of dysfunction of these signaling pathways in hepatocellular carcinoma have also been reconstructed.

*Acknowledgements:* Work 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.

### *References*

1. Ying C., Argentinis J.D., Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther.* 2016;38:688-701.
2. Krämer A., Green J., Pollard Jr.J., Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics.* 2014;30(4):523-530.
3. 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):D607-613.
4. Nikitin A., Egorov S., Daraselia N., Mazo I. Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics.* 2003;19(16):2155-7.
5. Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics.* 2020;21:228.
6. 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.
7. 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 Systems Biol.* 2015;9(Suppl 2):S2.
8. Levayer R. Solid stress, competition for space and cancer: The opposing roles of mechanical cell competition in tumour initiation and growth. In: *Seminars in cancer biology.* 2020;63:69-80.
9. Hong E., Lee E., Kim J., Kwon D., Lim Y. Elevated pressure enhanced TRAIL-induced apoptosis in hepatocellular carcinoma cells via ERK1/2-inactivation. *Cell Mol Biol Letters.* 2015;20(4):535-548.

# Tumor mechanobiology and related prospects for targeted therapies *in silico* simulations and analysis

Kalinin V.

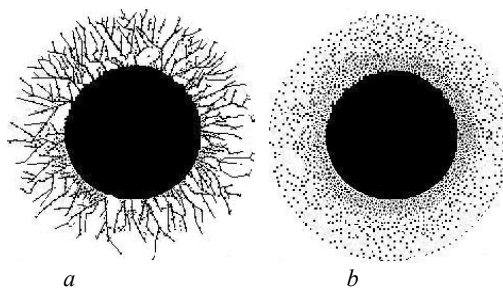
R&D Sector of TMA, Dublin, Ireland

vladimir.kalinin@tma-science.com

**Key words:** Glioblastoma multiform, Multicellular Tumor Spheroids, targeted therapy

Glioblastoma multiform (GBM) is one of the most lethal types of human cancers with a median survival time slightly above a year. The current work is focused on mechanics of glioma cells in Multicellular Tumor Spheroids (MTS) in order to determine Mechanical Modulators of glioma – possible alterations in mechanical properties or structure of tumor that may have a critical effect on glioma progression. Unlike typical targeting therapies designed to ultimately destroy or damage cancer cells, the mechanical modulation may affect tumor growth on the structural level, slowing down its progress. The work presents analysis based on *in-silico* MTS simulations of glioma growth and also available *in vitro* results.

Mechanical factors of tumor growth represent rheological properties of Cell- ECM system and also combination of the hydraulic mobility and motility of glioma cells can be seen as Mechanical Modulators of glioma progression. The difference between the last two cell characteristics is that the first one- mobility characterizes purely passive cell motion under intercellular pressure resulting from excessive cell proliferation, while cell motility relates to their individual activity providing their motion through ECM. The effect of glioma suppression, where these modulators were ultimately targeted has been demonstrated experimentally [1], as well as on computer models [2]. The pathways of these *modulatory* effects have been studied in the current work.



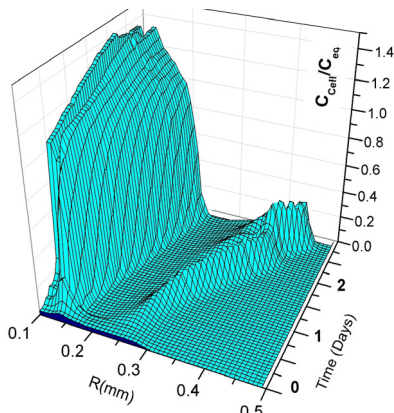
**Fig. 1.** Stylized schemes of the two typical forms of invasive patterns: Radial (*a*) and Disperse (*b*). Patterns of 'Radial' glioma invasion typically obtained in *in vitro* tests demonstrated 'moderate' cell activity. Aggressive cell lines often exhibit dispersal invasion [3, 4]

It was demonstrated in number of studies that different glioma cell lines exhibit different invasive strategies. This difference is especially noticeable between highly malignant glioma cell lines and less aggressive cells [3, 4] and can be observed optically on invasive patterns. An example of the two characteristic types of glioma invasive patterns is shown on Fig. 1. As shown earlier [2] the transition from 'Radial' to 'Dispersal' type of invasion may be related to formation of subtumors, where distant cell clusters are

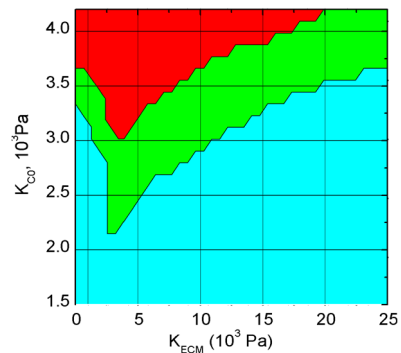


being formed in the invasive zone. The current study demonstrates that sub-tumor appearance results from radial instability in tumor core (Fig. 2) and is conditioned by the ratio of  $K_{Co}/K_{ECM}$ , where  $K_{ECM}$  and  $K_{Co}$  bulk moduli of ECM and glioma cells correspondingly. The Fig. 2 shows formation of a single cluster in the periphery of invasive zone and its growth. This cluster suppresses radial invasion, but initiates formation of ‘star burst’ pattern (see Fig. 1, *b*) typical for aggressive gliomas. Stiff cells may form more than one distant cluster (Fig. 3). Therefore, the model demonstrates a transition from relatively moderate radial strategy of invasion to the aggressive dispersed structure. The diagram on Fig. 3 covers the range of moduli values much wider than the one typical for *in vitro* and *in vivo* gliomas. This range is shown to demonstrate the principal diagram shape with most occurrences of dispersed invasive pattern at  $K_{Co}/K_{ECM} \sim 1$ .

The model assumptions are to be outlined briefly to demonstrate the capabilities of the toolkit used. It considers three cell phenotypes: proliferative cells with low or zero motility, highly motile hypoxic cells with low or zero proliferation rate and necrotic cells, which do not exhibit any proliferation or motility, while show the same hydraulic mobility as all other cells do, being under interstitial pressure. The local oxygen content determines all transitions between cell phenotypes within predetermined transition times. Viscoelastic compressible cell continuum moves through elastic compressible extracellular scaffold driven by individual cell motility and stress distribution in the composite system. The model is based on continuum approximation with the following assumptions. Extracellular Matrix (ECM) is isotropic, compressible, linear elastic, porous material. The cells exert attractive and repulsive forces on each other depending on intercellular distances. The lower limit of cell concentration in the colony is determined by the critical distance between them. Beyond this threshold cells cannot sense or forcibly affect each other. The model has been described in details in [2], including mathematical problem statement, initial and boundary conditions and model validation based on experimental data.



**Fig. 2.** Calculated radial distribution of glioma cell density  $C_{Cell}$  as function of time in MTS growth;  $C_{eq} = 3.5 \times 10^8$  cells/cm<sup>3</sup>. This dynamics demonstrates the formation of sub-tumor in front of MTS



**Fig. 3.** Calculated zones of sub-tumor formation, as they depend on ECM and cell colony bulk moduli, based on the current model of glioma invasion. Blue zone- radial invasive strategy; Green- formation of a single sublayer around MTS; Red- two or more sublayers lead to dispersed structure of invasive pattern

Modulation of tumor growth through variation of cell-ECM system rheology has been demonstrated experimentally [5, 6]. The model also allows us to observe alterations of tumor aggressiveness induced by modulations of cell-ECM dynamics. This collective cell dynamics of 3D tumor growth is not being replicated in typical invasion/migration assays. Though the assays produce valuable data for further *in silico* simulations of glioma growth. Cell motility and hydraulic mobility, as their key invasive characteristics, are individual mechanical attributes of each cell line and can be extracted from the assay results. The results of transwell assays obtained in [1] for A172, U87-MG, SW1088 and H4 glioma cells have been used for *in silico* simulations of glioma MTS growth in the current work. The simulation results are in line with the available data on aggressiveness of tumors formed by these glioma cells. The cell modifications carried out in [1] resulted in their new mechanical characteristics were also used for glioma MTS simulations. The simulations confirmed glioma suppression for most aggressive lines A172 and U87-MG assumed by authors of [1], showing for the first case even a transformation of the invasive structure from dispersal type to radial one, as illustrated on Fig. 1. At the same time, the model does not confirm the expected promotion of tumor activity for modified SW1088 and H4 cells. This result illustrates one more time that invasion assays do not replicate 3D tumor growth. It's important to note that the modulatory effect on dynamic cell characteristics depends dramatically on cell-ECM rheological conditions, as shown on Fig. 3. An immediate consequence of this effect is that 3D *in vitro* test results of tumor growth may differ completely from *in vivo* tumor dynamics, as the rheological properties of typical gels used for *in vitro* tests, matrigel in particular, differ substantially from *in vivo* rheology. However, *in silico* simulations were shown as an efficient complementary toolkit capable of predicting the modulatory effects on 3D tumor growth. It can be potentially a part of computer aided mechanobiological targeting strategy for potential anticancer therapy.

### References

1. Louca M., Stylianou A., Minia A. et al. *Sci Rep.* 2019;9(7782):1-13.
2. Kalinin V. *J Integr Bioinform.* 2022;20200027. DOI 10.1515/jib-2020-0027.
3. Vinci M., Box C., Eccles S.A. *J Vis Exp.* 2015;(99):e52686.
4. Stein A., Demuth T., Mobley D. et al. *Biophysical J.* 2007;92(1):356-365.
5. Ulrich T.A., Pardo E. de Juan, Kumar S. *Cancer Res.* 2009;69(10):4167-74.
6. Ananthanarayanan B., Kim Y., Kumar S. *Biomaterials.* 2011;32(31):7913-7923.

## Impact of negative feedbacks on *de novo* pyrimidines biosynthesis in *E. coli*

Khlebodarova T.M.<sup>1,2\*</sup>, Akberdin I.R.<sup>1,3,4,5</sup>, Kozlov K.N.<sup>6</sup>, Kazantsev F.V.<sup>1,2,3</sup>, Fadeev S.I.<sup>7,#</sup>, Likhoshvai V.A.<sup>1,#</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>5</sup> *Sirius University, Sochi, Russia*

<sup>6</sup> *Peter the Great St.Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>7</sup> *Institute of Mathematics, SB RAS, Novosibirsk, Russia*

# *Deceased*

\* *tamara@bionet.nsc.ru*

Regulatory processes via negative feedback loops play an essential role in biological systems like metabolic pathways and molecular genetic circuits. They are able to orchestrate intracellular processes, sustain the cellular homeostasis and possess living systems to adapt to variable environmental conditions. Theoretical analysis of mathematical models describing cellular processes controlled by negative feedback loops demonstrates the possibility of oscillatory dynamics in these systems, while a combination of positive-negative feedbacks and doubled negative feedbacks contribute to the manifestation of more complex dynamic modes including quasiperiodic, chaotic and even hyper chaotic. Furthermore, features of the dynamic modes depend not only on the number and type of regulatory feedbacks, but also on their length, i.e. a number of intermediate steps between regulatory element and its target.

From this perspective, an investigation of dynamic modes of key metabolic systems regulated by negative feedbacks, such as a system of pyrimidines biosynthesis, is of particular research interest. Herein, a new mathematical model of pyrimidine nucleotides biosynthesis is presented that includes all experimentally verified negative feedback loops regulating enzymatic reactions. Computational analysis of the model demonstrated the existence of stationary and oscillatory modes of the system behavior under particular sets of kinetic parameters. It is worth to note that values of the parameters fit into physiological bounds of the studied metabolic system. It was shown that an emergence of the oscillatory pattern in the biosynthesis depends on a variation of two parameters indicating Hill coefficient ( $h_{UMP_1}$ ) for nonlinear impact of UMP on the carbamoylphosphate synthase activity and parameter  $r$  that designates an impact of the non-competitive mechanism of UTP inhibition in the regulation of an enzymatic reaction of UMP phosphorylation. Finally, it was proposed that the biosynthesis of pyrimidine nucleotides in *E.coli* follows the oscillatory pattern which emerges depending on the regulatory mechanism of the UMP kinase activity.

*Acknowledgements:* This work was supported by the budget project No. FWNR-2022-0020).

## Computation and tracing of periodic solutions of Marchuk–Petrov model

Khristichenko M.\*, Nechepurenko Yu., Grebennikov D., Bocharov G.

*Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

\* *misha.hrist@gmail.com*

**Key words:** delay differential equations, periodic solutions, relaxation method, Newton type method, Fourier series expansion, tracing along a parameter

*Motivation and Aim:* Models represented by systems of delay differential equations are widely used to study the dynamics of infectious diseases and immune response. Stable periodic and non-trivial stationary solutions to such systems are interpreted as chronic diseases. Therefore, to study chronic diseases, it is necessary to compute the stationary and periodic solutions with sufficiently high accuracy and study their stability. Efficient methods for computing stationary solutions of time-delay systems and analyzing their stability were proposed in [1, 2]. These methods were tested on two well-known models, namely, the model of the dynamics of experimental infection caused by lymphocytic choriomeningitis viruses (LCMV) proposed in [3], and the Marchuk–Petrov model of antiviral immune response (MPM) proposed in [4]. It was shown that the proposed methods made it possible to obtain all stationary solutions of each model for given values of its parameters, trace these solutions along a selected parameter, and study their stability. The report discusses methods for computing periodic solutions of delay systems, proposed in [5, 6] using the Marchuk–Petrov antiviral immune response model as an example, and presents a new method for tracing a given periodic solutions along a selected parameter.

*Methods and Algorithms:* A periodic solution is suggested to be searched for in vicinity of a given unstable stationary solution in 3 stages: computation of an approximate periodic solution by the relaxation method, estimation of the least period of the approximate periodic solution by the Fourier series expansion, refining of the approximate periodic solution and the least period by a Newton type method. Tracing of a given periodic solution along a selected parameter is performed by constructing a grid along this parameter and computing the solution at each new node of the grid by the Newton type method starting with the solution computed at previous node.

*Results:* The proposed technology was applied to compute and trace periodic solutions for the Marchuk–Petrov model with parameter values corresponding to hepatitis B. The dependence of the amplitude and period of the solutions on the virus and immune response parameters of the model was examined in detail. The identified relationships provide relevant information for understating the mechanisms underlying relapsing forms of the infection and spontaneous recovery.

*Conclusions:* The proposed methods make it possible to obtain a periodic solution of the Marchuk–Petrov model in the vicinity of a given unstable stationary solution and to trace the periodic solution along a selected parameter. Application of the methods to a systematic analysis of the model parameter space for existence of qualitatively and quantitatively different types of chronic infection dynamics enables novel understanding of the hepatitis B infection pathogenesis.

*Acknowledgements:* This work was supported by the Russian Science Foundation (Grant 22-11-00025).

### *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 Series S*. 2020;13(9):2385-2401.
2. 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 Modelling*. 2020;35(2):95-110.
3. Bocharov G.A. Modelling the dynamics of LCMV infection in mice: conventional and exhaustive CTL responses. *J Theor Biol*. 1998;192(3):283-308.
4. 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.
5. Khristichenko M.Yu., Nechepurenko Yu.M. Computation of periodic solutions to models of infectious disease dynamics and immune response. *Russ J Numer Anal Math Modelling*. 2021;36(2):87-99.
6. 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).

## Sirius platform as environment for evolutionary development of mathematical models of complex biological systems

Kolpakov F.A.<sup>1,2,3\*</sup>, Kiselev I.N.<sup>1,2,3</sup>, Kutumova E.O.<sup>1,2,3</sup>, Akberdin I.R.<sup>1,3,4</sup>,  
Ryabova A.S.<sup>1,3</sup>, Yevshin I.S.<sup>1,3</sup>, Sharipov R.N.<sup>1,3,4</sup>, Zhatchenko S.A.<sup>1,3</sup>

<sup>1</sup> *Sirius University, Sochi, Russia*

<sup>2</sup> *FRC for Information and Computational Technologies, Novosibirsk, Russia*

<sup>3</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>4</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *fedor@biouml.org*

**Key words:** mathematical modeling, complex biological systems, evolutionary development, BioUML, GitLab, Sirius platform

*Motivation and Aim:* Development of mathematical models of complex biological systems is impossible from scratch. Such models can be developed only step-by-step, with incremental increasing of their complexity. This process is similar to the evolutionary one. However, this process is very time- and labor-consuming. For example, the most modern modular model of the human cardiovascular and renal systems for blood pressure regulation [1] still contains some part of the original Guyton model for overall regulation of circulation [2], that is already 50 years old. A rapid evolutionary development of mathematical models of complex biological systems requires a specialized software platform where models can evolve like living creatures (Table 1).

**Table 1.** Basic evolutionary processes for living creatures and mathematical models

Living creatures	Mathematical models
Multiplying	Easy copying from a repository
Living	Reproducibility – possibility to simulate and reproduce the original results
Mutagenesis	A model editing
Recombinations	Composing complex models from modules
Natural selection	Passing tests, correspondence to experimental results

Thus, the corresponding software environment should provide:

- a repository from which mathematical models can be easily copied. Thorough documentation about the models structure and parameters is also essential for further models development;
- editor (preferably both visual and textual) for model editing;
- support of modular approach for composing complex models from modules;
- simulation engine;
- test engine that checks the compliance of the model with certain conditions and experimental data;

- parameters fitting.

*Methods and Algorithms:* BioUML (<https://www.biouml.org>) – is a web-based integrated platform for systems biology and data analysis [3]. It supports visual modeling and development of hierarchical biological models that allow us to construct the most complex modular models of blood pressure regulation [1], skeletal muscle metabolism [4], and COVID-19 epidemiology [5].

Jupyter notebook/lab/hub (<https://jupyter.org/>) – Jupyter notebooks are widely used in biosciences for interactive and reproducible data analysis and visualization. We have integrated Jupyter notebooks into the BioUML platform using the Jupyter hub technology and developed Jupyter kernels for R, Python and JavaScript languages that provide access to BioUML functionality inside the corresponding notebooks. Jupyter notebooks can be used to reproduce the original results and model testing.

BioStore (<https://bio-store.org/>) – is the central hub, which provides collaborative research using the BioUML platform. For this purpose, it contains a registry of users, their projects, access rights to projects (several users can have access to the same project) and a list of publicly available servers where the BioUML platform or its derivatives are installed. Sirius (<https://sirius-web.org/>) – an information platform for projects related to the data analysis and modeling for educational and research projects of Sirius University. This platform combines the possibilities of the BioUML platform for modeling and data analysis and GitLab for project storage, management and documentation.

*Results:* To demonstrate the suggested approach, we have deposited into the Sirius platform the following complex mathematical models:

- blood pressure regulation [2]  
<https://gitlab.sirius-web.org/virtual-patient/blood-pressure-regulation>
- skeletal muscle metabolism [3]  
<https://gitlab.sirius-web.org/virtual-patient/muscle-metabolism>
- COVID-19 epidemiology [5]  
<https://gitlab.sirius-web.org/covid-19/dde-epidemiology-model>

A user can easily clone the corresponding model from the git repository. A repository contains not only the model SBML code, but also detailed description, as well as Jupyter notebooks to reproduce the original results. Using the appropriate links from README file on the start page of the respective git project, a user can open the model in one click in the BioUML platform, edit it visually, simulate its behavior, as well as reproduce original simulation results using Jupyter notebooks.

*Acknowledgements:* The study was supported by Sirius University.

## References

1. Kutumova E., Kiselev I., Sharipov R., Lifshits G., Kolpakov F. Thoroughly Calibrated Modular Agent-Based Model of the Human Cardiovascular and Renal Systems for Blood Pressure Regulation in Health and Disease. *Front Physiol.* 2021;12:746300.
2. Guyton A.C., Coleman T.G., Granger H.J. Circulation: overall regulation. *Annu Rev Physiol.* 1972;34:13-46.
3. Kolpakov F., Akberdin I., Kiselev I., Kolmykov S., Kondrakhin Yu., Kulyashov M., Kutumova E., Pintus S., Ryabova A., Sharipov R., Yevshin I., Zhatchenko S., Kel A. BioUML – towards a universal research platform. *Nucleic Acids Res.* 2022. In press. doi: 10.1093/nar/gkac286.
4. Akberdin I.R., Kiselev I.N., Pintus S.S., Sharipov R.N., Vertyshev A.Yu., Vinogradova O.L., Popov D.V., Kolpakov F.A.. *IJMS.* 2021;22(19):10353. doi: 10.3390/ijms221910353.
5. Kiselev I.N., Akberdin I.R., Kolpakov F.A. A Delay Differential Equation approach to model the COVID-19 pandemic. *medRxiv.* 2021.09.01.21263002. doi: 10.1101/2021.09.01.21263002.

# Applicability of the Oxford Nanopore sequencing technology to the analysis of DNA modifications

Konanov D.N.<sup>1\*</sup>, Babenko V.V.<sup>2</sup>, Ilina E.N.<sup>1,2</sup>

<sup>1</sup> *Institute of System Biology and Medicine of Rosпотrebnadzor, Moscow, Russia*

<sup>2</sup> *Federal Research and Clinical Center of Physical and Chemical Medicine of FMBA, Moscow, Russia*

\* *konanovdmitriy@gmail.com*

**Key words:** Oxford Nanopore sequencing, DNA modifications

*Motivation and Aim:* The Oxford Nanopore sequencing technology makes it possible not only to detect the four canonical bases, but also to additionally catch any of their modified forms that can significantly change the sequencing signal. However, the most popular tool used for DNA modification analysis called Tombo has a number of limitations, especially when processing low-represented DNA motifs. In this work, we attempt to improve existing methods for the analysis of DNA modifications and demonstrate their applicability on two highly methylated *H. pylori* genomes.

*Methods and Algorithms:* Using the Oxford Nanopore technology, both native and whole genome amplified (WGA) samples of two strains of *H. pylori* J99 and A45 were sequenced. The sequencing results were processed with Guppy and resquiggled with Tombo. To detect modified sites, the raw signal distributions of native and WGA samples were compared using the Kolmogorov-Smirnov test in combination with artificial data balancing to compensate the differences in the k-mers abundance. In addition, the Cohen's d-effect size was used to estimate the magnitude of the signal shift. A new greedy algorithm has been developed to extract a minimum set of modified sites based on observed signal shifts. The results obtained for J99 were compared with PacBio data.

*Results:* We have developed a new statistical approach for detecting modified positions and a new greedy algorithm for extracting potential modification motifs. Unlike Tombo, our method showed results close to the PacBio method on the *H. pylori* J99 strain. A new *H. pylori* strain A45 was analyzed by the same method and revealed the appearance of several methylation motifs other than J99.

*Conclusion:* Oxford Nanopore sequencing has a great potential for the analysis of epigenetic modifications and can successfully compete with PacBio technology. However, the currently existing software has certain drawbacks caused by the inconsistency of the statistical methods used. The authors hope that their efforts to improve the existing methods will be helpful to ONT users interested in DNA modifications analysis.

*Acknowledgements:* The study was supported by Government contract No. 122030900064-9.



# Creation of the mathematical model for the prognostic analysis of human pluripotent stem cells based on their morphological portrait

Krasnova O.A.<sup>1</sup>, Chabina A.S.<sup>1</sup>, Kulakova K.A.<sup>1</sup>, Alekseenko L.L.<sup>1</sup>, Gursky V.V.<sup>1,2</sup>, Neganova I.E.<sup>1\*</sup>

<sup>1</sup> *Institute of Cytology, RAS, St. Petersburg, Russia*

<sup>2</sup> *Ioffe Institute, RAS, St. Petersburg, Russia*

\* *irina.neganova@incras.ru*

**Key words:** human pluripotent stem cells, mathematical model, pluripotency

The ability of human pluripotent stem cells (hPSCs) for unlimited proliferation and self-renewal promotes their application in the fields of regenerative medicine, new drugs creation, *in vitro* human diseases modeling, cell banks development. The ‘quality control’ of the colony requires information on how morphological parameters of hPSCs are associated with pluripotency maintenance. We analyzed ten morphological parameters extracted from the phase-contrast images of hESC line (H9), control induced pluripotent stem cell line (iPSC AD3), and iPSC CaSR. Each colony was classified, by visual analysis, according to its potential for pluripotency and clonality maintenance, thus defining the colony phenotype as control parameter. Mathematical analysis of the variance of the morphological parameters in groups with different phenotype allows us to select parameters that were potentially informative in determining the phenotype. Features such as Area, Feret’s diameter, Compactness, and Shape factor are among the most prominent under various conditions. These parameters were used as predictors in the phenotype classification model, which was trained to produce high classification accuracy (79 % for colony parameters and 90 % for cellular parameters). Further, we performed qRT-PCR analysis for the expression of 14 pluripotency genes in groups of hPSCs with the same morphological portrait. By addressing the expression variance of the selected genes, we analyzed which ones can be described as the most informative for the separation of different morphological phenotype. Models of phenotype and clonality classification based on these genes expression predict the phenotype and clonality with 100 % accuracy based on the expression of only four genes (*SALL4*, *NANOG*, *REX1*, and *SOX2*). Our results indicate the fundamental possibility of constructing a morphological portrait of a colony informative for the automatic identification of the phenotype and link this portrait to expression of the pluripotency markers.

*Acknowledgements:* This research supported by the RSF grant No. 21-75-20132 for IN.

# Reconstruction of mathematical frame models of bacterial transcription regulation based on transcriptional regulatory networks

Lakhova T.<sup>1\*</sup>, Kazantsev F.<sup>1,2</sup>, Lashin S.<sup>1,2</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* tlakhova@bionet.nsc.ru

**Key words:** mathematical model, transcriptional regulatory network, transcription factor

*Motivation and Aim:* In recent years, an enormous amount of full genome sequencing data has been accumulated, including bacterial data. However, the analysis of these data lags behind the rate of their accumulation, particularly in the context of transcription regulation. For example, for the most studied model organism *Escherichia coli*, the complete description of genetic regulation has not yet been completely solved [1]. Even less information is available for other bacteria. The use of mathematical modelling techniques makes it possible to assess the impact of transcription regulation processes on the behaviour of a model system in dynamics. For example, simulations can provide general answers to the following questions: how gene expression changes at the transcription level with or without the participation of selected transcription factors (TFs); what contribution a series of TFs have to the system behaviour when their mutual regulation is taken into account, etc. Only experiment can answer these questions more accurately.

*Methods and Algorithms:* The developed method makes it possible to derive fusion rate functions from the available structure of the operon promoter region, which specifies the interaction variants between TFs and their binding sites (TFBSs). The equations are constructed in terms of generalized Hill functions [2]. Resulting models of transcription processes and their regulation are described by differential equations. The program is implemented in the Java programming language.

*Results:* This work presents the development of a previously developed algorithm for reconstruction of frame models based on the structure of bacterial transcriptional regulatory networks (TRNs) [3]. The new iteration takes into account the structure of operon – TFs roles and possible locations of TFBSs on the promoter region. These parameters more accurately specify the effects of regulation. The output contains the operon regulation model equations in text format, i.e. one file – one model. This is done for the convenience of further analysis. The algorithm was tested on bacterial genome sequencing data from the database of the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS.

*Conclusion:* The developed tool complements previously developed algorithms [3, 4] for genome analysis.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia (075-15-2019-1662).

## References

1. Santos-Zavaleta A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* 2019;47(D1):D212-D220.
2. Likhoshvai V., Ratushny A. Generalized Hill function method for modeling molecular processes. *J Bioinform Comput Biol.* 2007;5.(02b):521-531.
3. Lakhova T.N. et al. Construction of transcriptional regulatory networks based on found transcription factors and their binding sites in annotated bacterial genomes. *Systems Biology Bioinformatics (SBB-2021)*. 2021;27-27.
4. Lakhova T.N. et al. Development of algorithms for finding transcription factors and their binding sites in annotated bacterial genomes. *Systems Biology Bioinformatics (SBB-2020)*. 2020;87-87. (in Russian)

# On organizing a software platform for seeking biotechnologically important features in bacteria

Lashin S.<sup>1,2\*</sup>, Demenkov P.<sup>1,2</sup>, Ivanisenko V.<sup>1</sup>, Kazantsev F.<sup>1,2</sup>, Mukhin A.<sup>1</sup>, Kolchanov N.<sup>1,2</sup>

<sup>1</sup>*Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* [lashin@bionet.nsc.ru](mailto:lashin@bionet.nsc.ru)

**Key words:** microbial biotechnology, software platform, systems biology, mathematical model

*Motivation and Aim:* Modern biotechnology combines approaches of genetic engineering with technologies of computer analysis and modeling of the molecular-genetic organization of microorganisms and computer design of experiments to construct superproductive strains with specified properties. Taken together, this leads to the development of genetic and metabolic engineering technologies for the design of superproducing strains of target products.

The aim of this work is to develop a software platform to assist in the cycle of research from the search for new promising for biotechnology species of microorganisms to the construction of superproductive strains of target products. The platform is based on the use of developed bioinformatic approaches: software for assembly and annotation of sequenced microorganism genomes, computer analysis and modeling of molecular-genetic organization of microorganisms.

*Methods and Algorithms:* We used a wide range of methods and approaches of bioinformatics analysis (genome assembly and annotation, search of functional sites within genomes, reconstruction of gene regulatory networks [1] and metabolic pathways), systems biology (automatic generation of mathematical and computational models of microbial metabolism and genetic regulation [2]) as well as approaches and technologies of software development (client-server and micro services architectures, databases etc.) to implement the platform.

*Results and Discussion:* We have developed a suit of instruments, which provide a microbial genome analysis chain starting from genome assembly, functional annotation, search for various functional regions (operons, promoters, transcription factor binding sites) within genome and more bioinformatics analysis through the reconstruction of various biological networks (associative gene networks, transcription regulation networks, metabolic pathways), to the construction and analysis of mathematical and computational models of molecular-genetics processes in microbes, which include flux balance models and dynamical models written in the form of ordinary differential equations.

*Availability:* <https://cgimu.sysbio.cytogen.ru>.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia (075-15-2019-1662).

## References

1. Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. *BMC Bioinformatics*. 2020;21(S11):228. doi: 10.1186/s12859-020-03557-8.
2. Kazantsev F.V. et al. MAMMOTH: A new database for curated mathematical models of biomolecular systems. *J Bioinform Comput Biol*. 2018;16(01):1740010. doi: 10.1142/S0219720017400108.

# Study of the mechanism of reactions catalyzed by the cytochrome C complex with cardiolipin activated by coumarin dyes

Levchenko I.N.<sup>1\*</sup>, Volodyaev I.V.<sup>2</sup>, Boyarchenkov A.S.<sup>3</sup>

<sup>1</sup> Pirogov Russian National Research Medical University, Moscow, Russia

<sup>2</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Russia

\* [irnlevchenko@yandex.ru](mailto:irnlevchenko@yandex.ru)

**Key words:** cytochrome C, peroxidases, nanomachine, peroxidases, mathematical modeling, quantum yield

*Motivation and Aim:* Modern technologies and programming tools make it possible not to simply solve the problem, but also, using simulation and complex modeling methods, to create a virtual reality for work in which researchers can consider the topic of interest in all its diversity [1]. Here we analytically solved kinetic equations, describing the oxychemiluminescent reactions in the suspension of mitochondrial membranes, and compared the obtained results with experimental data [2, 3]. The system contained cytochrome C, forming a complex with cardiolipin, activated by the coumarin dye C-525 during apoptosis. The free radical formation was constantly monitored via the oxychemiluminescence kinetics. For data analysis, we developed a special program to calculate reaction rate constants based on the known reaction scheme, and compare different calculation algorithms.

*Methods and algorithms:* For solving the direct problem, the law of acting masses was used; the system of differential equations for reaction rates was solved by the Euler method; the rate constants were selected manually according to the method developed elsewhere [2]. The inverse problem was solved based on the above data and the law of quantum yield constancy. Several optimization algorithms in C++ were considered and implemented. Each of them was compared step by step with the experimental data, and the most successful algorithm was selected based on the calculated oxychemiluminescence quantum yields [3].

As is known, the chemiluminescence quantum yield in the presence of coumarin dyes is several orders of magnitude higher than for spontaneous oxychemiluminescence [4, 6]. Thus, analyzing electronic states in the process of photoexcitation [5], kinetic rate constants can be determined, which subsequently allows to determine the desired reaction scheme. To visualize the process, the photon emission was calibrated in conventional units so that the calculated luminescence kinetics and rate constants could be compared with the experimental data.

*Conclusion:* A method has been developed for analyzing the substrate of the peroxidase reaction, which can serve as an activator of chemiluminescence when studying the effect of the cytochrome C complex with cardiolipin on cancer cells.

*Acknowledgements:* We would like to express our gratitude to the supervisors for their valuable contribution to the planning of the study, namely, Academician, Professor Yuri Andreevich Vladimirov, Faculty of Fundamental Medicine, Moscow State University

and Professor Anatoly Nikolaevich Osipov, Faculty of Medicine and Biology, Pirogov Russian National Research Medical University.

### *References*

1. Malikov R.F. Fundamentals of the development of computer models of complex systems.
2. Vladimirov Yu.A., Batov L.P. Studies of changes in the cytochrome function with the method of mathematical modeling.
3. Vladimirov Yu.A., Vladimirov G.K. Structure and peroxidase function of the cytochrome C complex with cardiolipin in an aqueous medium and in a non-polar environment.
4. Vladimirov Yu.A., Proskurnina E.V., Izmailov D.Yu. Chemiluminescence as a method of detection and investigation of free radicals in biological systems. *BEBiM*. 2007;144(3):390-6.
5. Vladimirov Y.A., Sharov V.S., Driomina E.S., Reznitchenko A.V., Gashev S.B. Coumarin derivatives enhance the chemiluminescence accompanying lipid peroxidation. *Free Radical Biol Medicine*. 1995;18(4):739-45.
6. Zhuravlev A.I. Quantum biophysics of animals and humans: Textbook. 3rd ed. M.: MGAVMiB, 2009.
7. Levitus M. Tutorial: measurement of fluorescence spectra and determination of relative fluorescence quantum yields of transparent samples. *Methods and Applications in Fluorescence*.
8. Vladimirov Y.A. et al. Mechanism of activation of cytochrome C peroxidase activity by cardiolipin. *Biochemistry (Mosc)*. 2006;71(9):989-997.

## Study of the mechanism of reactions catalyzed by complexes of cytochrome C with cardiolipin

Levchenko I.N.<sup>1\*</sup>, Volodyaev I.V.<sup>2</sup>, Boyarchenkov A.S.<sup>3</sup>

<sup>1</sup> Pirogov Russian National Research Medical University, Moscow, Russia

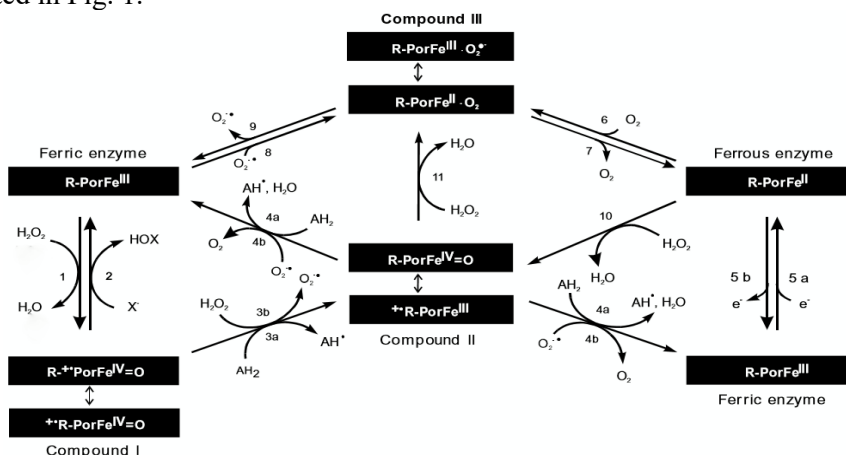
<sup>2</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Russia

\* irnlevchenko@yandex.ru

**Key words:** cytochrome C, peroxidases, nanomachine, peroxidases, mathematical modeling

*Motivation and Aim:* When studying objects of the living world, we often encounter phenomena that are well described by algorithms from the category of nanomachines. Models of this class have existed for more than 50 years and effectively help mankind in understanding many phenomena and solving practical problems. Thus, in many cases, simplified models of nanomachines allow a fairly good understanding of complex multi-component biological processes. For instance, the complex of cytochrome C with the phospholipid cardiolipin, interpreted as a nanomachine, quite well describes one of the key nodes in triggering apoptosis – interaction of cytochrome C with the inner membrane of mitochondria, activating its peroxidase activity and further release into the cytoplasm. Cytochrome C is a heme-containing protein that promotes electron transfer between mitochondrial respiratory complexes 3 and 4, and participates in triggering apoptosis. As shown [1], cytochrome C forms a complex with the membrane phospholipid cardiolipin, which activates its peroxidase activity and initiates lipid peroxidation in the mitochondrial membranes. This breaks the membrane barrier properties, releasing cytochrome C into the cytoplasm, where it binds to the protein Apaf-1, forming apoptosomes and triggering apoptosis. The general scheme of its peroxidase action is presented in Fig. 1.



**Fig. 1.** General reaction scheme of mammalian peroxidases

Regarded as a nanomachine, the Cytochrome C / cardiolipin complex is especially valuable as a trigger, switching between normal cell functioning and apoptosis

depending on its peroxidase activity. Generally speaking, the latter should be measured by the rate of free radical generation, but due to their extremely high reactivity, it cannot be done directly. Thus, it is estimated by the intensity of oxychemiluminescence, accompanying further reactions of the generated free radicals, and has to be done in dynamics – which demands the use of the so-called kinetic chemiluminescence (see [5]). *Methods and Algorithms:* Having the above experimental data available, in this work, we describe the mathematical model, creating an automated solution to the direct and inverse problems to study the mechanism of cytochrome C participation in apoptosis. Importantly, in this work we consider only the control process, which is carried out by changing the protein/lipid structural ratio and membrane viscosity.

The reaction rates were expressed as the product of the concentration of the reactants and the corresponding rate constant (see Fig. 1). The rate constants were selected manually. The initially formed system of differential equations for reaction rates was solved by the Euler method.

Further, specific optimization was proposed for solving the inverse problem. Several algorithms, each of which could serve as the basis for optimizing the selection process both for the rate constants and for the concentrations of the given substances, were progressively compared with the experiment, with the best one finally selected.

*Conclusion:* A program for the automatic determination of the rate constants of chemical reactions was created and debugged, corresponding to the formulation of the inverse problem.

Unsatisfactory solutions were screened, and the best model selected.

The rate constants of the reactions proceeding in the system for measuring the activity of the luminescence intensity were determined.

*Acknowledgements:* We would like to express our gratitude to the supervisors for their valuable contribution to the planning of the study, namely, Academician, Professor Yuri Andreevich Vladimirov, Faculty of Fundamental Medicine, Moscow State University and Professor Anatoly Nikolaevich Osipov, Faculty of Medicine and Biology, Pirogov Russian National Research Medical University.

## References

1. Vladimirov Y.A. et al. Mechanism of activation of cytochrome C peroxidase activity by cardiolipin. *Biochemistry (Mosc)*. 2006;71(9):989-97.
2. Cortese J.D., Voglino A.L., Hackenbrock C.R. The ionic strength of the intermembrane space of intact mitochondria is not affected by the pH or volume of the intermembrane space. *Biochim Biophys Acta*. 1992;1100(2):189-197.
3. Nantes I.L. et al. Modifications in heme iron of free and vesicle bound cytochrome C by tert-butyl hydroperoxide: a magnetic circular dichroism and electron paramagnetic resonance investigation. *Free Radic Biol Med*. 2000;28(5):786-796.
4. Dawson J.H. Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science*. 1988;240(4851):433-439.
5. Furtmuller P.G. et al. Kinetics of interconversion of redox intermediates of lactoperoxidase, eosinophil peroxidase and myeloperoxidase. *Jpn J Infect Dis*. 2004;57(5):S30-S31.
6. Weng L.C., Baker G.M. Reaction of hydrogen peroxide with the rapid form of resting cytochrome oxidase. *Biochemistry*. 1991;30(23):5727-5733.
7. Kagan V.E. et al. A role for oxidative stress in apoptosis: oxidation and externalization of phosphatidylserine is required for macrophage clearance of cells undergoing Fas-mediated apoptosis. *J Immunol*. 2002;169(1):487-499.
8. Kagan V.E. et al. Cytochrome C acts as a cardiolipin oxygenase required for release of proapoptotic factors. *Nat Chem Biol*. 2005;1(4):223-232.



# Data lake platform of the bacterial strain properties for microbiology

Mukhin A.<sup>1,2\*</sup>, Demenkov P.<sup>1,2</sup>, Kazantsev F.<sup>1,2,3</sup>, Lakhova T.<sup>1,2</sup>, Klimenko A.<sup>1,2,3</sup>, Lashin S.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [mukhin@bionet.nsc.ru](mailto:mukhin@bionet.nsc.ru)

**Key words:** database, data-lake, microbiology, SQL, java, genomes, annotation

*Motivation and Aim:* Modern high-throughput sequencing methods lead to generation of large amounts of genomic data. These data then processed with generating additional bunches of data. All of it should be collected in a structured way like a data warehouse. In developing such repository, one must also consider that data may be described either incompletely, or may not be fully annotated, or may be described using non-standard or erroneous formats [1]. Because of these problems, it is impossible to reuse the primary study results in subsequent computational experiments. All of these things must be taken into account in solving the research tasks of finding biotechnologically meaningful properties of microorganisms. The work is aimed at the tasks of the Kurchatov Genomics Center of the ICG SB RAS that collects large amounts of bacterial strains with biotechnological potential and performed DNA sequencing.

*Methods and Algorithms:* In this work we used Java and Kotlin programming languages, web framework Spring Boot to build the server part of the system. HTML, CSS and JavaScript programming language are used to build client part using VueJS framework. PostgreSQL is used as the database management system. The Data Lake architectural concept was chosen to create the repository. It stores both primary and derived structured data within a single software system. The reuse data problem is solved with taking into account FAIR [2] principles. The task of comprehensive describing the bacterial strain properties is important from both a scientific point of view and an engineering one (creating producers of certain substances or maximizing biomass).

*Results:* The developed system is named MicrobialDataLake. It includes data pipelines from a series of bash-based scripting tools: bacterial genome assembly and annotation, taxonomic identification of the strain, generation of structural (graph) and mathematical frame models. These results form the informational basis of the MicrobialDataLake system and are located in the raw data repository. To describe the results of analysis and links to the raw data, a JSON schema format has been implemented. The PostgreSQL database management system was also deployed to describe in detail biotechnologically relevant properties of microbial strains (such as the lists of enzymes synthesized by the strain, taxonomic affiliation, gene and protein sequences, etc.). To parse and upload the structured properties of each microorganism, a script was implemented in the Kotlin programming language. A natural tool for working with genomes is the genome browser; the JBrowse2 browser for genome visualization was integrated into the software system. A web service providing access to the data via REST API was also implemented using Java and Spring Boot framework. The Kotlin tools and the Java web service were

developed using common domain classes and database accesses in order to develop a unified platform.

At the moment, the data of 1004 bacterial strains are presented in the system.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### *References*

1. Roche D.G., Kruuk L.E.B., Lanfear R., Binning S.A. Public data archiving in ecology and evolution: how well are we doing? *PLOS Biol.* 2015;13(11):e1002295. doi: 10.1371/journal.pbio.1002295.
2. Wilkinson M.D. et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3(1):1-9. doi: 10.1038/sdata.2016.18.

# Evolutionary analysis of gene networks with Orthoweb software

Mustafin Z.<sup>1,2\*</sup>, Mukhin A.<sup>1</sup>, Anikin D.<sup>2</sup>, Kazantsev F.<sup>1,2</sup>, Lashin S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *mustafinzs@bionet.nsc.ru*

**Key words:** phylostratigraphic analysis, gene networks, web-service

*Motivation and Aim:* Phylostratigraphic analysis is a new trend in evolutionary analysis, aimed to determine the time of gene origin [1]. The method based on analysis of taxonomic classification of genes and their orthologs and finding the common taxon. With such method of sequences analysis, as dN/dS ratio [2], it becomes possible to detect the time of gene's origin and the type of selection affecting to it. The methods of such analysis, implemented in different software packages (myTAI [3], phylostratr [4], Orthoscape [5], etc.), require the user to know the programming languages (mostly R) or how to use another software, like Cytoscape [6]. Most methods also aimed to analyze the gene lists, but it is interesting to look not only to genes, but also to their interactions and possible clusters in gene networks.

*Methods and Algorithms:* Orthoweb software has been developed in the Java programming language. To develop the server part, we used the Spring framework; VueJS and Webix frameworks have been used for the client part. Networks visualizing is implemented with cytoscape.min.js library. The document-oriented database MongoDB is used to store the data from KEGG database (taxa, list of orthologous, coding sequences, etc.) and intermediate analysis results. It is significantly increases the speed of future work with the same data.

To store the result of analysis, we have developed the PostgreSQL database. Using REST API, the user can retrieve the data from previous analysis of gene with exactly same parameters.

*Results and Discussion:* We present a web-based service Orthoweb, which allows to analyze both gene sets and gene networks, using only web browser. It is possible to import a gene network to Orthoweb (from one of two external databases, from manually generated file or from automatically generated file by, for example, STRING [7]). Orthoweb visualizes the results of analysis by coloring and sizing the nodes and their borders, and additionally upload the gene ontology terms, associated with genes. Orthoweb allows interacting with nodes and edges and saving the result. Orthoweb also allows to get the results of analysis for model organisms using API of PostgreSQL database to get the data without analysis.

*Availability:* The web-service is available on <https://orthoweb.sysbio.cytogen.ru>.

*Acknowledgements:* This study was supported by the Budgetary Project FWNR-2022-0020 and by RFBR No. 20-04-00885.

## References

1. Domazet-Lošo T., Brajković J., Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 2007;23(11):533-539. doi: 10.1016/j.tig.2007.08.014.
2. Kryazhimskiy S., Plotkin J.B. The population genetics of dN/dS. *PLoS Genet.* 2008;4(12). doi: 10.1371/journal.pgen.1000304.

3. Drost H.-G., Gabel A., Liu J., Quint M., Grosse I. myTAI: evolutionary transcriptomics with R. *Bioinformatics*. 2018;34(9):1589-1590. doi: 10.1093/bioinformatics/btx835.
4. Arendsee Z., Li J., Singh U., Seetharam A., Dorman K., Wurtele E.S. Phylostratr: A framework for phylostratigraphy. *bioRxiv*. 2018;360164. doi: 10.1101/360164.
5. 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(S1):1-9. doi: 10.1186/s12859-016-1427-5.
6. 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.
7. Szklarczyk 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. doi: 10.1093/nar/gky1131.

## Modern trends in personalized hemodynamics

Parshin D.\*, Tikhvinsky D., Kuianova Iu., Chupakhin A.

*Lavrentyev Institute of Hydrodynamics, SB RAS, Novosibirsk, Russia*

\* *danilo.skiman@gmail.com*

**Key words:** cerebral hemodynamics, bypass, aneurysm, intravascular device

*Motivation and Aim:* Vascular diseases hold the leading place in deaths as a result of diseases (according to WHO [1]). Cerebral aneurysms or abdominal aortic aneurysms are common diseases and occur in 2–5 % of the population depending on gender and race [2]. The most promising approach in the treatment of such pathologies is the approach that includes the use of personalized patient data or personalized preoperative modeling. However, such modeling in some cases faces both computational and expert challenges, and therefore requires a versatile approach – from numerical experiments with idealized formulations to formulations taking into account nonlinear models of vessel wall elasticity. The paper presents various approaches to preoperative modeling of hemodynamics, signs of development or stabilization of pathologies and, as a result, allowing medical decisions to be made.

*Methods and Algorithms:* A numerical decision on the installation of a vascular anastomosis was considered both in the model [3] and in the patient-specific case; the model problem of the growth of an abdominal aortic aneurysm [4], the problem of correct endothelialization of flow-redirecting devices in the treatment of cerebral aneurysms [5], and the problem of embolization of a model configuration of arteriovenous malformation according to experimental data on the rheology of real embolisates. All calculations were performed using the ANSYS 17.2–20.2 software packages.

*Results:* Based on the results of the work, an optimal kgol anastomosis was found, the parameters of normal endothelialization of the flow-redirecting device were calculated for the first time, and an increase in pressure proximal to the abdominal aortic aneurysm was proved, which is confirmed in clinical practice.

*Conclusion:* The results of the work make a significant contribution to the solution of various problems of cerebral and abdominal hemodynamics and can significantly increase the understanding of hemodynamics by real users of such models – practicing surgeons.

*Acknowledgements:* The study is supported by a grant from Russian Science Foundation, (Projects No. 20-71-10034 concerning cerebral hemodynamics and No. 21-15-00091 concerning hemodynamics of aorta).

### References

1. World Health Organization: <https://www.who.int/ru>.
2. Feigin V.L. et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet*. 2014;383(9913):245-254.
3. Kuyanova Y.O. et al. Numerical study of the tee hydrodynamics in the model problem of optimizing the low-flow vascular bypass angle. *J Appl Mech Tech Phys*. 2019;60(6):1038-1045.
4. Tikhvinskii D.V. et al. Numerical study of the effect of aortic bifurcation region geometry on the abdominal aortic aneurysm hemodynamics. *Russ J. Numer Anal Math Model*. Submitted.
5. Tikhvinskii D.V. et al. Numerical Assessment of the Risk of Abnormal Endothelialization for Diverter Devices: Clinical Data Driven Numerical Study. *J Pers Med*. 2022;12(4):652.

## **Age-related alterations in the mouse liver circadian clock are caused by an imbalance of the NAD<sup>+</sup> consumption system**

Podkolodnyy N.L.<sup>1,2\*</sup>, Tverdokhleby N.N.<sup>1</sup>, Podkolodnaya O.A.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

\* [pnl@bionet.nsc.ru](mailto:pnl@bionet.nsc.ru)

Data on the role of NAD<sup>+</sup>-dependent histone deacetylase SIRT1 in the integration of circadian rhythm regulation and metabolic pathways, as well as on the function of NAD<sup>+</sup> as a "metabolic oscillator" have opened a promising direction in this area. Especially important is the experimentally observed decrease in the cellular level of NAD<sup>+</sup> in both chronological and biological aging in various tissues in experimental animals and humans. The liver is the main metabolic organ of mammals, providing de-novo NAD<sup>+</sup> synthesis and NAD recycling. Therefore, the analysis of the interaction of the liver circadian clock with the NAD<sup>+</sup> consumption system and the modeling of age-related changes associated with these interactions is the actual task.

A mathematical model of the interaction of the circadian oscillator in the mouse liver with the NAD<sup>+</sup> consumption system, including the enzyme SIRT1, PARP1, CD38, whose activity changes with age, has been developed and tested on experimental data.

We included in our model the genes *Arntl*, *Clock/Npas2*, *Per1/2*, *Cry1/2*, *Rora/γ*, *Rev-Erba/β*, *Dbp*, *Nfil3*, *Nampt*, which are important for the functioning of the circadian oscillator in the mammalian liver (mouse). To verify the model, we used experimental data with various types of impacts, knowledge about the intervals of possible values of model parameters and the effect of knockout of various genes on the dynamics of the model and the oscillation period, the relationship between amplitudes, phases of model variables.

The simulation showed a pronounced circadian rhythm of the CB38, PARP1 and SIRT1 activity, and the fact that age-related changes in the activity of these enzymes lead to increased NAD<sup>+</sup> catabolism and its depletion. Such disorders can be one of the causes of dysfunction of circadian oscillators in the liver and contribute to disruption of circadian rhythms in the body as a whole.

*Acknowledgements:* Supported by FWNR-2022-0020, and No. 0251-2021-0004.

## Reconstruction and analysis of the gene network of hepatocellular carcinoma under mechanical stress of the cell

Revva P.<sup>1,2,3\*</sup>, Demenkov P.S.<sup>1,2</sup>, Ivanisenko V.A.<sup>1,2,3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

\* p.revva@g.nsu.ru

**Key words:** hepatocellular carcinoma, mechanical stresses, apoptosis

*Motivation and Aim:* Hepatocellular carcinoma is a leading cancer worldwide. Its incidence is increasing, and is closely related to advanced liver disease. Cirrhosis represents the greatest risk factor for this malignancy, and is the main indication for screening and surveillance. Due to the absence of symptoms in 60 % of patients, hepatocellular carcinoma is diagnosed late, cases – against the background of multiorgan metastasis. Despite advances in medical, locoregional and surgical therapies, hepatocellular carcinoma remains one of the most common causes of cancer-related death globally [1]. Alterations to the mechanical properties of the microenvironment are a hallmark of cancer. Elevated mechanical stresses exist in many solid tumors and elicit responses from cancer cells. Uncontrolled growth in confined environments gives rise to elevated solid compressive stress on cancer cells. Cancer cells in the tumor respond to elevated mechanical stresses, and these responses are important for cancer cell survival. Solid compressive stress alters cancer cell behaviors such as proliferation, invasion, and apoptosis. Mechanical stress in tissue cells should lead to apoptosis. However, this is not observed in tumor cells [2]. The aim of this work was to reconstruct and analyze the gene network of hepatocellular carcinoma under mechanical stress of the cell and to build frame models of the regulation of apoptosis by hepatocellular carcinoma.

*Methods and Algorithms:* The gene networks were built using the ANDSystem system for automatic reconstruction of associative gene networks. The databases of Gene Ontology were used to search for genes associated with cell response to mechanical stress and ClinVar to search for single nucleotide polymorphisms associated with hepatocellular carcinoma.

*Acknowledgements:* Work 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.

### References

1. Hartke J. et al. The diagnosis and treatment of hepatocellular carcinoma. *Semin Diagn Pathol.* 2017;34:153-159.
2. Purkayastha P. et al. Molecular cancer cell responses to solid compressive stress and interstitial fluid pressure. *Cytoskeleton.* 2021;78.

## Gene network reconstruction and functional association studies for oncological diseases using online bioinformatics tools

Shakirova R.R.<sup>1\*</sup>, Elistratova M.G.<sup>1</sup>, Korobeynikova A.V.<sup>1</sup>, Mozyleva Y.A.<sup>1</sup>, Orlova N.G.<sup>2,3</sup>, Orlov Y.L.<sup>1,4</sup>

<sup>1</sup> *The Digital Health Institute, I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

<sup>3</sup> *Financial University under the Government of the Russian Federation, Moscow, Russia*

<sup>4</sup> *Moscow State Technical University of Civil Aviation, Moscow, Russia*

\* [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** biomedicine, oncology, complex disorders, gene network, bioinformatics, databases, e-Health

*Motivation and Aim:* Functional annotation of genes and gene ontology analysis for complex diseases using available bioinformatics tools provide background to search new targets for therapy. We aimed develop computer pipelines to study set of complex heterogeneous disease such as cancers and mental disorders using online bioinformatics tools. Examples of gene lists studied are related to glioma, glioblastoma, cystic fibrosis, Kaposi's sarcoma, Parkinson's disease and dementia. Thus, cystic fibrosis a systemic hereditary disease caused by a mutation in the gene for the transmembrane regulator of cystic fibrosis and characterized by damage to the external secretion glands, severe respiratory dysfunction. Parkinson's disease and dementia present socially important complex diseases. Functional annotation of genes related to the disease could be retrieved based on genetic databases and cross-validated by integrating complementary experimental data. Gene network reconstruction for a set of genes (proteins) proved to be effective approach to study mechanisms underlying disease progression.

*Methods and Algorithms:* We used online bioinformatics tools such as GeneCards, GeneMANIA (<https://genemania.org/>) and STRING-DB (<https://string-db.org/>) for annotation of gene list for the disease, reconstruction of gene network and comparative analysis of gene ontology categories. The Internet-resource OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) and GeneCards suit (GeneCards.org) were used for gene list reconstruction. DAVID (<https://david.ncifcrf.gov/>) and PANTHER resources (Protein Analysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) were used to analyze the categories of gene ontologies. The issue of developing bioinformatics courses is related to the adaptation of the training to the educational profile of students and trainees. Examples of the tasks in medical informatics, which can be solved using only online tools – the formation of lists of genes associated with the disease, reconstruction of the associative gene network, visualization of the gene network, calculation of categories of gene ontologies, overrepresented in a given list of genes, determination of the spatial structure of the protein. These tasks students are able to solve independently, using online tools, with the preparation of abstracts and their own publications as it was discussed at series of students workshops [6] and Younf Scientists Schools on bioinformatics [8].

*Results:* We reconstructed gene networks for series of oncological diseases – glioma, lymphoma, and Kaposi's sarcoma. The work on gene network models include



glioma [2], Parkinson's disease [5], and metabolic syndrome [9]. The steps include collection of a list of genes associated with the development of the disease (e.g., Parkinson's disease), analyze gene ontology categories for such a list, and reconstruct the gene network. For the key disease genes derived from the gene network structure analysis, drug search options are considered [5]. This approach has already been tested for teaching the course "Fundamentals of Bioinformatics and Databases Management" at the Sechenov University [1, 2]. In many cases, it requires advanced training, additional education in modern genomics technologies, such as high-throughput genomic and transcriptomic sequencing, data analysis, metagenome sequencing, non-coding RNA studies, Artificial Intelligence (AI) applications in medicine, which greatly expand the range of classical bioinformatics tasks [4].

*Conclusion:* Overall, the goal of the work is to use bioinformatics online tools to solve practical tasks in biomedicine. The development of next-gen sequencing technologies poses new challenges which can be solved by already existing online tools without the use of programming skills, in remote access. The issue of developing bioinformatics courses is related to the need to adapt training depending on the educational profile of the trainees [6]. The transition to distance learning requires the development of new methods of teaching, with the ability of students to perform independent work using only Internet resources, which in turn meets a number of challenges due to limited access to a number of online services in Russia (for example, data storage in Google cloud storages) [3].

*Acknowledgements:* The study is supported by the Potanin Foundation grant for masters' teachers (ГК22-000797).

### References

1. Dergilev A.I., Orlova N.G., Dobrovolskay O.B., Orlov Y.L. Statistical estimates of transcription factor binding site clusters in plant genomes based on genome-wide data. *J Integr Bioinform.* 2021;18:20200036. doi: 10.1515/jib-2020-0036.
2. Gubanova N.V., Orlova N.G., Dergilev A.I., Oparina N.Y., Orlov Y.L. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools. *J Integr Bioinform.* 2021;18:20210031. doi: 10.1515/jib-2021-0031.
3. Gusev A.V., Ivshin A.A., Vladzimirsky A.V. Russian mobile health apps: systematic search in app stores. *Russ J Telemed E-Health.* 2021;7(3):21-31. doi: 10.29188/2712-9217-2021-7-3-21-31. (in Russian)
4. Koshechkin K., Lebedev G., Radzievsky G., Seepold R., Martinez N. Review of blockchain technology projects to provide telemedical services. *J Med Internet Res.* 2021;23(8):e17475. doi: 10.2196/17475.
5. Orlov Y.L. et al. Reconstruction of the Parkinson's disease gene network to search for target genes. *Biomed Chemistry.* 2021;67(3):222-230. doi:10.18097/PBMC20216703222. (In Russian)
6. Orlov Y.L., Bakulina A.Y. Development of Education in Bioinformatics Based on Student Conferences ISSC-2018, School of Molecular Modeling and Hackathon in Novosibirsk. *Vestnik Novosibirsk State University. Series: Information Technologies.* 2018;16(3):5-6. (In Russian)
7. Shaderkin I.A., Shaderkina V.A. Remote medical consultations for patients: what has changed in Russia in 20 years. *Russ J Telemed E-Health.* 2021;7(2):7-17. doi: 10.29188/2712-9217-2021-7-2-7-17. (In Russian)
8. Tatarinova T.V., Tabikhanova L.E., Eslami G., Bai H., Orlov Y.L. Genetics researches at the "Centenary of human population genetics" conference and SBB-2019. *BMC Genetics.* 2020;21(Suppl 1):109. doi: 10.1186/s12863-020-00906-7.
9. Tiis R.P., Osipova L.P., Galieva E.R. et al. N-acetyltransferase (NAT2) gene polymorphism and gene network analysis. *Biomeditsinskaya Khimiya.* 2021;67(3):213-221. doi: 10.18097/PBMC20216703213.

## Modeling of contractile activity-induced fatigue in human skeletal muscle

Son A.<sup>1\*</sup>, Akberdin I.<sup>2,3,4</sup>, Kiselev I.<sup>2,3,5</sup>, Pintus S.<sup>2,3,5</sup>, Vertyshev A.<sup>6</sup>, Popov D.<sup>7</sup>,  
Kolpakov F.<sup>2,3,5</sup>

<sup>1</sup> Department of Physics, Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Department of Computational Biology, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology, Sochi, Russia

<sup>3</sup> BIOSOFT.RU, LLC, Novosibirsk, Russia

<sup>4</sup> Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia

<sup>5</sup> Laboratory of Bioinformatics, Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia

<sup>6</sup> JSC "Sites-Tsentr", Moscow, Russia

<sup>7</sup> Institute of Biomedical Problems, RAS, Moscow, Russia

\* a.son1@g.nsu.ru

**Key words:** mathematical model, skeletal muscle, physical exercise, muscle fatigue, modular approach, BioUML

Intensive and/or prolong muscle contractions induce muscle fatigue leading to a decline in exercise performance. This phenomenon arises from central and peripheral changes limiting power output to avoid energy depletion which can cause cell death. The development of muscle fatigue is mainly regulated on the cellular level by both metabolic pathways engaged in ATP synthesis and ATP hydrolysis. However, despite the significant amount of experimental data on metabolic changes contributing to a decrease in power production, a comprehensive mathematical model of the muscle fatigue, which embraces all necessary cellular processes and includes both muscle fiber types, still does not exist. We markedly modified a previous modular model [1] comprising metabolic, signaling and gene expression levels by introducing a new function describing the dependency of power output on the phosphate concentration and pH which affect activation and sensitivity of myosin-actin filaments. To validate the extended version of the model, biochemical data obtained during an intermittent aerobic exercise (rhythmic knee extension for 4 minutes) was used [2]. Numerical analysis of the model was carried out in the BioUML platform [3].

### References

1. Akberdin I.R. et al. *IJMS*. 2021;22(19):10353. doi: 10.3390/ijms221910353.
2. Sundberg C.W. et al. *J Physiol*. 2019;597(19):4943-4957. doi: 10.1113/JP278684.
3. Kolpakov F.A. et al. *Nucleic Acids Res*. 2019;47(W1):W225-W233. doi: 10.1093/nar/gkz440.

## Variability and robustness of the functioning of gene regulatory networks: role in the evolution of early embryogenesis

Spirov A.<sup>1\*</sup>, Myasnikova E.<sup>2</sup>

<sup>1</sup> *I.M. Sechenov Institute of Evolutionary Physiology & Biochemistry, RAS, St. Petersburg, Russia*

<sup>2</sup> *Peter the Great St.-Petersburg Polytechnical University, St. Petersburg, Russia*

\* *Alexander.Spirov@gmail.com*

**Key words:** gene regulatory networks, dynamic modelling, pattern formation, robustness and variability, embryogenesis, evo-devo

The early *Drosophila* development is one of the best studied evo-devo processes. In the first few hours of a fruit fly development, processes of biochemical blueprint take start, predefining the segment organization of the embryo body. Broad spatial gradients of maternally-derived factors sequentially activate cascades of embryonic segmentation genes. These patterns determine the segmented insect body plan in more and more detail. This process should be extremely robust as it determines the correct further development and survival of an individual.

Researchers have long noted that segmentation genes are expressed with greater spatial precision than the maternal patterns. Computational modelling of the gene-gene regulatory interactions provides a mechanistic understanding for this robustness to maternal variability in wild-type (WT) patterning. Functioning of the segmentation gene network is also robust to many of experimental perturbations and a number of mutations. These observations once again raise a long-standing question in evolutionary developmental biology: how such, highly robust to mutations, gene ensembles nevertheless are still evolvable?

What levels of variability and what types of variability are still capable of “pushing out” such a system beyond the WT limits? The more fundamental question is: Are there any pathways in gene regulations which are more vulnerable to losing the WT form? We analyze these problems using our specific models of three segmentation gene ensembles. These are gap, pair-rule, and segment polarity genes. We investigate the question of what alternative (or pathological) dynamic forms are close to WT? Accordingly, if robustness is overcome by a specific perturbation (primarily by a specific mutation), what alternative forms of segmentation patterns are most likely to be realized? Numerically manipulating with maternal patterns in several ways, we have revealed common scenarios in the loss of WT. They consist, typically, in the loss of one or more anterior or posterior segments. So far, we have not been able to find simple scenarios for the emergence of new segments. We discuss these results in terms of the evolvability of insect segmentation, and in terms of experimental perturbations and mutations which could test the model predictions.

*Acknowledgements:* The study is supported by the Russian Foundation for Basic Research grant 20-04-01015.

# Investigation of various hemodynamic regimes in elastic vessels

Vasyutkin S.A.<sup>1,2\*</sup>, Chupakhin A.P.<sup>1,2\*</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Lavrentyev Institute of Hydrodynamics, SB RAS, Novosibirsk, Russia

\* s.vasyutkin@g.nsu.ru, \* alexander190513@gmail.com

**Key words:** one-dimensional model of hemodynamics, traveling waves, blood flow in vessels

The study of fluid flow in an elastic tube is of considerable interest both for fundamental hydrodynamics and for numerous applications: blood flow in vessels, fluid transport in pipelines, etc. An accurate description of this problem includes both finding the velocity and pressure of the fluid, displacements of the pipe walls and stresses in them, and finding the position of the walls. This is a problem with an unknown boundary, the conditions for the coupling of velocity and stresses are set on it. The complexity of this problem initiates the derivation and study of approximate models. One of the most common is the long tube model (one-dimensional model of hemodynamics), in which it is assumed that the length of the tube exceeds its cross-section. The introduction of a small parameter into the problem – the ratio of these quantities – allows us to move to an approximate model in which the velocity is averaged over the pipe section. A lot of works devoted to the one-dimensional model are devoted to both the mathematical side of the issue and physical applications [1, 2]. The characteristic modes of motion in such a system are traveling waves, the study of which was started in [3].

In this case, the system of equations is reduced to a system of ordinary differential equations. The principle is the closure of this system by a kind of equation of state, which sets the dependence of pressure on the radial displacement of the pipe wall. The paper investigates solutions of the traveling wave type for a pipe given by a Coiter shell. In this model, the pressure is represented as a function of the displacement of the wall and the derivatives of this value. The report examines the influence of elastic and viscous factors on the behavior of the solution near the singular point of the corresponding dynamical system. Conditions for different modes of blood flow in the vessels were found. The influence of various physico-chemical characteristics of the vessel walls and blood, which make the greatest contribution to the change of blood flow regimes in the vessels, has been studied. In a numerical experiment, it is shown that solutions of the traveling wave type can describe the deviation of the vessel wall and the characteristics of the blood flow with sufficient accuracy.

## References

1. Čanić S., Tambača J., Guidoboni G., Mikelić A., Hartley C.J., Rosenstrauch D. Modeling viscoelastic behavior of arterial walls and their interaction with pulsatile blood flow. *SIAM J Appl Math.* 2006;67(1):164-193.
2. Il'ichev T., Sumskoj S.I., Shargatov V.A. Unsteady Flows in Deformable Pipes: The Energy Conservation Law. In: Proceedings of the Steklov Institute of Mathematics. 2018;300:68-77.
3. Barlukova A.M., Cherevko A.A., Chupakhin A.P. Traveling waves of a one-dimensional model of hemodynamics. *J Appl Mech Tech Phys.* 2014;55(6):917-926.

## 2.2 Section “Modeling of population and ecological systems and processes”



# Chlorophyll quota and energy-intensive substance for prime production estimation of the aquatic ecosystem

Abakumov A. \*, Pak S.

*Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*

\* [abakumov@dvo.ru](mailto:abakumov@dvo.ru)

**Key words:** chlorophyll, energy-intensive substance, aquatic ecosystem

*Motivation and Aim:* The photosynthesis process is investigated with help of the mathematical modelling. Many researchers are working in this area [1, 2]. We are using of the chlorophyll quota that is mass part chlorophyll in phytoplankton biomass. We modelling of the process of formation of energy-intensive (ATP) substances. These substances are a basic for the growth of the phytoplankton biomass.

*Methods and Algorithms:* The two models are construed from the general idea based on Droop's model. These models have a target for the estimation of the biological production of aquatic ecosystems. The first model uses the chlorophyll quota to describe dynamic changes in phytoplankton due to photosynthesis. The second model uses the concentration of energy-intensive substance too.

*Results:* The chlorophyll quota has very variable values. These values depend on the phytoplankton state and environment. The chlorophyll quota depends from cell quota in our model. Both models have complex dynamics when the stable equations don't exist. These dynamics are subjected to the attractor of the Droop's submodel. The solution dynamics can have complex properties on this attractor. The dynamics could have periodic or quasistochastic properties. This is a result for a stable environment. Our models have a different reaction to any feature of the environment. The big dark period is one of these features. The impact of environmental are shown. The year period for ordinary dynamics of photosynthetic activity radiation and temperature is studied.

*Conclusion:* The models are constructed from the general idea based on Droop's model. Prime production depends on photosynthesis intensity. Our models characterize photosynthesis intensity by chlorophyll content and by energy-intensive substance. This value depends on the phytoplankton state and environment. Our models can have complex dynamics. They could be used for detail description of prime production in aquatic ecosystem.

*Acknowledgements:* The work is supported of the Institute of Automation and Control Processes, FEB RAS (121021700006-0).

## References

1. Platt T., Geider R., Sciandra A., Copin-Montegut C., Bouman H., Sathyendranath S. Dynamics and interactions of autotrophs, light, nutrients and carbon dioxide. In: Robinson A.R., Brink K.H. (Eds.). *The global coastal ocean: multiscale interdisciplinary processes*. Harvard University Press, 2005.
2. Nikolaou A., Hartmann P., Sciandra A., Chachuat 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.

## Closure and stability of trophic chains

Bartsev S. \*, Degermendzhi A.

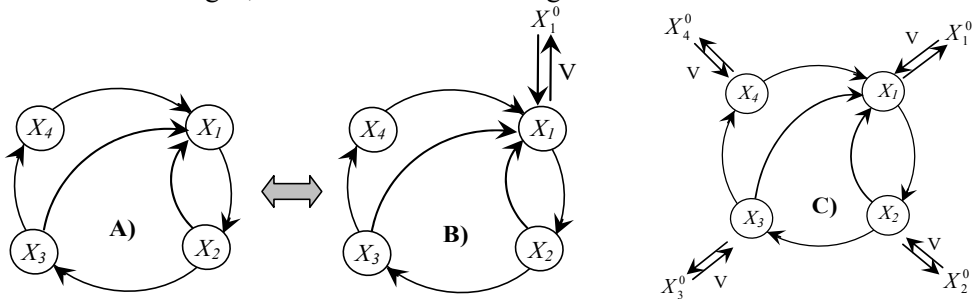
*Institute of Biophysics FRC "Krasnoyarsk Science Center", SB RAS, Krasnoyarsk, Russia*

\* *BartsevSI@ibp.ru*

**Key words:** closed trophic chains, closure and stability, ecosystem stability prediction

*Motivation and Aim:* The emergence of oscillatory regimes in fully or partially closed trophic chains of various lengths is considered in detail in the monograph [1, pp. 242-268]. Three types of trophic functions were used for the analysis: linear (Volterra), hyperbolic and S-shaped. However, there are quite a few trophic functions of a different type, which well describe the experimental data [2]. The question arises whether it is possible to say something about the change in the stability of the trophic chain with a change in the degree of closure of the nutrient flow without specifying the type of trophic functions and for different modes of external biogen exchange.

*Methods and Algorithms:* Consider first a closed trophic chain that has one exchange channel with the environment (Fig. 1, B). The ecosystem of an isolated lake, where carbon acts as a biogen, can serve as a real analogue of such a chain.



**Fig. 1.** Various options of matter exchange of a closed trophic chain with an environment

In general, this scheme can be described by the following system of equations:

$$\begin{cases} \dot{X}_1 = F_1(\vec{X}) + V(X_1^0 - X_1) \\ \dot{X}_i = F_i(\vec{X}) \end{cases} \quad (1)$$

where  $V$  is the rate constant of exchange with the environment;  $\vec{X}$  – vector of concentrations of substances or numbers of organisms in the ecosystem;  $X_1^0$  – the concentration of a substance or the number of organisms of a given species in the environment.

The stationary state of the system is reached at  $X_1^* = X_1^0$ . Note that for system (1) the stationary state is the same for any non-zero values of  $V$ . Let us formally consider the process of increasing closure in this system, which corresponds to a decrease in the exchange rate constant  $V$ . In the limit, at  $V = 0$ , the system takes the form:

$$\dot{X}_i = F_i(\vec{X}) \quad (2)$$

Note that system (2) is qualitatively different from the original system (1), since the order of the system decreases by one due to the conservation law. Let us consider the stability of the stationary points of these two systems. The eigenvalues of the systems can be found by solving the following equations respectively:

$$\Delta_{NC}^n \equiv \begin{vmatrix} \frac{\partial F_1}{\partial X_1} - V - \lambda & \frac{\partial F_1}{\partial X_2} & \dots & \frac{\partial F_1}{\partial X_N} \\ \frac{\partial F_2}{\partial X_1} & \frac{\partial F_2}{\partial X_2} - \lambda & \dots & \frac{\partial F_2}{\partial X_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial F_N}{\partial X_1} & \frac{\partial F_N}{\partial X_2} & \dots & \frac{\partial F_N}{\partial X_N} - \lambda \end{vmatrix} = 0 \quad \Delta_C^{n-1} \equiv \begin{vmatrix} \frac{\partial F_2}{\partial X_2} - \frac{\partial F_2}{\partial X_1} - \lambda & \frac{\partial F_2}{\partial X_3} - \frac{\partial F_2}{\partial X_1} & \dots & \frac{\partial F_2}{\partial X_N} - \frac{\partial F_2}{\partial X_1} \\ \frac{\partial F_3}{\partial X_2} - \frac{\partial F_3}{\partial X_1} & \frac{\partial F_3}{\partial X_3} - \frac{\partial F_3}{\partial X_1} - \lambda & \dots & \frac{\partial F_3}{\partial X_N} - \frac{\partial F_3}{\partial X_1} \\ \dots & \dots & \dots & \dots \\ \frac{\partial F_N}{\partial X_2} - \frac{\partial F_N}{\partial X_1} & \frac{\partial F_N}{\partial X_3} - \frac{\partial F_N}{\partial X_1} & \dots & \frac{\partial F_N}{\partial X_N} - \frac{\partial F_N}{\partial X_1} - \lambda \end{vmatrix} = 0$$

where  $NC$  means incompletely closed,  $C$  means closed trophic chains,  $n$  is the order of the determinant.

Let's formally describe a smooth transition from a partially closed to a completely closed system, which in this case undergoes a qualitative change – it loses one variable. To do this, we need to find a relationship between the above determinants. Shown, that

$$\Delta_{NC}^n = -V\Delta_{NC}^{n-1} - \lambda \Delta_C^{n-1} \quad \text{and} \quad V\Delta_{NC}^{n-1} + \lambda \Delta_C^{n-1} = 0 \quad (3)$$

is the equation for finding the eigenvalues of the food web model.

*Results.* It can be seen from this equation that if the connection with the external environment ( $V$ ) weakens for the system (1), then one of the eigenvalues also tends to zero, that is, one degree of freedom disappears. In this case, the equation for calculating the eigenvalues takes the form  $\Delta_C^{n-1} = 0$ . An equation with similar properties was also

obtained for the case when each of the components of the ecosystem exchanges with the environment with the intensity specified by the parameter  $V$  (see Fig. 1, C).

Since we do not know the specific form of the right-hand sides of the equations, we cannot calculate the eigenvalues, and can only ask in which direction the eigenvalues will shift when  $V$  changes. The Gershgorin's circle method allows localizing eigenvalues by elements of the determinant and thereby estimating the stability of the system [3]. With an increase in  $V$ , the Gershgorin's circle corresponding to the first variable shifts to the left towards negative values. This does not guarantee decreasing probability of oscillatory and chaotic modes, since the eigenvalues are localized in the union of all Gershgorin's circles, however, the stability of the system does not exactly decrease. At the same time numerous computational experiments with have shown that the stability of the system with increasing  $V$ , increases as a rule. In the case of a trophic web with the exchange of all components (see Fig. 1, C) all diagonal elements tend to shift to the negative region with increasing  $V$ , however a change in  $V$  causes a change in stationary values and, potentially, a loss of stability can be observed in some part of the change in  $V$ . However computational experiments have not confirmed that this sometimes happens. Nevertheless it is clear that at very large  $V$ , the values of the food chain variables are determined by the values of the corresponding variables in the environment. That is, the variables themselves become unchanged, but at the same time it is hardly possible to speak of a trophic chain or an ecosystem itself – it has disappeared.

## References

1. Svirezhev Yu.M. Nonlinear waves, dissipative structures and catastrophes in ecology. M., 1987.
2. Tyutyunov Y.V., Titova L.I. *Biol Bull Rev.* 2020;10:167-185.
3. Horn P., Johnson C. *Matrix Analysis*. 2nd ed. NY: Cambridge University Press, 2013.



# Modelling of eco-evolutionary dynamics in microbial communities with haploid evolutionary constructor

Klimenko A.I. \*, Matushkin Yu.G., Kolchanov N.A., Lashin S.A.

*Kurchatov Genomics Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

*Novosibirsk State University, Novosibirsk, Russia*

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [klimenko@bionet.nsc.ru](mailto:klimenko@bionet.nsc.ru)

**Key words:** eco-evolutionary modelling, agent-based modelling, ecological modelling, multi-scale modelling, multi-layer modelling, model integration, simulation tools

*Motivation and Aim:* Being compact ecosystems of small organisms, microbial communities act as a brilliant trade-off between biological complexity and controllability. These communities exhibit significant range of essential life processes while providing broad opportunities to investigate different layers of their biological organization. Meanwhile, microorganisms are known for a rapid mutation rate and volatile genomes affected by such processes as horizontal gene transfer and loss of genetic material. Thus, microbial communities present a fascinating object for investigating eco-evolutionary dynamics.

*Methods and Algorithms:* “Haploid evolutionary constructor” (HEC), a software developed for the modelling of eco-evolutionary processes in microbial communities, considers the following levels of biological organization: genetic, metabolic, cellular, population, and ecological ones. It simulates both high-level (cellular chemotaxis and diffusion, substrates flow and diffusion) and low-level (mutations, horizontal gene transfer, gene regulation and metabolism) processes allowing combining various mathematical modeling approaches (agent-based modeling, differential equations, automata, etc.) in one model. HEC employs the concept of multi-layer modelling to integrate the models describing different biological levels into a unified modelling framework.

*Results:* Haploid evolutionary constructor modelling approach and simulation software package have been extended to describe eco-evolutionary dynamics in spatially structured environments more accurately. HEC has been used to address such problems as investigation of genetic variation in the models of populations inhabiting spatial environment of different structure, modelling of horizontal gene transfer between bacterial cells in changing environmental conditions [1], simulation of evolution of bacterial communities under phage infection [2], eco-evolutionary scenarios in self-organizing microbial communities [3], and competition of motile and sedentary bacterial populations in aquatic environments [4] among others.

*Conclusion:* Haploid evolutionary constructor modelling approach meets the challenge of the integration of cellular and population level models of microbial communities to get the knowledge of how population heterogeneity and complex behavior arises from the variation on individual level. Using multiscale data integration, including but not limited to multi-(meta-)omics data, as a foundation for such a research is essential for the further development in this field starting from the whole-cell models and up to a population level.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### *References*

1. Klimenko A.I. et al. Modeling evolution of spatially distributed bacterial communities : a simulation with the haploid evolutionary constructor. *BMC Evol Biol.* 2015;15(Supp. 1):S3.
2. Klimenko A.I. et al. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: A simulation study. *BMC Microbiol.* 2016;16(1):S10.
3. Klimenko A.I. et al. Spatial heterogeneity promotes antagonistic evolutionary scenarios in microbial community explained by ecological stratification: a simulation study. *Ecol Modell.* 2019;399(2):66-76.
4. Klimenko A. et al. Leave or stay: simulating motility and fitness of microorganisms in dynamic aquatic ecosystems. *Biology (Basel).* 2021;10(10):1019.

## **Bifurcation mechanisms leading to the genetic divergence of two populations coupled by migration**

Kulakov M.P., Frisman E.Ya.\*

*Institute for Complex Analysis of Regional Problems, FEB RAS, Birobidzhan, Russia*

\**frisman@gmail.com*

**Key words:** genetic divergence, population, mathematical modeling, dynamics, migration, bifurcations

*Motivation and Aim:* The study of evolution in biological populations has a long and rich history. One of the main problems of such works is devoted to the search for factors leading to primary genetic divergence, microevolution and speciation. This work, based on a mathematical model, studies the bifurcation mechanisms of the emergence of the genetic divergence, i.e. stable genetic differences between two adjacent populations coupled by migration.

*Methods and Algorithms:* We consider the following simple situation. The populations are panmictic with Mendelian rules of inheritance. The fitness of individuals is strictly determined by a single diallelic locus. We use a dynamic model that describes the change in the concentration of one of the alleles in each population and the ratio (weight) of first population to the total size. To study the model, we used the qualitative theory of differential equations studies.

*Results:* We showed that the genetic divergence is reached only with reduced fitness of heterozygotes and is the result of the subcritical pitchfork bifurcation (period doubling) of the unstable equilibrium state corresponding to polymorphic populations. In free populations with unlimited growth in numbers, solutions corresponding to the genetic divergence are always unstable and are just a part of the transient process to state corresponding to the monomorphic populations. Genetic divergence is stable only for populations that maintain the ratio of their numbers at a constant level. For example, this is possible for populations with limited or fully synchronous dynamics in adjacent sites. In this case, the divergence is preceded by a saddle-node bifurcation. The dynamic is quad-stable, i.e. depending on the initial conditions, the populations are homogeneous in their genetic structure, or there are significant differences between them [1, 2].

*Conclusion:* Thus, the appearance of genetic divergence is impossible without a significant ecological limitation of population growth in addition to genetic factors.

*Acknowledgements:* This study was performed in the framework of the State targets of the Institute of Complex Analysis of Regional Problem FEB RAS

### *References*

1. Kulakov M.P., Frisman E.Y. Simple and complex dynamics in the model of evolution of two populations coupled by migration with non-overlapping generations. *Izv VUZ Appl Nonlinear Dyn.* 2022;30(2):208-232. (In Russian)
2. Frisman E.Y., Kulakov M.P. On the genetic divergence of two adjacent populations living in a homogeneous habitat. *Izv VUZ Appl Nonlinear Dyn.* 2021;29(5):706-726. (In Russian)

## Alpine resort or battle for survival: A didactic story about short-lived perennials and their matrix models

Logofet D.O.\*, Kazantseva E.S., Belova I.N.

A.M. Obukhov Institute of Atmospheric Physics, RAS, Moscow, Russia

\* danilal@postman.ru

**Key words:** *Androsace alba*, *Eritrichium caucasicum*, population viability, pattern-multiplicative average

*Motivation and Aim:* Population structures of two short-lived perennial species, *Androsace alba* and *Eritrichium caucasicum*, have been monitored, in terms of ontogenetic stages, annually on permanent plots in the alpine belt of the North-West Caucasus during 13 years (2009–2021). Data from each pair of successive years ( $t, t+1$ ) provided for the accurate calibration of  $\mathbf{L}(t)$ , the annual population projection matrix (PPM), with its dominant eigenvalue  $\lambda_1$  considered as the asymptotic growth rate. If  $\lambda_1 < 1$ , then the population declines, otherwise it grows exponentially. This advantage of the matrix formalism turns into a serious problem as different PPMs have different  $\lambda_1$ s corresponding to respectively different, sometimes controversial, forecasts of population viability. The challenge was therefore to invent a technique summarizing the whole period of monitoring. The technique was actually invented about 30 years ago within the paradigm of the *stochastic growth rate* ( $\lambda_s$ ) of a population in the random environment [1]. The problem is however an artificial nature of the randomness model adopted in the current literature on  $\lambda_s$ , such as the *i.i.d.* (independent, identically distributed) environments [2]. The aim is therefore to find an alternative method of viability forecast summarizing the whole period of monitoring.

*Methods and Algorithms:* The alternative method relies on the original concept of pattern-multiplicative average for nonnegative matrices [3]. If  $\mathbf{x}(t)$  denotes the vector of population structure and  $\mathbf{L}(t)$  is the annual PPM, then 12 annual projections in the form of

$$\mathbf{x}(t+1) = \mathbf{L}(t)\mathbf{x}(t), t = 2009, \dots, 2020,$$

provide for the following projection from  $\mathbf{x}(2009)$  to  $\mathbf{x}(2021)$ :

$$\mathbf{x}(2021) = \mathbf{L}(2020)\mathbf{L}(2019) \dots \mathbf{L}(2009)\mathbf{x}(2009).$$

The logic behind the concept of *pattern-multiplicative average* suggests that the average matrix  $\mathbf{G}$  ought to do exactly the same projection, whereby we have

$$\mathbf{G}^{12} = \mathbf{L}(2020)\mathbf{L}(2019) \dots \mathbf{L}(2009), \quad (*)$$

and  $\mathbf{G}$  has the same zero/nonzero pattern as the annual PPMs. Consider this matrix equation as a system of scalar algebraic equations for the unknown elements of  $\mathbf{G}$ . Since the pattern of  $\mathbf{G}$  is nontrivial (it corresponds to the life cycle graph of organisms), the number of unknowns is substantially greater than the number of scalar equations in the system behind (\*), i.e., the system is *overdetermined*. Therefore, there is no exact solution to (\*), and we have to content with an approximate one. The algorithm reduces to solving the corresponding nonlinear constraint minimization problem for the approximation error, the constraints being determined by the set of annual PPMs. An eternal obstacle we encounter along this way is the local vs. global character of the

solution. What MATLAB<sup>®</sup> proposes to break through this obstacle is called GlobalSearch [4] and suggests certain adjustment of the local routine.

*Results:* The dominant eigenvalue of the pattern-multiplicative average matrix  $\mathbf{G}$  for *A. albana* has turned out equal  $\lambda_1(\mathbf{G}) = 0.83580$ , which is markedly less than that of the arithmetic mean of the annual PPMs [3]. Because of reproductive uncertainty in the data of *E. caucasicum*, we have  $\lambda_1(\mathbf{G}) \in [0.945004, 0.959403]$ , whereas the arithmetic mean interval goes beyond  $\lambda_1(\mathbf{G}) = 1$  by its right bound, thus leaving the viability forecast uncertain [3].

*Conclusion:* The pattern-multiplicative average has revealed that the local population of each species actually fights a battle for survival in the long term rather than rests in the alpine resort. This conclusion is consonant with those obtained by a traditional method within the paradigm of stochastic growth rate ( $\lambda_s$ ), albeit with realistic models of randomness correlated to variations in local weather indexes [5, 6]. Minor quantitative differences between  $\lambda_1(\mathbf{G})$  and  $\lambda_s$  can be attributed to the approximation error inherent in solving the averaging problem.

*Acknowledgements:* The study is supported by the Russian Scientific Foundation (22-24-00628).

### References

1. Tuljapurkar S.D. Population dynamics in variable environments. Springer New York, 1990.
2. Caswell H. Matrix population models: Construction, analysis and interpretation. 2nd ed. Sinauer Associates: Sunderland MA USA, 2001.
3. Logofet D.O. Does averaging overestimate or underestimate population growth? It depends. *Ecol Modell.* 2019;411:108744.
4. MathWorks. <https://www.mathworks.com/help/gads/globalsearch.html>. Accessed 28.03.2022.
5. Logofet D.O., Golubyatnikov L.L., Ulanova N.G. Realistic choice of annual matrices contracts the range of  $\lambda_S$  estimates. *Mathematics.* 2020;8(12):2252. doi: 10.3390/math8122252.
6. Logofet D.O. et al. "Realistic choice of annual matrices contracts the range of  $\lambda_S$  estimates" under reproductive uncertainty too. *Mathematics.* 2021;9(23):3007. doi: 10.3390/math9233007.

## Evolutionary dynamics of predator in a community of interacting species

Neverova G.P.<sup>1\*</sup>, Zhdanova O.L.<sup>1</sup>, Frisman E.Ya.<sup>2</sup>

<sup>1</sup> *Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*

<sup>2</sup> *Institute for Complex Analysis of Regional Problems, FEB RAS, Birobidzhan, Russia*

\* *galina.nev@gmail.com*

**Key words:** eco-genetic model with discrete time, diploid population, dynamics modes, multistability

*Motivation and Aim:* This paper proposes an ecological-genetic model of “predator-prey” community that explicitly describes the Mendelian inheritance mechanism for a trait of predator lifecycle. When modeling, we were motivated by the “arctic fox – mouse-like rodents” community which is an example of interspecies interaction of “prey – predator” type. To describe such a community, one has to consider along with ecological variables the predator genetic structure as well, as litter size being one of the most important traits of arctic fox lifecycle is genetically determined by a single diallelic locus [1]. Thus using a single locus diallelic model of inheritance is not unnatural for the considered case. With limited food resources, which corresponds to a low density of rodent population, pups' survival in large and small litters of Arctic foxes differs, therefore the considered locus determines both the birth rate of Arctic fox population and the survival rate of its offspring. At the same time, the survival rate of Arctic foxes, especially in large litters, significantly depends on prey population size as their food supply. To model it, we introduce a functional dependence of the fitness of predator genotypic groups on the prey population size instead of the constant ones used in classical papers [2, 3]. This study focuses on the explicit dependence of predator genotype composition and, accordingly, its birth rate on the prey abundance.

*Methods and Algorithms:* We propose a discrete-time model of predator-prey dynamics with Holling type II response function and Verhulst low of density limitation of each population in the community [4]. Predator evolution is considered by taking into account Mendelian inheritance of a lifecycle trait of the diploid predator by a single diallelic locus. In accordance with r-K-selection, the predator genotypes' fitnesses functionally depends on the prey abundance, thus its genotypes differing in reproductive potentials adapt unequally to food resource limitation. The proposed model is analytically and numerically studied, including research on stability, the construction of phase and parametric portraits, and bifurcation analysis. We use dynamic mode maps to study stability domains of fixed points, bifurcations, and evolution direction for the model.

*Results:* The proposed model demonstrates periodic, quasiperiodic and chaotic oscillations as well depending on the values of model parameters. Bistability of the monomorphic states is shown to appear provided a reduced fitness of the heterozygote, thus current genetic composition of the predator determines not only direction of its genetic evolution but the mode of population dynamics too. The model also demonstrates multistability with several alternative coexisting attractors, and initial conditions determine which of them attracts the system trajectory. Note that the attraction basins of coexisting monomorphic states can show fractal structure. Thus even little variation of predator or prey abundance due to accidental external influence can change the dynamics

mode of the community populations' numbers. Such a variation does not change the evolution direction, because it does not influence the predator genetic composition. However, a drastic change in the predator abundance with restructuring of its genetic structure can shift the evolution direction. The possibility of polymorphism existing at reduced heterozygote fitness with fluctuations of both the community populations' sizes and predator genetic composition turned out to be unexpected, because this phenomenon violates the classical evolutionary scenario. Such a polymorphism appears due to the genotype fitness dependence on a current prey abundance that can either stabilize or fluctuates about equilibrium changing the genotypes' fitnesses. With a heterogeneous predator in its fertility and adaptation to limited food resources, different phases of prey dynamics correspond to opposite direction of natural selection in the predator population. Thus fluctuating prey abundance can result in oppositely directed natural selection in the predator population that alters the ratio of its genotypes' fitnesses. Here one can see radically different scenarios of the predator evolution caused by various birth rate of the prey. Note that such a phenomenon is seemed to be characteristic for natural populations when more fecund genotypes of the predator require more food resources like in arctic fox populations, for instance. So, with a low birth rate in the prey population, eliminating of homozygotes with high reproductive potential and food requirements occurs in the predator population with monomorphism establishment to follow. A higher birth rate of the prey reverses the natural selection direction against those homozygotes that were winners in the previous conditions, and more fecund prey creates conditions for the predator to fix both its alleles. Such a polymorphism is characterized by long processes of elimination and accumulation of alleles without complete loss of any of them, in addition, one can see significantly different dynamics of the population sizes in the different phases of the process, which can be recognized by an observer as a change in the population dynamics mode of any species of the community.

*Conclusion:* Such a fluctuating polymorphism seems to exist in natural populations of inland arctic fox, because recessive inheritance type of small litters in this species excludes possibility of overdominance resulting in stable polymorphism, and only vast fluctuations of rodents' abundance being the main food of arctic foxes change natural selection direction in predator population and prevent loss of its genetic variety. At the same time, in the opposite dynamics phases of predator genetic composition, its consumption affects the prey population in different ways, which can already change the prey dynamics mode. It is an influence which can be a reason for the observed change in the dynamics mode in natural populations of rodents.

*Acknowledgements:* The work was completed within the frameworks of the State tasks of the Institute for Automation and Control Processes FEB RAS and the Institute for Complex Analysis of Regional Problem FEB RAS.

### *References*

1. Axenovich T.I. et al. Inheritance of litter size at birth in farmed arctic foxes (*Alopex lagopus*, Canidae, Carnivora). *Heredity*. 2007;98(2):99-105.
2. Kostitzin V.A. *Biologie Mathématique*. Librairie Armand Colin. Paris, 1937.
3. Pimentel D. Population regulation and genetic feedback. *Science*. 1968;159:1432-1437.
4. Neverova G.P., Zhdanova O.L., Frisman E.Y. Evolutionary dynamics of predator in a community of interacting species. *Nonlinear Dyn.* 2022;108:4557-4579. doi 10.1007/s11071-022-07372-z.

## Modeling dynamics of a plankton community using a discrete-time model

Neverova G.P.\*, Kan V.A., Zhdanov V.S., Zhdanova O.L.

*Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*

\* *galina.nev@gmail.com*

**Key words:** phytoplankton, zooplankton, predator-prey community, Arditi–Ginzburg response function, dynamics modes

*Motivation and Aim:* Modeling the dynamics of phytoplankton and zooplankton is an important part of mathematical ecology since plankton is a key component in aquatic ecosystems, as it is food for various animals: from invertebrates and crustaceans to fish and large marine mammals. This study proposes a two-component discrete-time model of a plankton community, describing the features of the development and interaction of phytoplankton and zooplankton. Reason for applying such models is the day-night rhythm. Indeed, many processes occurring in the plankton community are consistent with circadian rhythm, that is, cyclic fluctuations in the intensity of various biological processes due to the alternation of day and night. Most field observations and measurements of plankton have a day step. Also, discrete-time models allow us to describe a delay naturally.

*Methods and Algorithms:* To describe the dynamics of each species in the community, we use a discrete analog of the Verhulst equation, which allows us to take into account the autoregulation process. The trophic function covers a lower phytoplankton density due to its consumption by zooplankton. The growth and survival rates of zooplankton depend on the success of its feeding. Zooplankton death caused by its high density or a higher concentration of toxic substances released by phytoplankton is included in the limitation processes. The reproduction and survival of zooplankton depend on the abundance of food supply, namely phytoplankton density, which is described by the trophic function. We use the Arditi–Ginzburg response function based on the study [1], in which different trophic functions were identified for experimental estimates of two species of phytophagous rotifers (*Brachionus calyciflorus* and *Philodina acuticornis*). The proposed model is analytically and numerically investigated, including research on stability, the construction of phase and parametric portraits, and bifurcation analysis. The dynamic mode maps is used to study the dynamics modes of the model

*Results:* The study analyzes scenarios of the transition from stationary dynamics of community to fluctuations for various values of parameters determining the dynamics of phytoplankton and zooplankton and their interaction. The model analysis shows the stability loss of non-trivial fixed point, corresponding to the coexistence of species in the community, can occur via period-doubling bifurcation and the Neimark-Sacker one leading to the emergence of quasi-periodic fluctuations. In the multistability areas, dynamics mode shift is possible due to a change in the initial condition values. The proposed model of the plankton community reveals the occurrence of long-period fluctuations, which are the alternation of peaks and falls in the abundance of species as a result of the “predator-prey” interaction. Such behavior is in good agreement with the hypothesis that blooming species are those able to escape control by microzooplankton



through a combination of predation avoidance mechanisms at the beginning of the bloom [2]. Note that such dynamics occurs without artificially complicating the coefficients and the model [3].

*Conclusion:* Thus, despite its simplicity, the proposed discrete-time model of plankton community dynamics reveals adequate dynamics of interacting species. This model can be developed using other types of trophic functions with the expansion of the community components by including additional equations in the model.

*Acknowledgements:* The study is supported by Russian Science Foundation, project No. 22-21-00243, <https://rscf.ru/en/project/22-21-00243/>.

### *References*

1. Tyutyunov Yu.V. et al. Trophic function of phytophagous rotifers (rotatoria). Experiment and modeling. *Biol Bull Rev.* 2010;71(1):52-62. (In Russian)
2. Irigoien X., Flynn K.J., Harris R.P. Phytoplankton blooms: a 'loophole' in microzooplankton grazing impact? *J Plankton Res.* 2005;27(4):313-321.
3. Neverova G.P., Zhdanova O.L., Abakumov A.I. Discrete-time model of seasonal plankton bloom. *Math Biol Bioinf.* 2020;15(2):235-250. doi: 10.17537/2020.15.235. (In Russian)

## Influence of breeding strategy on genetic diversity of individuals

Poroshina A.A.

*Limnological Institute, SB RAS, Irkutsk, Russia*

\* *a.poroshina@lin.irk.ru*

**Keywords:** Microsatellite markers, object-oriented modeling, asexual reproduction, parthenogenesis

The most widespread method of asexual reproduction in animals is parthenogenesis, which is also found among highly organized organisms [1]. Even in stable natural populations of multicellular organisms, the phenomenon of the formation of offspring due to asexual reproduction is quite often encountered, when there is no recombination and unification of the genetic material of different organisms. This process is most widespread among plants and primitive animals (annelids, coelenterates, etc.). Often, the appearance of a new generation in such populations occurs according to a mixed type, some individuals are formed as a result of sexual reproduction, and some – during asexual reproduction. There are situations when populations, depending on environmental conditions, can switch from sexual reproduction to asexual reproduction and vice versa [2, 3]. In such cases, violations of the patterns of genetic diversity should be expected.

In populations with a sexual process with neutral molecular evolution – fixation of new alleles due to gene drift is established by the Hardy - Weinberg principle.

Hardy-Weinberg's principle is the position of population genetics, which states that in a population of infinitely large size, in which natural selection does not work, there is no mutation process, there is no exchange of individuals with other populations, there is no gene drift, all crosses are random – the frequencies of genotypes for which -or a gene (if there are two alleles of this gene in the population) will be kept constant from generation to generation and will correspond to equation [4].

The formation of genetic diversity in populations with sexual reproduction occurs due to the recombination and independent segregation of genetic material in several generations. This leads to the formation of new combinations of alleles. Some of the new alleles that have arisen as a result of spontaneous mutations are fixed or eliminated from the population due to gene drift and/or natural selection. In vegetatively propagating populations, the process of recombination and segregation of alleles is absent. For this reason, during vegetative reproduction, the genotype is transferred from one individual to another completely. In this case, genetic diversity is formed only due to the mutational process [5, 6] and competition/drift of clones. The mutation appears independently in different alleys.

In populations with a sexual process with neutral molecular evolution, fixation of new alleles occurs due to gene drift, and the equilibrium of frequencies by the allele is maintained following the Hardy-Weinberg principle.

Violation of this ratio, as a rule, indicates the impact on the population of any processes that dramatically change genetic diversity. Such processes can be: natural selection, fixing new alleys, expansion or decline in numbers ("bottleneck"), as well as migration. The bottleneck effect is common in many invertebrates because their number is greatly reduced when the temperature regime changes (autumn-winter period). The sharp drop

in numbers leads to significant difficulties in the analysis of various genetic markers. In highly organized animals, this effect can lead to the low viability of the species. Individuals also have to resort to inbreeding, and this, in turn, leads to inbred depressions within the species.

In our work, using a computer simulation model, we tried to investigate how the transition from sexual reproduction to vegetative reproduction will affect the population of diploid organisms with a neutral character of molecular evolution. At the same time, special attention was paid to the specificity of microsatellite markers. This specificity is a consequence of the multi-allele nature of microsatellite repeats and the high rate of their mutation.

### *References*

1. Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987;236:787-792.
2. Combosch D.J. Mixed asexual and sexual reproduction in the Indo-Pacific reef coral *Pocillopora damicornis*. *Ecol Evol*. 2013;3(10):3379-3387.
3. Delmotte F. Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Mol Ecol*. 2002;11(4):711-723.
4. Wigginton J.E. A note on exact tests of Hardy–Weinberg equilibrium. *Am J Hum Gen*. 2005;76(5):887-893.
5. Page R.D.M. Molecular evolution: A phylogenetic approach. Wiley-Blackwell, 2009.
6. Wang I.J. Isolation by environment. *Mol Ecol*. 2014;23(23):5649-5662.

## A heuristic approach to estimation of ecosystem stability

Saltykov M.\*, Bartsev S.

*Institute of Biophysics, SB RAS, Krasnoyarsk, Russia*

\* [saltykoff.mixail@yandex.com](mailto:saltykoff.mixail@yandex.com)

**Key words:** ecosystems stability, matrix stability

*Motivation and Aim:* Theoretical forecasting of ecosystem stability is important in our time. Stability of ecosystem means in general coexistence of all ecosystem species without extinction[1]. Stability can be predicted with mathematical model, however modern ecological models can be solved only numerical due to their high complexity.

*Methods and Algorithms:* In our work we establish a heuristic approach to estimate ecosystem model stability based on Gershgorin's Circles theorem. To check proposed approach, stochastic numerical study of models was used. Models were solved numerically many times with set of random parameters and then results are compared with heuristic method prediction.

*Results:* It was shown that there is a positive correlation between number of Gershgorin's Circles with centers on negative semi-plane and probability to have Lyapunov stable solution of equations. It is an important that Gershgorin's Circles theorem do not predict this result – to increase probability to have stable solution is not necessary to move all circles to negative semiplane.

*Conclusion:* The results obtained in the research can be widely used in modern ecology. Proposed heuristic approach can be used to predict mechanisms of ecosystems stability and “key species” with high role in stability maintenance.

### References

1. Ives A.R., Carpenter S.R. Stability and diversity of ecosystems. *Science*. 2007;317(5834):58-62. doi: 10.1126/science.1133258.

# Application of convolutional neural networks in the processing of hemispherical photos of the canopy

Shubin A.\*, Frolov P.

*Institute of Physicochemical and Biological Problems in Soil Science, Pushchino Center for Biological Research, RAS, Pushchino, Russia*

\* shanld@rambler.ru

**Key words:** Neural Networks, hemispherical photography, light, global light index

*Motivation and Aim:* The study of solar radiation transmittance in the forest canopy is necessary for a wide range of tasks, including identification of natural regeneration conditions, analysis of the biological cycle, analysis of the impact of global climate change on forest ecosystems [1, 2] and monitoring the condition of forest ecosystems [3]. Making direct measurements of light transmittance of the forest canopy is difficult due to the high level of spatial and temporal variability of the measured parameters [1]. Consequently, to estimate the solar radiation transmittance of the forest canopy, it is necessary to use indirect methods, one of which is hemispherical photography. The first stage of processing such photographs is binarization, i.e., separating the pixels related to the sky from all other pixels. The resulting black and white image is then used as a mask to calculate which fraction of incoming light passes through this mask, allowing the transmission of diffuse and direct radiation to be estimated.

In the classical approach used for binarization of images, only the blue RGB channel is analyzed. This is due to the fact that foliage elements have much lower reflectance and transmittance in the blue region of visible electromagnetic spectrum [4, 5, 3], while the brightness difference between clouds and blue sky in this region of the spectrum is minimal. Due to the fact that the manually selected threshold value can be a source of errors due to its subjectivity [6, 7], a large number of methods have been developed for automatic calculation of the binarization threshold. The most common in the processing of canopy photographs are the Ridler-Calvard algorithm [8] based on the image histogram analysis, and the Nobis-Hunziker algorithm [9] based on the edge detection. As these methods are based on pixel brightness analysis, they appear to be sensitive to photographic exposure. For example, in images taken with automatic exposure, these algorithms underestimate the fraction of gaps in low-density canopies, while overestimating the fraction of gaps in medium- and high-density canopies [5]. Currently, the use of convolutional neural networks (NN) in tasks related to image segmentation is widespread. This is due to the fact that this type of NN is characterized not only by the accuracy of image processing, but also by its stability in the presence of various image defects, such as extraneous noise, blurred borders, or improperly chosen exposure. However, the quality of such NN strongly depends on the volume and quality of the training data set, while the creation of a high-quality set of masks of hemispherical canopy photos is an extremely time-consuming process.

*Methods and Algorithms:* We developed a convolutional NN based on U-Net, whose architecture consists of a constricting path to capture the context and a symmetric expanding path that allows precise localization. To train the developed NN, a dataset consisting of only 240 hemispherical photographs taken in different stands was used,

whereas typically the minimum training sample size for the NN is 3000-5000 images. In order to compensate for the small number of images used in training, we applied a data augmentation image preprocessing technique, which included random crop to obtain a random image fragment of a given dimensionality, as well as contrast and brightness changes. This allowed us to qualitatively expand the training dataset and avoid overtraining the model. To validate the proposed NN at the points of continuous monitoring of photosynthetically active radiation, we obtained hemispherical photographs of the canopy taken at different exposures. Binary masks were created from these photos using the classical approach (the Nobis-Hunziker algorithm) and the developed NN, and then the weighted average radiation transmittance global light index was calculated from these masks, which was compared with the measured data.

*Results and Conclusion:* The validation showed that the use of the proposed approach allowed us to qualitatively train the developed NN for the binarization of hemispherical photographs of the canopy on a small training sample. The developed NN turned out to be stable to varying exposure of photographs and blurring of image edges arising as a result of lens distortions. At the same time, it should be noted that in the presence of the Sun in the frame, there are extensive areas in which the brightness of all pixels is maximal. Consequently, such images cannot be correctly binarized either by classical methods or by methods based on the application of NN.

*Acknowledgements:* The study is supported by the Institute of Physicochemical and Biological Problems in Soil Science, Pushchino Center for Biological Research of the Russian Academy of Sciences (122040.500037-6.) and the RSF, project 18-14-00362.

### References

1. Chen J.M., Rich P.M., Gower S.T., Norman J.M., Plummer S. Leaf area index of boreal forests: Theory, techniques, and measurements. *J Geophys Res Atmos.* 1997;102(D24):29429-29443.
2. Macfarlane C., Hoffman M., Eamus D., Kerp N., Higginson S., McMurtrie R., Adams M. Estimation of leaf area index in eucalypt forest using digital photography. *Agric For Meteorol.* 2007;143(3):176-188.
3. Macfarlane C. Classification method of mixed pixels does not affect canopy metrics from digital images of forest overstorey. *Agric For Meteorol.* 2011;151(7):833-840.
4. Leblanc S.G., Chen J.M., Fernandes R., Deering D.W., Conley A. Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agric For Meteorol.* 2005;129(3-4):187-207.
5. Zhang Y., Chen J.M., Miller J.R. Determining digital hemispherical photograph exposure for leaf area index estimation. *Agric For Meteorol.* 2005;133(1-4):166-181.
6. Cescatti A. Indirect estimates of canopy gap fraction based on the linear conversion of hemispherical photographs: Methodology and comparison with standard thresholding techniques. *Agric For Meteorol.* 2007;143(1-2):1-12.
7. Jarčuška B., Kucbel S., Jaloviar P. Comparison of output results from two programmes for hemispherical image analysis: Gap Light Analyser and WinScanopy. *For Sci.* 2010;56(4):147-153.
8. Ridler W., Calvard S. Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cyber.* 1978;8:260-263.
9. Nobis M., Hunziker U. Automatic thresholding for hemispherical canopy-photographs based on edge detection. *Agric For Meteorol.* 2005;128(3-4):243-250.

# Genetic polymorphism under cyclical selection in long-lived species: the complex effect of age structure and maternal selection

Zhdanova O.<sup>1\*</sup>, Frisman E.<sup>2</sup>

<sup>1</sup> *Institute for Automation and Control Processes, FEB RAS, Vladivostok, Russia*

<sup>2</sup> *Institute for Complex Analysis of Regional Problem, FEB RAS, Birobidzhan, Russia*

\* *axanka@iacp.dvo.ru*

**Key words:** maternal selection, cyclical selection, genetic polymorphism, age structured population, diploids

*Motivation and Aim:* The maintenance of genetic polymorphism in heterogeneous environments continues to be a topic of research interest in evolutionary genetics for a long time. Storage effect is known [1] to maintain genetic polymorphism in fluctuating environments via buffering factors limiting the impact of negative selection in temporarily unfavorable environments, thereby avoiding allele extinction. A major buffering factor is overlapping generations due to age (stage)-specific selection. Another type of buffering factors that promote the maintenance of polymorphism are genetic ones, such as natural selection on sex-limited traits, maternal (or paternal) selection, and various maternal genetic effects as well (i.e. genomic imprinting or delayed inheritance) [2]. Despite this framework has a long history and the significant progress has been achieved in understanding the conditions for maintaining genetic variety, the complex effect of the principal buffering factors under study: overlapping generations and genetic (in particular maternal selection) has not been discussed in details. At that, in populations of long-living species, both of these factors act simultaneously and the result of their joint action is not obvious.

Our study basically tries to model a situation demonstrated by arctic fox populations, where the litter size being a major life history trait is a sex-limited female trait [3]. It is influenced by maternal selection with cyclic fluctuations because of oscillations in food abundance [4]. Note, arctic fox is a long-lived species having an age structure. An evolutionary mechanism that promote the maintenance of genetic variation in litter size of this species is not entirely clear.

*Methods and Algorithms:* We consider the maintenance of polymorphism under the complex influence of maternal selection, age structure, and fluctuating environments. We also consider two models of natural selection in a population with and without age structure [5] and the model of maternal selection in a population without age structure [6] as reference points for the age-structured model with maternal selection.

*Results:* For a population with non-overlapping generations, the analysis results of the influence of different cycle lengths and the ratio of reproductive potentials of genotypes on maintaining polymorphism turned out to be quite expected. Maternal selection always dilates polymorphism region vs. classic selection. The longer an external cycle, the narrower polymorphism region, whereas the greater difference in genotypes' reproductive potentials, the wider polymorphism region. The complex effect of maternal selection and overlapping generations happened to be as follows [7]: overlapping

generations always tend to dilate the polymorphism region, odd and even external cycles produce various types of polymorphism regions. Particularly, odd external cycles expand but only slightly the coexistence region at low survival of reproductive individuals; with increasing survival of reproductive individuals, this region expands together with gradual decreasing extension of the region due to maternal selection. On the contrary, even cycles produce the greatest extension of the polymorphism region at low survival of reproductive individuals, at that, with large difference in reproductive potential of genotypes maternal selection dilates the region of alleles' coexistence, but classic selection becomes more effective in maintaining polymorphism with low values of this parameter.

*Conclusion:* Maternal selection [2] and overlapping generations [8] can facilitate the stable coexistence of alleles under temporally fluctuating environment. Using mathematical models, we considered the complex effect of these two factors on the maintenance of genetic polymorphism in cyclically changing environments. We focused on asymmetric cyclic selection, which allows us to describe fluctuations of environments similar to food resources cycles with rare peaks and prolonged decline of prey abundance. Such environment provides conditions for r-K selection, at which some genotypic groups rapidly propagate and therefore exhibit a high reproductive potential r (r-strategy), whereas other groups adapt to limited resources and present a high capacity of the ecological niche K (K-strategy) [9, 10]. With the asymmetric external cycle, dominant genotypes that produce more offspring most often concede to recessive genotypes with low reproductive potential to be able to adapt to limited food resources. One can observe similar situation in arctic foxes we focused on in our models. The obtained results showed that compared with the simple Mendelian inheritance in the classic model, maternal selection and overlapping generations increases the chances of maintaining polymorphism in populations of arctic foxes.

*Acknowledgements:* The work was completed within the frameworks of the State tasks of the Institute for Automation and Control Processes FEB RAS and the Institute of Complex Analysis of Regional Problem FEB RAS.

## References

1. Hedrick P.W. Genetic polymorphism in a temporally varying environment: effects of delayed germination or diapause. *Heredity*. 1995;75:164-170.
2. Yamamichi M., Hosono M. Roles of maternal effects in maintaining genetic variation: Maternal storage effect. *Evolution*. 2017;71(2):449-457.
3. Axenovich T.I. et al. Inheritance of litter size at birth in farmed arctic foxes (*Alopex lagopus*, Canidae, Carnivora). *Heredity*. 2007;98:99-105.
4. Volodin I.A. et al. Structure of arctic fox (*Alopex lagopus beringensis*) colonies in the northern extremity of Bering Island. *Biol Bull*. 2013;40:614-625.
5. Zhdanova O.L., Frisman E.Ya. Mathematical modeling of the mechanism of a reproductive strategies differentiation in natural populations (on the example of arctic fox, *Alopex lagopus*). *Comput Res Model*. 2016;8(2):213-228. (in Russian)
6. Zhdanova O.L., Frisman E.Ya. Mathematical modeling of selection by sex-limited trait: to the question of maintenance of litter size polymorphism in natural populations of arctic foxes. *Rus J Genet*. 2021;57(2):227-237.
7. Zhdanova O.L., Frisman E.Ya. Genetic polymorphism under cyclical selection in long-lived species: the complex effect of age structure and maternal selection. *J Theor Biol* 2021;512:110564.
8. Turelli M., Schenk D.W., Bierzychudek P. Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: protecting the blues in *Linanthus parryae*. *Evolution*. 2001;55:1283-1298.
9. Pianka E.R. *Evolutionary Ecology*, 2nd ed. Harper and Row. New York, 1978.
10. Frisman E.Y. et al. Mathematical modeling of population dynamics based on recurrent equations: results and prospects. Part II. *Biol Bull Russ Acad Sci*. 2021;48:239-250.



# Zooplankton effect on the dynamics of competing phytoplankton species

Zhdanova O. \*, Zhdanov V., Kan V., Neverova G.

*Institute for Automation and Control Processes, FEB RAS, Vladivostok, Russia*

\* [axanka@iacp.dvo.ru](mailto:axanka@iacp.dvo.ru)

**Key words:** community dynamics, Ricker's model, phytoplankton, zooplankton, predator-prey interaction, competing phytoplankton populations, multistability

*Motivation and Aim:* The phenomenon of phytoplankton blooming expressed as a dramatic increase in phytoplankton density is currently under active research. At that toxic blooming negatively affects aquaculture, coastal tourism, and human health. The most interesting results in modeling phytoplankton bloom were obtained based on a modification of the classical system of phytoplankton and zooplankton interaction. The modifications use delayed equations, as well as piecewise continuous functions with a delayed response to intoxication processes, making it possible to obtain adequate phytoplankton dynamics like in nature. Recurrent equations, which are widely used in modeling populations and communities [e.g., reviews 1, 2] and which allow us to describe lag effects naturally, are practically not used for modeling plankton communities' dynamics. Using the classical modeling technique, we propose a discrete-time three-component model of the plankton and zooplankton community. The community consists of zooplankton and two phytoplankton species competing for resources: toxic and non-toxic. This work continues our previous study [3], where a simple model of a phytoplankton-zooplankton community consisting of two equations with discrete-time has been proposed.

*Methods and Algorithms:* We propose a three-component model of the phytoplankton-zooplankton community with discrete-time, in which toxic and non-toxic species of phytoplankton compete for resources. We use the Holling functional response of type II to describe an interaction between zoo- and phytoplankton. Also, we use the Ricker competition model to describe the restriction of phytoplankton biomass growth by the availability of external resources (mineral nutrition, oxygen, light, etc.). Many phytoplankton species, including diatom algae, are known not to release toxins if they are not damaged. Zooplankton pressure on phytoplankton is shown to decrease in the presence of toxic substances. For example, Copepods are selective in their food choices and avoid consuming toxin-producing phytoplankton [4]. Therefore, in our model, zooplankton (predator) consumes only non-toxic phytoplankton species being prey, and toxic species phytoplankton only competes with non-toxic for resources. We carried out an analytical and numerical study of the proposed model, including stability analysis, the construction of phase and parametric portraits, and bifurcation analysis. We use dynamic mode maps to study stability domains of fixed points, bifurcations, and the evolution of the community.

*Results:* Based on methods of mathematical modeling we estimated a zooplankton influence on the dynamics of phytoplankton species coexistence. The study analyzed scenarios of the transition from stationary dynamics to fluctuations in the size of phytoplankton and zooplankton for various values of intrapopulation parameters

determining the nature of the dynamics of the species constituting the community, and the parameters of their interaction. The focus is on exploring the complex modes of community dynamics. In the framework of the model used for describing the dynamics of the phytoplankton community in the absence of interspecific interaction, phytoplankton dynamics undergoes a series of period-doubling bifurcations. At the same time, with zooplankton appearance, the cascade of period-doubling bifurcations in phytoplankton and the community as a whole is realized earlier (at lower reproduction rates of phytoplankton cells) than in the case when phytoplankton develops in isolation. The proposed dynamic model of the phytoplankton and zooplankton community allows us to observe long-period oscillations, which is consistent with the results of field experiments. As well, the model has multistability areas, where a variation in initial conditions with the unchanged values of all model parameters can result in a shift of the current dynamic mode.

*Conclusion:* In the absence of zooplankton, the proposed model transforms to the two-species competition model proposed by A.P. Shapiro [5] in 1972 and by R. May [6] in 1974 yrs. Detailed study of this model showed [7] that the dynamics of two species competing for a resource can be extremely various and complicated having multistability regions as well. With strong interspecific competition of phytoplankton species for resources, both the ratio of reproduction rates and initial conditions determine their survival. As a result, zooplankton removing a part of one phytoplankton species abundance changes the current population sizes of the competitors thus allowing the population with the lower reproductive potential to eliminate the more fecundate competitor. With weak interspecific competition of phytoplankton species and sufficiently high growth rates of their populations, it is difficult to predict the scenario of community development, although the species will coexist. The zooplankton changing the initial condition can shift the dynamic mode and oscillation phase observed in the phytoplankton community.

*Acknowledgements:* The work was financially supported by Russian Science Foundation, project No. 22-21-00243, <https://rscf.ru/en/project/22-21-00243/>.

### References

1. Frisman E.Ya. et al. The key approaches and review of current researches on dynamics of structured and interacting populations. *Comp Res Model.* 2019;11(1):19-151. (In Russian)
2. Frisman E.Y. et al. Mathematical modeling of population dynamics based on recurrent equations: results and prospects. Part II. *Biol Bull Russ Acad Sci.* 2021;48(3):239-250.
3. Neverova G.P., Zhdanova O.L., Abakumov A.I. Discrete-time model of seasonal plankton bloom. *Math Biol Bioinf.* 2020;15(2):235-250. doi: 10.17537/2020.15.235. (In Russian)
4. DeMott W.R., Moxter F. Foraging on cyanobacteria by copepods: responses to chemical defenses and resource abundance. *Ecology.* 1991;72:1820-1834.
5. Shapiro A.P. Discrete model of competition between two populations. *DAN USSR.* 1974;218:699-701. (In Russian)
6. Kulakov M., Neverova G., Frisman E. The Ricker competition model of two species: dynamic modes and phase multistability. *Mathematics.* 2022;10(7):1076. doi: 10.3390/math10071076.
7. May R.M. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science.* 1974;186(4164):645-647.

## Study, storage and visualization of data on the ontogenesis of dwarf shrubs and herbaceous plants (CAMPUS program)

Zubkova E.V.\* , Frolov P.V.

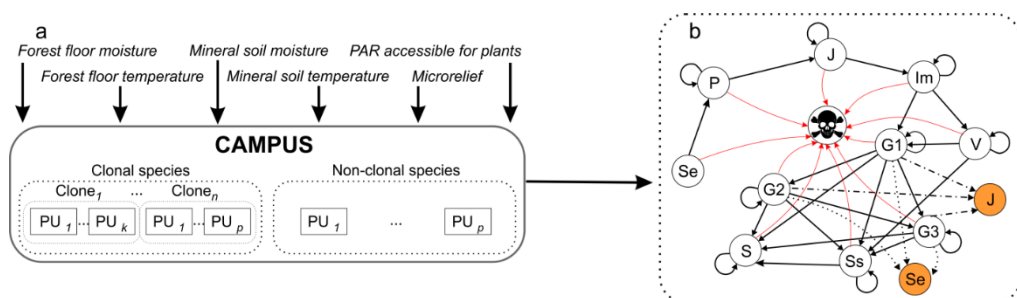
*Institute of Physicochemical and Biological Problems in Soil Science, Pushchino Center for Biological Research, RAS, Pushchino, Russia*

\* elenazubkova2011@yandex.ru

**Key words:** forest, plant growth, plants of ground cover, simulation modeling

**Motivation and Aim:** Geobotanists have described more than 2000 plant ontogenies (the period of development from zygote to death). The concept of discrete description of individual plant development was formulated in the middle of the twentieth century [1, 2]. Studies have also been conducted on the modularity of plant structure [3, 6]. The combination of these two approaches with the modeling of cellular automata has become the basis of a mathematical model that allows you to visualize the development and predict the timing of the occupation of the territory by plant cenopopulations and predict interactions between different plants.

**Methods and Algorithms:** The CAMPUS model (Cellular Automata Model of Plants' United Spread) is an individual-based and spatially-explicit discrete simulation model [4, 5]. Since the model is intended to be able to treat a dynamics of population of as clonal plants as well as seeds plants we use a term "plant unit" (PU) for either a partial formation (i.e. a shoot together with rhizome) or for an individual plants (Figure). To be realistic enough for many species of forest ground vegetation the spatial resolution of CAMPUS model should be no less than  $1 \times 1$  cm. In our program it is limited only by computer capacity. The time step of the model is 1 calendar month which is appropriate to represent all growth phases of plants. The features of the model are as follows: plant ontogenesis is encoded by 10 ontogenetic states; plants of different stages and different life forms may have different schemes of morphological structures, both aboveground and underground (rhizomes, roots); morphological structures may vary in size and shape, probabilities of dying off or rejuvenate during ontogenesis.



**Fig. 1.** CAMPUS model organization diagram (a) and the flowchart of the CAMUS-S model individual step (b). Se, P, J, Im, V, G1, G2, G3, Ss, and S are PU development stages which in the model are based on ontogeny theory. Solid lines show possible transitions between ontogenetic stages, dotted lines show seed reproduction, dotted-dashed lines show vegetative reproduction

*Results:* The parameterization of the model has been carried out for eight perennial species that dominate coniferous and coniferous-deciduous forests. The species belong to different life-form: deciduous dwarf shrub *Vaccinium myrtillus* L., evergreen dwarf shrub *Vaccinium vitis-idaea* L., Wintergreen rhizomatous summer green plant *Convallaria majalis* L.; rhizomatous winter green plants: *Aegopodium podagraria* L., *Carex pilosa* Scop., *Oxalis acetosella* L.; rhizomatous fern *Pteridium aquilinum* (L.) Kuhn ex Decken; bunch grass *Calamagrostis arundinacea* (L.) Roth. Data about primroses: *Anemone ranunculoides* L., *Corydalis solida* (L.) Clairv. and other plants were collected.

*Conclusion:* The CAMPUS-S computer program allows to represent detailed information about the morphology of plants (height of shoots, diameter of the aboveground part, length of rhizomes, number and frequency of their branching), time indicators of ontogenetic development (duration of stages), flexibility of growth patterns. Such data make it possible to model the spatial development of plants and predict competition for the space of cenopopulations of different species.

*Acknowledgements:* This study was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (theme No. 122040.500037-6). The research on determination of heterogeneity of soil conditions was supported by RSF (project No. 18-14-00362-P).

### References

1. Rabotnov T.A. The life cycle of perennial herbaceous plants in meadow cenosis. In: Tr. Bin. USSR Academy of Sciences, Geobotany. M.-L., 1950;3(6):3-204. (In Russian)
2. Uranov A.A. Age spectrum of phytocenopopulation as a function of time and energy wave processes. *Biol Sci.* 1975;2:7-34. (In Russian)
3. Zaugolnova L.B., Zhukova L.A., Komarov A.S., Smirnova O.V. Cenopopulations of plants: article of population biology. M.: Nauka, 1988. (In Russian)
4. Frolov P.V., Shanin V.N., Zubkova E.V., Bykhovets S.S., Grabarnik P.Ya. CAMPUS-S – The model of ground layer vegetation populations in forest ecosystems and their contribution to the dynamics of carbon and nitrogen. I. Problem formulation and description of the model. *Ecoll Modell.* 2020;431:109184. doi: 10.1016/j.ecolmodel.2020.109184.
5. Frolov P.V., Zubkova E.V., Shanin V.N., Bykhovets S.S., Mäkipää R., Salemaa M. CAMPUS-S – The model of ground layer vegetation populations in forest ecosystems and their contribution to the dynamics of carbon and nitrogen. II. Parameterization, validation and simulation experiments. *Ecoll Modell.* 2020;431:109183. doi: 10.1016/j.ecolmodel.2020.109183.
6. Harper J.L. Population biology by plants. Acad. Press., 1977.

# 3 Symposium “Structural biology and pharmacology: computational and experimental approaches”



- 3.1 Section “The role of synchrotron radiation and advanced instrumental techniques for macromolecular crystallography and pharmacology” [261](#)
- 3.2 Section “Structural biology of proteins and membranes” [266](#)
- 3.3 Section “Chemoinformatics and chemical biology” [340](#)

## Protein-lipid nanoparticles for studying G-protein coupled receptors functional properties

Khorn P.A. \*, Luginina A.P., Gusach A.Yu., Borshchevskiy V.I., Mishin A.V.

*Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow University of Physics and Technology (NRU), Dolgoprudny, Russia*

\* *horn\_pa@mail.ru*

**Keywords:** membrane modeling media, GPCR, functional activity

*Motivation and Aim:* G-protein coupled receptors (GPCRs) constitute the largest superfamily of membrane proteins in the human genome and are therapeutic targets for over 30 % of drugs on the pharmaceutical market. [1]. Structural and biophysical research of GPCRs is aimed at finding the mechanisms of ligand recognition and signaling for this protein superfamily. Such studies are extremely challenging due to the low stability of GPCRs outside their natural membrane environment. Detergent micelles and nanodiscs based on derivatives of human apolipoprotein A-1 are the most common membrane mimetic systems (MMS) and are widely used for structural studies, as well as for studying the functional activity of GPCRs and other membrane proteins by NMR, single molecules fluorescence, mass spectrometry, *etc.* Nevertheless, the choice of the optimal MMS for various applications is still an actual task [2].

*Methods and Algorithms:* First of all, MMS must satisfy a number of criteria such as being compatible with the procedure for isolating and purifying the protein under study, thermally stabilizing, and allowing investigation of the structural and functional properties for the protein, along with retaining protein native properties. These requirements can be met by a new generation of MMS – nanodiscs based on the lysosomal protein saposin A (Salipro®) and nanodiscs based on a synthetic diisobutylene-maleic acid copolymer (DIBMALP). In case of Salipro® and DIBMALP, the formation of stabilized in solution protein-lipid nanoparticles occurs directly from the membrane, which qualitatively distinguishes these systems from all others and makes it possible to study GPCRs in presence of lipids surrounding them in the native membrane.

*Results:* In this work, we performed a comparative analysis of the stability and tendency to aggregation for the human A<sub>2A</sub> adenosine receptor and cysteinyl leukotriene receptors (CysLT<sub>1</sub>R, CysLT<sub>2</sub>R) embedded in new and classic MMS.

*Conclusion:* The results obtained will allow to define MMS, the most suitable for physical and chemical studies of GPCRs *in vitro*.

*Acknowledgements:* The work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-00337-20-03, project FSMG-2020-0003).

### References

1. Shchepinova M.M. et al. Chemical biology of noncanonical G protein-coupled receptor signaling: Toward advanced therapeutics. *Curr Opin Chem Biol.* 2020;56:98-110.
2. Majeed S. et al. Lipid membrane mimetics in functional and structural studies of integral membrane proteins. *Membranes (Basel).* 2021;11(9):685.

## $\beta$ -turns parameters refinement using neutron diffraction data

Korobkov A.A.<sup>1</sup>, Khurmuzakiy A.A.<sup>2</sup>, Esipova N.G.<sup>1</sup>, Tymanyay V.G.<sup>1</sup>,  
Anashkina A.A.<sup>1, 2</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> I.M. Sechenov First Medical University (Sechenov University), Moscow, Russia

**Key words:**  $\beta$ -turn, conformation, polypeptide chain,  $\beta$ -hairpin

*Motivation and Aim:* Beta-turns are an insufficiently studied object in proteins, since the stabilization mechanism of energetically unfavorable conformation still remains unclear. Classification of beta-turns usually carried out according to conformational angles  $\phi$  and  $\psi$  of amino acid residues  $i + 1$  and  $i + 2$ . However, analysis of the Ramachandran map of amino acid residues  $i + 1$  and  $i + 2$  indicates a high conformational stress arising in the turn, which should be compensated by additional interactions. We assume that in order to stabilize the beta-turn, an additional hydrogen bond must exist, the energy of which compensates for the stress of the Beta-turn. This additional hydrogen bond couldn't be taken into account because of X-ray methods of protein analysis used to determine prior beta-turns classifications. Our main aim was to analyze protein structures determined by neutron diffraction and cluster beta-turns conformational angles with satisfying criteria for hydrogen bond.

*Methods and Algorithms:* We used PDB database to collect all protein structures which were determined by neutron diffraction (as of November 2020). To calculate hydrogen bonds in protein structures and  $\beta$ -turns HBPLUS software was used [2]. The turns were selected according to the Venkatachalam's criteria for a  $\beta$ -turns of the presence of a closing hydrogen bond between the oxygen of the main chain of the  $i$ -th amino acid residue and the nitrogen of the main chain of the  $i+3$  residue [6]. This is a more rigorous and unambiguous approach to identifying  $\beta$ -turns than estimating the distance between  $C\alpha$  atoms. In addition to the four amino acid residues that form the turn itself, two flanking residues were also considered to the right and left of the turn. Thus, fragments of eight amino acid residues containing a kink with a closing hydrogen bond in the middle were selected. A total of 3717 such fragments were found, i.e. an average of 22 turns in one structure. The conformational angles  $\phi$ ,  $\psi$ , and  $\omega$  of the six central residues were determined in the fragment structure.

*Results:* It was found that all 3717 turns can be easily classified by  $\phi$ ,  $\psi$  angles of  $i + 1$  and  $i + 2$  residues of the turn into eight classes only. 4 of them correspond to classical beta turn types, so the central values  $\phi$ ,  $\psi$  of type I, type I', type II and type II' have been refined. The other four classes of peptides have corresponding classes in Shapovalov [5] and de Brevern [1] classifications. We did not find additional stabilizing bifurcated hydrogen bond formed between the CO oxygen atom of the  $i$ -th residue group and the nitrogen atom of the  $i$ -th residue + 2.

*Conclusion:* These results indicate the absence of additional stabilization of the strained conformation of the beta-turn at the level of the protein secondary structure by the bifurcated hydrogen bond. Forced-conformation segments, taken by themselves, must be conformationally labile. Indeed, they cannot have a preferred conformation that is the same as the one they should adopt in the protein, since this conformation is met with

stereochemical obstacles, but it is better if this segment has a whole set of conformations (with favorable energy, of course). The most unfavorable case is when the fragment has one conformation, moreover, with a favorable energy. Then it is especially difficult for external factors to give this part of the protein chain the required shape. Based on the foregoing, one can expect increased conformational freedom from beta-turns.

*Acknowledgements:* This research has been supported by Russian Science Foundation Grant No. 22-24-00936.

### *References*

1. de Brevern A. Protein local conformations at the light of a structural alphabet. *Biophys J.* 2018;14(3):231a.
2. Hutchinson E.G., Thornton J.M. A revised set of potentials for beta-turn formation in proteins. *Protein Sci Pub Protein Soc.* 1994;3:2207-2216.
3. Lewis P.N. et al. Chain reversals in proteins. *Biochim Biophys Acta Proteins Proteom.* 1973;303(2):211-229.
4. Richardson J.S. The anatomy and taxonomy of protein structure. *Adv Protein Chem.* 1981;34:167-339.
5. Shapovalov M. et al. A new clustering and nomenclature for beta turns derived from high-resolution protein structures. *PLOS Comput Biol.* 2019;15:e1006844.
6. Venkatachalam C.M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers.* 1968;6:1425-1436.
7. Wilmot C.M., Thornton J.M. Analysis and Prediction of the Different Types of B-Turn in Proteins. *J Mol Biol.* 1988;203(1):221-32
8. Wilmot C.M., Thornton J.M.  $\beta$ -Turns and their distortions: a proposed new nomenclature. *Protein Eng.* 1990;3:479-493.



## Overview of lipid-sensing GPCR receptors structure and function

Mishin A.<sup>1\*</sup>, Luginina A.<sup>1</sup>, Gusach A.<sup>1</sup>, Lyapina E.<sup>1</sup>, Marin E.<sup>1</sup>, Safronova N.<sup>1</sup>, Khorn P.<sup>1</sup>, Shevtsov M.<sup>1</sup>, Borshchevskiy V.<sup>1</sup>, Cherezov V.<sup>1,2</sup>

<sup>1</sup>Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia

<sup>2</sup>Bridge Institute, Department of Chemistry, University of Southern California, Los Angeles, USA

\*mishinalexej@gmail.com

**Key words:** G Protein Coupled Receptors (GPCRs), lipidic ligands, cysteinyl leukotriene receptors (CysLT1,2), S1P<sub>5</sub>, X-ray structure, inverse agonist

**Motivation and Aim:** G protein-coupled receptors (GPCR) are the most important class of membrane proteins in the human body, responsible for carrying out a wide range of physiological functions, such as maintaining homeostasis of cell metabolism, modulating the transmission of nerve impulses and immune system activity, visual and olfactory perception, regulating processes in the cardiovascular system.

Endogenous ligands for GPCR represent a wide spectrum of chemical entities - amino acids and ions, proteins and peptides, nucleotides and biogenic amines, bioeffector lipids. We are aimed at structural and functional studies of lipid GPCRs that regulate various processes in the central nervous system (CNS). Examples of lipids that are ligands for GPCR are lysophospholipids, fatty acids, endocannabinoids, prostaglandins, sphingolipids, eicosanoids, leukotrienes, and other arachidonic acid derivatives.

Structural studies combined with extensive biophysical, biochemical, and pharmacological characterization help to develop a better understanding of these receptors' molecular mechanism of action and initiate rational drug development.

**Methods:** X-Ray Crystallography at microfocus beamlines of synchrotron sources and X-Ray Free Electron Lasers combined with functional assays, structure-activity relationship studies, molecular modeling and docking.

**Results:** In this report we will discuss the structural and functional peculiarities of lipid-sensing GPCRs, including the three representatives, for which the structure was recently obtained by our team: CysLT1 [1], CysLT2 [2], S1P<sub>5</sub> [3].

**Conclusions:** Further complex structural and functional studies of lipid-sensing receptors need to be carried out to find answers to fundamental questions about the molecular mechanisms of their functioning and the rational creation of drugs. A lot of works have appeared that expand the understanding of their biomedical significance, and new methods and approaches in research have been developed.

**Acknowledgements:** The work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-00337-20-03, project FSMG-2020-0003).

### References

1. Luginina A. et al. Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs. *Sci Adv.* 2019;5(10):eaax2518. doi: 10.1126/sciadv.aax2518.
2. Gusach A. et al. Structural basis of ligand selectivity and disease mutations in cysteinyl leukotriene receptors. *Nat Commun.* 2019;10:5573. doi:10.1038/s41467-019-13348-2.
3. Lyapina E. et al. Structural basis for receptor selectivity and inverse agonism in S1P<sub>5</sub> receptors. *bioRxiv.* 2022;PPR462137. doi: 10.1101/2022.02.25.480536.

## 3.2 Section “Structural biology of proteins and membranes”



# A model of SARS-CoV-2 SPIKE transmembrane domain linked to THE HR2 region: structural organisation and possible role in membrane fusion

Aliper E.T.<sup>1\*</sup>, Krylov N.A.<sup>1,2</sup>, Nolde D.E.<sup>1,2</sup>, Polyansky A.A.<sup>1,3</sup>, Efremov R.G.<sup>1,2,4</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> National Research University Higher School of Economics, Moscow, Russia

<sup>3</sup> Department of Structural and Computational Biology, Max Perutz Labs, Vienna, Austria

<sup>4</sup> Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia

\* [la.marelle@gmail.com](mailto:la.marelle@gmail.com)

The spike (S) protein of SARS-CoV-2 effectuates membrane fusion and virus entry into target cells. Its transmembrane domain (TMD) is a homotrimer of  $\alpha$ -helices anchoring the spike in the viral envelope. Although S-protein models available to date include TMD, its precise configuration has only been given brief consideration. Understanding viral fusion entails realistic TMD models, while no reliable approaches towards predicting the 3D structure of transmembrane (TM) trimers exist. Immediately upstream of the TMD is a domain known as HR2. Apart from being involved in the complex refolding of the spike, HR2 is believed to interact with the viral envelope, as it is brought into contact with the target membrane, which has been observed for HIV's protein gp41 from the same family of fusion proteins. Here, we adopted diverse computational tools to model the spike TMD (S-TMD) based solely on its primary structure and used the resulting model to study a larger element of the fusion machinery, TMD linked to HR2, i.e., an anchor plus a “membrane-fixing factor”.

We performed amino acid sequence pattern matching and compared molecular hydrophobicity potential (MHP) distribution on the helix surface against TM homotrimers with known 3D structures. Eventually, the TMD of the tumour necrosis factor receptor 1 (TNFR-1) was selected for template-based modelling. Adjusting so-called “dynamic MHP portraits” and residue variability motifs, we iteratively built an all-atom homotrimer model of the S-TMD, whereof each helix possessed two overlapping interfaces interacting with either of the remaining helices. The interfaces included conservative residues like I1216, F1220 and M1229. The stability of this model was tested in all-atom molecular dynamics (MD) simulations in a POPC bilayer mimicking the viral envelope and compared to several alternative configurations, including a recent NMR structure of a trimerised peptide.

Our model trimer remained extraordinarily tightly packed over a microsecond-range MD and retained its stability when palmitoyl chains were added at cysteine residues located downstream in accordance with experimental data. Overall, the resulting model of S-TMD conforms to the known basic principles of TM helix packing.

S-TMD was then linked to the HR2 region, modelled based on an NMR structure of the same domain in SARS-CoV's spike. Given that HR2 is a canonical coiled with an angle between helical axes of  $\sim 17^\circ$ , whilst in our model trimer this angle amounted to  $\sim 41^\circ$ , a hinge area between the two domains was required, implying they are not part of a continuous coiled coil encompassing both domains. This is in agreement with the predicted functional versatility of the N-terminal aromatic-rich region in the TMD and C-terminal residues in HR2, which would also entail structural flexibility. Models of

monomeric and trimeric HR2 both on its own and linked to the TMD were subjected to MD simulations to evaluate the behaviour and mutual influence of the two domains when part of one molecule, possibly relevant to understanding viral fusion. Trimeric HR2 remained intact in water, and so did trimeric HR2 linked to TMD in a model POPC bilayer: throughout the MD simulation, both domains remained stable. Previously SARS-CoV-2's spike HR2 had been experimentally demonstrated to have limited affinity for lipid bilayers, and our MD simulations agree with these observations: on its own, monomeric HR2 did not remain membrane-bound. Linking it to a monomer of TMD, on the other hand, resulted in more frequent contacts with the bilayer. In both cases, HR2 tended to form hairpin-like structures, intramolecular contacts seemingly preferable to contacts with the membrane. Thus, even though the TM anchor significantly increases the ability of HR2 to be adsorbed on a model membrane mimicking the viral envelope, such interaction is still insufficient to consider HR2 as a peripheral membrane domain. We also conducted Monte Carlo (MC) simulations with an implicit hydrophobic slab mimicking a membrane, and data derived therefrom corroborate the results we obtained in the course of MD. In the lowest energy states, individual monomeric HR2 did not land onto the membrane interface. Meanwhile, the corresponding MC-states identified for monomeric HR2 linked to TMD corresponded to the protein anchored in the hydrophobic slab via the latter, but HR2 still failed to adsorb on the membrane surface, although it interacted with the hydrophobic medium rather strongly compared to HR2 without the TMD. It is thus likely that other factors than the TMD come into play for proper binding of HR2 to the membrane to take place during the early stages of fusion.

*Acknowledgements:* This work was supported by RSF grant 18-14-00375.

## Hierarchical structure of protein sequences – new insight of protein organization

Anashkina A.A.<sup>1</sup>, Nekrasov A.N.<sup>2</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

**Key words:** protein structure, hierarchy, protein sequences, ANIS method, protein design

*Motivation and Aim:* It should be noted that natural polypeptide chains (NPCs) have a number of properties that are unusual for polymers. In particular, NPCs are able to independently form stable and unique structures of spatial organization. This property allows "nature" to form many classes of "molecular machines" from NPCs, which ensure the flow of all physiological processes in living systems. Classical ideas about protein sequences are based on the idea that amino acid residues are the monomeric unit of NPCs. Such representations are natural and are associated with the chemical structure of polypeptide chains in which amino acid residues of different chemical type are linked by identical peptide groups. Initially, it was assumed that the properties of NPCs are determined by the patterns of distribution of amino acid residues along the polymer chain. However, repeated attempts to analyze the location of residues in the NPCs did not reveal significant patterns in the location. This suggested that there are other elementary structural units of NPCs. The aim of our work was to find the elementary units of the NPCs, allowing us to see the structural organization of the protein sequence.

*Methods and Algorithms:* To achieve the set task, the information entropy of a set of non-homologous natural protein sequences was studied. The study of the entropy characteristics of the NPCs showed that for groups of 5 successively located residues, a reduced level of informational entropy is observed and, therefore, blocks of precisely this size must be considered as elementary structural units of the NPCs. This approximation made it possible to develop a method that reveals the hierarchical structure of the NPCs – the Information Structure Analysis method (ANIS method) [10]. We assumed that the revealed hierarchical structures are associated both with the molecular evolution of NPCs and their functional activity.

*Results:* Further studies showed that the use of the ANIS method made it possible to solve a number of important practical problems in various fields [2] – the design of protein molecules, studies of intermolecular interactions, and the study of the mechanisms of protein functioning. In addition, it was shown on model structural blocks that, with certain combinations of amino acid residues, they can only be located in the predominant conformational given topology [9].

To date, the ANIS method has been successfully applied to solve a number of practical problems in the design of protein molecules. Thus, a truncated form of human 1-CYS peroxiredoxin was obtained, which retained about 95 % of the catalytic activity of the wild form of the protein [6]. In a modified form of human Interleukin-13, the ability to give a signal for the development of an allergic reaction was removed [7]. The specificity of oligopeptidase from *Serratia proteamaculans* was changed and the enzyme acquired the ability to cleave high molecular weight substrates [5]. A highly efficient hydrolase has been made from the bacteriophage  $\phi$ KZ polypeptide [3]. Minimal sites that retain

the functional activity of native proteins have been identified in human tumor necrosis factor [11] and human heat shock protein [12].

In addition, in the study of protein-protein interactions of hydrolase-inhibitor complexes, it was shown that the interface providing their interaction consists of almost 85 % pairs of sites with a low and high density of elements of the information structure of the lower level of the hierarchy [8]. Based on the revealed regularities in the formation of interchain interaction interfaces, a method for identifying immunogenic sites in proteins was proposed and an immune response was obtained for protein epitopes of *Aspergillus fumigatus* Asp f 2 and Asp f 3 [1]. Models for the functioning of a number of hydrolytic enzymes and heme-containing proteins have been proposed [4].

The totality of the presented theoretical and experimental data allows us to speak about the birth of a new paradigm of the structural organization of proteins, which is based on the entropy characteristics of the NPCs. This opens up fundamentally new opportunities not only for studying and modifying known NPCs, but also for "designing" fundamentally new "molecular machines" in terms of organization and function.

*Acknowledgements:* This research has been supported by Russian Foundation for Basic Research Grant No. 20-04-01085.

### References

1. Alekseeva L. et al. Cryptic B-cell epitope identification through informational analysis of protein sequences. *Vaccine*. 2007;25(14):2688-2697.
2. Anashkina A.A. et al. Entropy analysis of protein sequences reveals a hierarchical organization. *Entropy*. 2021;23(12):1647.
3. Briers Y. et al. The structural peptidoglycan hydrolase gp181 of bacteriophage phiKZ. *Biochem Biophys Res Commun*. 2008;374(4):747-751.
4. Chertkova R.V. et al. New insight into the mechanism of mitochondrial cytochrome C function. *PLoS One*. 2017;12(5):e0178280.
5. Mikhailova A.G. et al. Truncated variants of *Serratia proteamaculans* oligopeptidase B having different activities. *Biochemistry (Mosc.)*. 2015;80(10):1331-1343.
6. Nekrasov A.N. et al. The novel approach to the protein design: active truncated forms of human 1-CYS peroxiredoxin. *J Biomol Struct Dyn*. 2007;24(5):455-462.
7. Nekrasov A.N. et al. Design of a novel interleukin-13 antagonist from analysis of informational structure. *Biochemistry (Mosc.)*. 2009;74(4):399-405.
8. Nekrasov A.N. et al. Structural features of the interfaces in enzyme-inhibitor complexes. *J Biomol Struct Dyn*. 2010;28(1):85-96.
9. Nekrasov A.N. et al. A minimum set of stable blocks for rational design of polypeptide chains. *Biochimie*. 2019;160:88-92.
10. Nekrasov A.N. et al. Hierarchical structure of protein sequence. *Int J Mol Sci*. 2021;22(15):8339.
11. Shingarova L.N. et al. Expression and properties of human TNF peptide fragments. *Russ J Bioorganic Chem*. 2010;36(3):301-309.
12. Vlaskin et al. Activating effect of 70 kDa heat shock protein on human NK cells. *Immunologia (Russian)*. 2007;28(2):74-79.

## Comparison of hemoglobin sequences of normoxic species to hypoxia-resistant *Heterocephalus glaber* (naked mole rat)

Anashkina A.A.<sup>1\*</sup>, Melikhova A.V.<sup>2</sup>, Simonenko S.Yu.<sup>2</sup>, Orlov Yu.L.<sup>2</sup>,  
Petrushanko I.Yu.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

**Key words:** oxidative stress, hemoglobin, hypoxia, conservation, glutathione, cysteine

*Motivation and Aim:* Red blood cells containing hemoglobin are carriers of oxygen from the lungs to all tissues of our body. Hemoglobin binds oxygen in the lungs and releases it to peripheral tissues. Hemoglobin is very tightly packed in the erythrocyte and is surrounded by a layer of bound water [1, 2]. The binding and release of oxygen is cooperative in nature (Borh effect) [3]. In different species, hemoglobin may have different affinities for oxygen, may have a feature of oxygen release in tissues, and may also have different sensitivity to oxidative stress developing during hypoxia. Most animals live and breed at a normal level of atmospheric oxygen of about 20 % and experience hypoxia when it decreases. Under conditions of hypoxia, oxidative stress develops in tissues, which leads to oxidative modifications of proteins and cell damage. However, not all animals are equally resistant to lack of oxygen. There are hypoxia resistant animals, such as the naked mole rat, which not only show extreme resistance to hypoxia (5 % O<sub>2</sub>) [4], but are also capable of reproduction. The search for molecular reasons for such stability was the goal of this work. We attempted to find the structural determinants that determine the resistance of the naked digger to oxidative stress by comparing the sequences and structures of hemoglobins from different species.

*Methods and Algorithms:* Each hemoglobin molecule normally consists of four subunits, in humans two of them are formed by alpha chains, and two by beta chains. We searched the Uniprot database and compiled two sets of amino acid sequences separately for hemoglobin alpha and beta chains. The sequences of the corresponding hemoglobins of normoxic animals were assigned to the first group. The second group included hemoglobin sequences of the naked mole rat, which easily tolerate hypoxic conditions and are able to live and reproduce under such conditions. We built multiple alignments in each group by Uniprot (<https://www.uniprot.org/>) and Toffee (<https://tcoffee.crg.eu>) servers to find conserved amino acid residues in both cases and compared them between themselves. In addition, we mapped the variability map to the structure of hemoglobin to search for molecular causes of hypoxic resistance by the ConSurf server (<https://consurf.tau.ac.il/>).

*Results:* The naked mole rat, unlike many other mammals, contains several types of hemoglobin. So, in addition to two variants of the alpha chain, one of which contains two cysteine residues instead of one, it has four variants of the zeta chain close to it and the theta-1 chain similar to them. The beta chain has three different versions, in addition, there are epsilon-1 and gamma chains that are similar to it, and in all the cysteine residue 112 is replaced by valine. If within these groups the sequences of hemoglobin chains are quite similar to each other, then the naked mole rat has another hemoglobin alpha chain,

which differs greatly from both the first and second groups by the presence of a deletion of the first 32 amino acids from the N-terminus, an insertion of five amino acids TSGGA between amino acids 96 and 97 and a 27 amino acid insert QAGPYPCPLCLRPGSPPLPTVHLCIFE at the C-terminus. These substitutions, in addition to structural changes, introduce three additional cysteine residues into the structure of hemoglobin.

*Conclusion:* We believe that the diversity of different types of hemoglobin chains, as well as the replacement of the cysteine 112 residue by valine in all beta chains, contribute to the hypoxic resistance of *Heterocephalus glaber*, since cysteines are actively involved in the regulation of proteins through post-translational modifications of oxidative stress, such as glutathionylation, nitrosylation, or sulfhydrogenation. The biological role of the Cys112Val substitution remains to be elucidated.

*Acknowledgements:* This research has been supported by Russian Science Foundation Grant No. 19-14-00374.

### *References*

1. Latypova L., Barshtein G. et al. Oxygenation state of hemoglobin defines dynamics of water molecules in its vicinity. *J Chem Phys.* 2020;153:135101. doi: 10.1063/5.0023945.
2. Latypova L., Puzenko A. et al. Hydration of methemoglobin studied by *in silico* modeling and dielectric spectroscopy. *J Chem Phys.* 2021;155:015101.
3. Riggs A.F. The Bohr effect. *Annu Rev Physiol.* 1988;50:181-204. doi: 10.1146/annurev.ph.50.030188.001145.
4. Browe B.M., Vice E.N., Park T.J. Naked mole-rats: blind, naked, and feeling no pain. *Anat Rec.* 2020;303(1):77-88. doi: 10.1002/ar.23996.



# Interaction of D-cycloserine with a D-amino acid transaminase from *Haliscomenobacter hydrossis*

Bakunova A. \*, Matyuta I., Nikolaeva A., Boyko K., Popov V., Bezsudnova E.

*Bach Institute of Biochemistry, Research Centre of Biotechnology, RAS, Moscow, Russia*

\* [albakunova@mail.ru](mailto:albakunova@mail.ru)

**Key words:** D-amino acid transaminase, enzyme inhibition, D-cycloserine, substrate recognition

**Motivation and Aim:** Studying the structures of enzyme complexes with inhibitors is a great way to explore the substrate recognition mode of the enzyme. Recently, we characterized pyridoxal-5'-phosphate (PLP)-dependent D-amino acid transaminase (DAAT) from bacterium *Haliscomenobacter hydrossis* (Halhy) with a new type of the recognition site for the substrate  $\alpha$ -carboxyl group [1]. D-cycloserine (D-CS) is a cyclic analog of serine. Various PLP-dependent enzymes, including transaminases (TA), are able to accommodate D-CS in their active sites. D-CS is proposed to react with TA through the canonical sequence of reaction steps and finally forms irreversibly a stable aromatic compound with PLP or dissociates from the active site [2–5]. Determination of the structure of the Halhy complex with D-CS can shed light on both the mechanism of D-CS interaction with Halhy and the mode of substrate binding in the active site of the new type of DAAT.

**Methods and Algorithms:** We explored the interaction of D-CS with Halhy by the combination of UV-Vis spectroscopy, enzyme kinetics and X-ray crystallography.

**Results:** D-CS is a strong inhibitor of Halhy with  $IC_{50}$  value of 3  $\mu$ M. The inhibited enzyme shows an absorbance maximum near 330-340 nm, indicative of a completed catalytic half-reaction. In the crystal structure D-CS is covalently attached to the cofactor and presents in two different states: an intermediate product, external aldimine, and a stable aromatic compound. Contrary to expectations, D-CS is not held by numerous noncovalent interactions in the active site. The only hydrogen bond was found between the oxygen of D-CS carbonyl group and the  $\epsilon$ -amino group of the catalytic lysine.

**Conclusion:** D-CS binds PLP irreversibly; however, the excess of PLP restores the activity of Halhy. In the active site of Halhy D-CS interacts specifically with PLP, however, no specific interaction with the residues of the recognition site is observed. The modes of the natural substrates binding and D-CS coordination are different.

**Acknowledgements:** This work was funded by the Russian Science Foundation, grant No. 19-14-00164.

## References

1. Bakunova A.K., Nikolaeva A.Y. et al. The Uncommon Active Site of D-Amino Acid Transaminase from *Haliscomenobacter hydrossis*: Biochemical and Structural Insights into the New Enzyme. *Molecules*. 2021;26(16):5053. doi: 10.3390/molecules26165053.
2. Peisach D. et al. D-cycloserine inactivation of D-amino acid aminotransferase leads to a stable noncovalent protein complex with an aromatic cycloserine-PLP derivative. *J Am Chem Soc*. 1998;120:2268-2274. doi: 10.1021/ja973353f.
3. Amorim Franco T.M., Favrot L., Vergnolle O., Blanchard J.S. Mechanism-based inhibition of the mycobacterium tuberculosis branched-chain aminotransferase by D- and L-cycloserine. *ACS Chem Biol*. 2017;12:1235-1244. doi: 10.1021/acscmbio.7b00142.
4. de Chiara C., Homšak M. et al. D-Cycloserine destruction by alanine racemase and the limit of irreversible inhibition. *Nat Chem Biol*. 2020;16:686-694. doi: 10.1038/s41589-020-0498-9.
5. Dindo M., Grottelli S. et al. Cycloserine enantiomers are reversible inhibitors of human alanine:glyoxylate aminotransferase: implications for Primary Hyperoxaluria type 1. *Biochem J*. 2019;476:3751-3768. doi: 10.1042/BCJ20190507.

## **Bending compliance of lipid bilayer regulates the curvature threshold and pathway of membrane structural stability loss**

Bashkirov P.<sup>1,2\*</sup>, Kuzmin P.<sup>3</sup>, Arrasate P.<sup>4,5</sup>, Shnyrova A.<sup>4,5</sup>, Frolov V.<sup>4,5,6</sup>

<sup>1</sup> *Scientific Research Institute for Systems Biology and Medicine, Moscow, Russia*

<sup>2</sup> *Federal Research and Clinical Centre of Physical-Chemical Medicine, Moscow, Russia*

<sup>3</sup> *A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, RAS, Moscow, Russia*

<sup>4</sup> *Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country, Leioa, Spain*

<sup>5</sup> *Department of Biochemistry and Molecular Biology, University of the Basque Country, Leioa, Spain*

<sup>6</sup> *Ikerbasque, Basque Foundation for Science, Bilbao, Spain*

\* *bashkirov\_pv@niid.ru*

**Key words:** lipid bilayer structure, membrane elasticity, membrane pores

*Motivation and Aim:* Membrane fission and fusion generally involve highly curved membrane intermediates. Membrane bending not only allows to confine the membrane rearrangement sites but also creates the elastic stress widely regarded as a trigger of local lipid rearrangements. The experimental and theoretical analyses have revealed various lipid rearrangements induced by the critical bending stress pathways. However, it remains unclear which molecular and force factors control the pathways choice and enable leakage-free membrane rearrangements.

*Methods and Algorithms:* Here, we modeled the stress action experimentally and theoretically using a simple 1D model, the membrane nanotube (NT) constricted by proteins and physical forces. We used the electric field to measure the ionic conductance of the NT lumen and, at the same time, assess, in real-time, changes of the material parameters of the NT membrane during the NT constriction and scission.

*Results:* We found that the NT membrane retained linear bending elasticity up to the moment of scission. The scission, seen as an abrupt loss of the luminal connectivity, happened spontaneously upon reaching a certain critical curvature, a function of the lipid composition of the NT membrane. The theoretical analysis demonstrated that the corresponding instability was associated with a threshold value of elastic stress, the bending rigidity function, and the NT monolayers' intrinsic curvature. Analysis of the NT constriction dynamics by proteins revealed that the proteins sense the lipid elastic response and generally create subcritical or transient critical stresses, thus effectively avoiding causing membrane leakage.

*Conclusion:* We conclude that the nanoscale deformations of highly bent intermediates of membrane fusion and fission could be described in terms of linear macroscopic elasticity with the embedded elastic moduli ascribed to bilayer midplane. We further speculate that the material properties of the lipid bilayer are reflected in the functional design of proteins mediating cellular membrane remodeling: the membrane constriction action of the protein machinery is adapted to the lipidome.

*Acknowledgements:* The study is supported by Russian Science Foundation (grant No. 22-15-00265).

## Diversity of structural-dynamic properties of transmembrane domains among the insulin receptor family

Bershatsky Y.<sup>1,2\*</sup>, Kuznetsov S.<sup>1,3</sup>, Lesovoy D.<sup>1</sup>, Plashchinskaia D.<sup>1,2</sup>, Arseniev A.<sup>1</sup>, Efremov R.<sup>1,3</sup>, Bocharov E.<sup>1</sup>

<sup>1</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

<sup>2</sup> *Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow Region, Russia*

<sup>3</sup> *Higher School of Economics, Moscow, Russia*

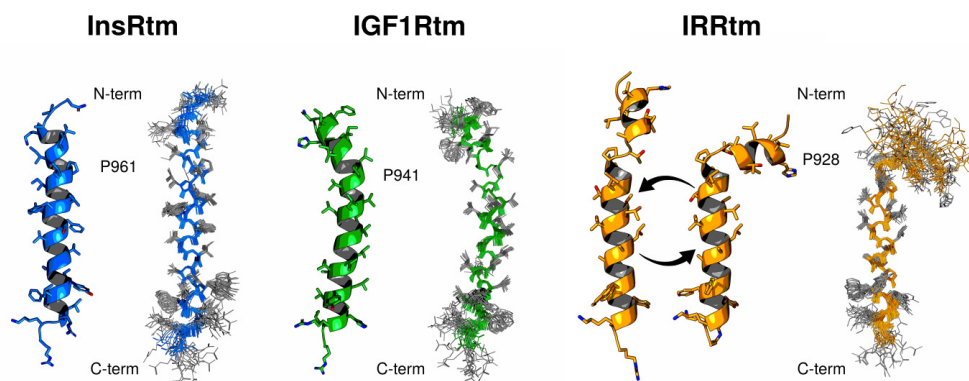
\* *yaroslav.bershatskiy@phystech.edu*

**Key words:** insulin receptors, NMR spectroscopy, transmembrane domain

*Motivation and Aim:* Receptor tyrosine kinases (RTKs) are ubiquitous receptors in the human organism which include the insulin receptor subfamily three highly homologous receptors: the Insulin receptor (IR), the Insulin-like growth factor-1 receptor (IGF1R) and an orphanic the insulin receptor-related receptor (IRR) [1]. The receptors of the insulin receptor subfamily are involved in fundamental physiological processes such as growth, division, differentiation and survival. Disruption of the receptors function has been associated with a wide range of diseases including diabetes, cancer and Alzheimer's disease. Although TMDs of receptors do not interact with ligands directly, they are an important part of the molecular mechanism of receptor activation. Due to the lack of information, the process of signal transduction across the plasmatic membrane via TMD remains poorly understood at the molecular scale.

*Methods and Algorithms:* For characterising the structure and dynamics of all TMDs of the insulin receptor family we assembled genetic constructs, expressed them in CF and purified the recombinant InsR<sub>tm</sub>, IGF1R<sub>tm</sub> and IRR<sub>tm</sub>. The peptides contained a hydrophobic TM segment flanked by short polar N-terminal and C-terminal regions. InsR<sub>tm</sub>, IGF1R<sub>tm</sub> and IRR<sub>tm</sub> were solubilised in the aqueous suspension of DPC micelles. We used a detergent/peptide (D/P) molar ratio of 120 to shift the predominant dynamic equilibrium state of the peptides into monomer form. The prepared samples were explored with a conventional <sup>15</sup>N/<sup>13</sup>C heteronuclear NMR technique.

*Results:* We conducted TMDs structural and dynamycs studies in membranes mimicking micelles using NMR spectroscopy (Fig. 1). Each of three obtained structures of TMDs has some distinguishing features. Both InsR and IGF1R<sub>tm</sub> structures form slightly curved along the entire length  $\alpha$ -helices with a kinks in the proline region. Moreover, IGF1R<sub>tm</sub> has increased flexibility in comparison to InsR<sub>tm</sub>. A distinctive feature of IRR<sub>tm</sub> is to form two short  $\alpha$ -helices connected through a flexible linker.



**Fig. 1.** Spatial structure of the InsRtm (blue; PDB 7PHT), IGF1Rtm (green; PDB 7PH8) and IRRtm (yellow; PDB 7PL4) monomers obtained in micellar membrane mimicking environments. (Left) The ribbon diagrams and (right) the ensemble of 20 NMR-derived structures after alignment of the backbone atoms monomers of the TMDs. Heavy atom bonds are depicted. Hydrogen atom bonds are shown only for polar groups

*Conclusion:* The structures of InsRtm and IGF1Rtm are similar. They have approximately same length of the TM helices. The high similarity of the structures of InsRtm and IGF1Rtm is consistent with the previously obtained MD, structural and biochemical data. IRRtm is significantly different from the other two family members. The IRRtm structure is represented by two helices. The proline sequence, as well as the TATP motif in front of it, locally increase the mobility to such an extent that one turn of the helix disintegrates. This difference of structures probably indicates a different mechanism of signal transduction across the membrane.

*Acknowledgements:* The study is supported by Russian Science Foundation (project No. 18-14-00375).

### References

1. Lemmon M.A., Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. 2010;141(7):1117-1134.

# Application of X-Ray, SAXS and essential dynamics simulations to study conformational transitions of oligopeptidase B

Bershatsky Y.<sup>1\*</sup>, Britikov V.<sup>2</sup>, Timofeev V.<sup>1,3</sup>, Petrenko D.<sup>4</sup>, Britikova E.<sup>2</sup>, Bocharov E.<sup>1</sup>, Mikhailova A.G.<sup>1</sup>, Rakitina T.<sup>1</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Institute of Bioorganic Chemistry, National Academy of Sciences of Belarus, Minsk, Belarus

<sup>3</sup> Federal Scientific Research Center “Crystallography and Photonics”, RAS, Moscow, Russia

<sup>4</sup> National Research Center “Kurchatov Institute”, Moscow, Russia

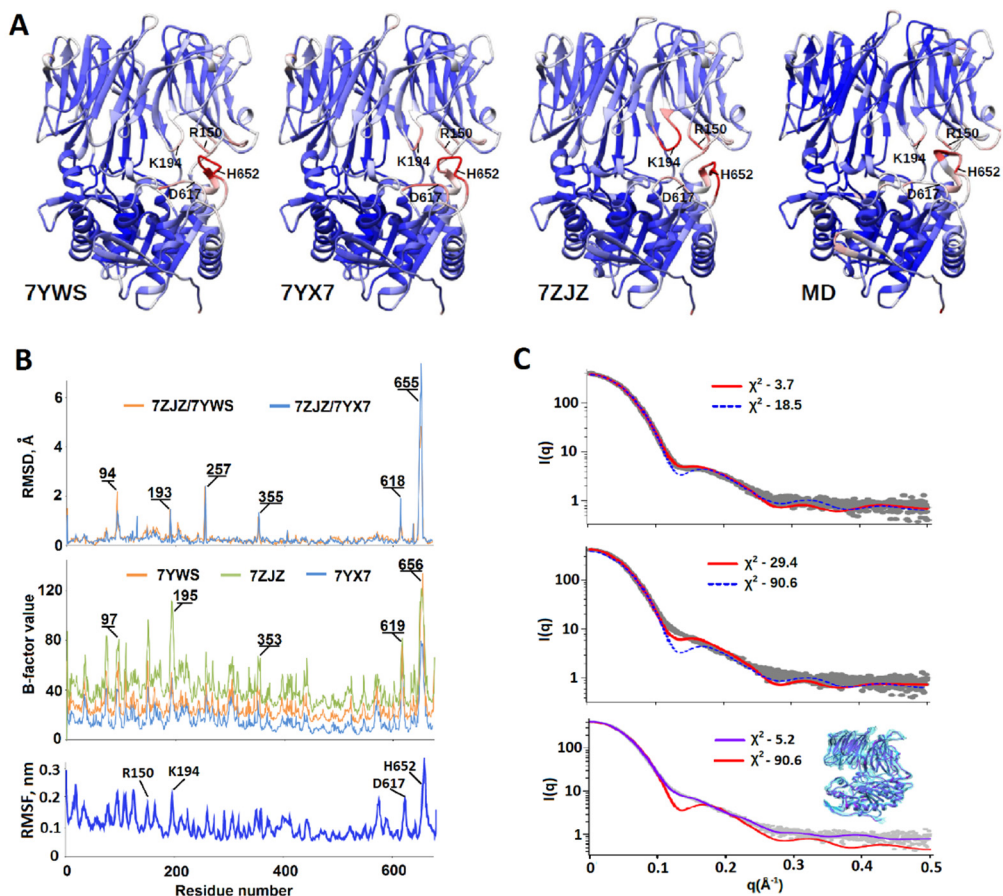
\* bershackyjaroslav@gmail.com

**Key words:** prolyl oligopeptidase, oligopeptidase B, intermediate state, Essential Dynamics Sampling

*Motivation and Aim:* Oligopeptidases B (OPBs) are trypsin-like serine peptidases from the family of prolyl oligopeptidases (POPs). OPBs are found only in bacteria and parasitic protozoa and represent pathogenesis factors of the corresponding infections and putative pharmacological targets. According to the X-Ray study, these two-domain enzymes exist in 3 conformations: closed, open, and intermediate. In first case, the domains and residues of the catalytic triad (S, D, and H) are brought together; in second, they are separated; while in third case, the domains closure preceded the assembly of the catalytic triad. Transitions between these states regulates catalytic activation and consequently provide an avenue for inhibiting the enzymatic activity. Previously, X-Ray diffraction analysis (XRD) showed that OPB from bacteria *Serratia proteamaculans* (SpOPB) crystallized in the presence of polyamine spermine in the intermediate conformations; according to small-angle X-ray scattering (SAXS), in solution SpOPB adopted the open state, while spermine induced its transition to the intermediate state [1].

*Methods and Algorithms:* To elucidate the conformational dynamics of SpOPB, we applied Essential Dynamics/Molecular Dynamics (ED/MD), which combines classical MD simulation with Essential Dynamics Sampling (EDS), which used to guide MD simulations. In brief, when a definition of the collective fluctuations with largest amplitude is obtained from an initial MD simulation, EDS is used to manipulate the position of a protein along collective coordinates (eigenvectors) stimulating the system to explore new regions along these collective coordinates. The results of this computational analysis were compared with those obtained by experimental X-Ray-based methods: XRD analysis and SAXS. Both previously published in [1] and new structures were included in the comparison.

*Results:* The first crystal structure of catalytically deficient SpOPB (SpOPBS532A) with the intact hinge sequence was obtained as well as new structures of the modified enzyme. Similarly to SpOPBs with modified hinges, SpOPBS532A was crystallized in the intermediate conformation. Despite the overall similarity of the crystal structures, an important difference was found in the arrangement of the catalytic D532, which could be the reason for the activity loss of the modified enzyme. The relationship between local fluctuations and conformational transitions of SpOPB was studied by ED/MD using the SpOPBS532A crystal structure (PDB ID 7ZJZ) as a starting point (Fig. 1). As a result of the EDS simulation, the enzyme with the intact hinge was transferred from the intermediate to open state.



**Fig. 1.** Comparison of the crystal SpOPB structures with the averaged structure obtained from classical MD (A, B). Superposition of the experimental and theoretical SAXS curves (C). In the top and medium graphs, the curves for averaged MD structures (red) and the crystal structure 7ZJZ (blue) were fitted to the experimental SAXS curve (gray dots) obtained in presence (top) and absence (medium) of spermine. At the bottom, experimental SAXS profile for spermine-free SpOPB was aligned with the theoretical curves for the best EDS structure (violet) and the crystal structure (red)

*Conclusion:* Summarizing, we can state that during the EDS simulation, which started from the intermediate conformation, the open conformation of the protein was reached and this open conformation correspond to the conformation observed in solution with a high degree of probability. Thus, the approach, which allowed simulation of the conformational transition, can be applicable to different systems for both study of the fundamental mechanisms of protein activity and structure-based drug design.

*Acknowledgements:* The Russian Science Foundation, grant No. 21-74-20154, funded the research. The Center for collective use "Bioorganika" of Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry RAS provided technical support.

## References

1. Petrenko D.E. et al. First Crystal Structure of Bacterial Oligopeptidase B in an Intermediate State: The Roles of the Hinge Region Modification and Spermine. *Biology*. 2021;10(10):1021.

## “Hot spots” for pathogenic gain-of-function mutations in transmembrane domains of receptor tyrosine kinases

Bocharov E.

*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

*Moscow Institute of Physics and Technology, MIPT, Dolgoprudny, Russia*

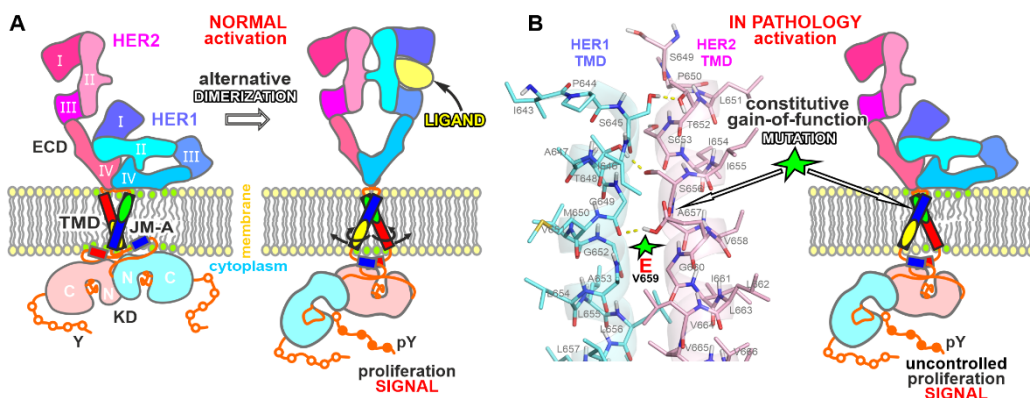
*edvbon@mail.ru*

**Key words:** receptor tyrosine kinases, transmembrane domain, protein-lipid and protein-protein interactions, signal transduction; pathogenic mutations

*Motivation and Aim:* Signal transduction by receptor tyrosine kinases (RTKs) and related receptors has been in the spotlight of scientific interest owing to the central role of these single-spanning membrane proteins in the regulation of development, cell motility, proliferation, differentiation, and apoptosis. Nowadays, the elucidation of high-resolution structure of full-size RTK having flexible multiple-domain composition is still a challenge. During signal transduction across plasma membrane, RTKs are activated by proper ligand-induced homo- and hetero-dimerization or by reorientation of monomers in preformed receptor dimers upon ligand binding. Specific helix-helix interactions of transmembrane domains (TMD) are believed to be important for RTK lateral dimerization and signal transduction. Either destroying or enhancing such helix-helix interactions, e.g. by mutagenesis, can result in many human diseases: developmental, oncogenic, neurodegenerative, immune, cardiovascular etc.

*Methods and Algorithms:* We experimentally determined the alternative dimeric conformations of the TMDs of several representatives of RTK and related receptors in different membrane-mimicking environments using integrative structural biology methods, including high-resolution NMR spectroscopy combined with MD-relaxation in explicit lipid bilayer. Based on the available mutagenesis and clinical data about localization of gain-of-function mutation, observed conformations correspond to the dormant and active states of these receptors.

*Results:* Fine adaptation of intermolecular polar and hydrophobic contacts that we found to accompany the different dimerizations of human epidermal growth factor receptor (HER) TMDs (observed in detergent micelles or in lipid bicelles) suggests that certain membrane properties can govern the TMD helix-helix packing and, thus, their alteration can trigger the receptor state. Whereas two distinct dimeric modes of growth hormone receptor (GHR) TMD (coexisting in micellar environment) revealed the functional role of juxtamembrane region rearrangements in alternation between protein-protein and protein-lipid interactions that can be initiated by ligand binding. Pathogenic transmembrane mutations found for the HER relatives are located as a rule in narrow regions within the specific TMD helix-helix interfaces assuming that the intermolecular interactions inside membrane are important for the RTK cell signaling dysfunction in human organism [1, 3]. Such regions can be characterized as a “hot spot” for gain-of-function mutations associated with different human pathologies.



**Fig. 1.** Normal and pathological activation of HER1/HER2 hetero-dimer as a model RTK. (A) Presumable structural rearrangements of HER1/HER2 hetero-dimer in the course of ligand-induced receptor activation according to the lipid-mediated rotation-coupled mechanism proposed in [1]. (B) Additional inter-monomeric hydrogen bonding caused by HER transmembrane polar mutations (e.g. V659E-HER2) can stabilize the N-terminal dimerization mode of TMD allowing release of cytoplasmic juxtamembrane regions (JM-A) from the membrane leaflet and activation of kinase domains (KD) independently of ligand binding and regardless of the states of extracellular domains (ECD) and surrounding lipids [3]

**Conclusion:** Observed the RTK TMD helix-helix packing diversity appears in favor of the recently proposed the lipid-mediated rotation-coupled activation mechanism [1], which implies that the sequence of structural rearrangements of RTK domains is associated with perturbations of the lipid bilayer in the course of ligand-induced receptor activation, considering the receptor together with its lipid environment as a self-consistent signal transduction system (Fig. 1A). The finding of the pathogenic mutation “hot spots” justifies a prediction that similar gain-of-function mutations, e.g. enhancing the TMD dimerization in certain conformation and thus activating the receptor independently of ligand binding (Fig. 1B), can be found for other RTK representatives and suggests searching for them is a promising idea for future clinical studies. It can also have potential therapeutic implications, broadening the spectrum of targets for pharmaceuticals by inclusion of plasma membranes and their constituents.

**Acknowledgements:** The study is supported by Russian Science Foundation (project No. 18-14-00375).

### References

1. Bocharov E.V. et al. Helix-helix interactions in membrane domains of bitopic proteins: Specificity and role of lipid environment. *Biochim Biophys Acta Biomembr.* 2017;1859(4):561-576.
2. Ou S.I., Schrock A.B., Bocharov E.V. et al. HER2 transmembrane domain (TMD) mutations (V659/G660) that stabilize homo- and heterodimerization are rare oncogenic drivers in lung adenocarcinoma that respond to afatinib. *J Thorac Oncol.* 2017;12(3):446-457.
3. Bocharov E.V. et al. Structural basis of the signal transduction via transmembrane domain of the human growth hormone receptor. *Biochim Biophys Acta Gen Sub.* 2018;1862(6):1410-1420.



## Structural features of bacterial luciferase related to temperature adaptation

Deeva A.A.<sup>1\*</sup>, Sukovatyi L.A.<sup>1</sup>, Lisitsa A.E.<sup>1</sup>, Nemtseva E.V.<sup>1,2</sup>

<sup>1</sup> Siberian Federal University, Krasnoyarsk, Russia

<sup>2</sup> Institute of Biophysics, SB RAS, Krasnoyarsk, Russia

\* adeeva@sfu-kras.ru

**Key words:** bacterial luciferase, mesophile, psychrophile, enzyme kinetics, structural features

**Motivation and Aim:** Luminous bacteria are widely distributed in nature, with the majority of species living in the ocean and only a few representatives described as freshwater and terrestrial [1]. Thus, they have adapted to live in various environmental conditions and at different temperatures as well. Based on optimal growth temperature, psychrophilic and mesophilic species of luminous bacteria can be distinguished. The features of bacteria adaptation to specific temperature range can be seen across all levels of organization, including enzymatic. One of the most studied enzyme of luminous bacteria is luciferase, which is responsible for bioluminescent reaction and widely applied as an element of biosensors in fundamental research, biotechnological applications, ecotoxicology, etc. [2, 3]. The aim of this study is to reveal the structural features of two luciferases from different subfamilies (*Photobacterium leiognathi* and *Vibrio harveyi*) related to temperature dependence of their function.

**Methods and Algorithms:** The set of 47 amino acid sequences of bacterial luciferase associated with fully or partially sequenced *lux*-operons from NCBI database was obtained manually. We used the set of sequences to analyze the amino acid substitutions specific to temperature adaptation with GroupSim and multi-Harmony program packages. The temperature dependence of reaction kinetics for *V. harveyi* and *P. leiognathi* luciferases was measured using stopped-flow technique with SX-20 analyzer (Applied Photophysics) in the range 5–45 °C. Additionally, we have run 100-ns molecular dynamics simulations for both enzymes at 5, 10, 27, 45, 60 °C using the GROMACS 2020.4 software package.

**Results:** Sequence analysis revealed that *P. leiognathi* luciferase is highly homologous to luciferases from psychrophilic strains, while *V. harveyi* luciferase is very similar to the thermostable luciferase from *V. campbellii*. According to experimental results, *V. harveyi* luciferase does have a wider temperature optimum shifted to higher temperatures compared to *P. leiognathi* luciferase. However, the total quantum yield is lower in the reaction catalyzed by *V. harveyi* luciferase compared to the *P. leiognathi* one. Kinetics analysis indicated that *P. leiognathi* luciferase better stabilizes 4a-hydroperoxyflavin intermediate in the temperature range of 20–35 °C, and is less effective at temperatures of 40–45 °C.

Computational simulation showed that *P. leiognathi* structure is more dynamic, but the shape of the protein does not remarkably change with increasing temperature, which is typical for psychrophilic enzymes. *V. harveyi* luciferase undergoes minor overall fluctuation at studied temperatures indicating its higher rigidity, which is characteristic for thermostable enzymes. We also analyzed the flexibility of separate structural elements of the proteins at different temperatures. The fluctuations of a mobile loop

located near the active site becomes higher for both enzymes with increasing temperature. Additionally, a turn in  $\beta$ -sheet of *V. harveyi* luciferase (149–151 a.a.) is more mobile at 45 and 60 °C indicating that stabilizing interactions of this fragment is not the same as those in *P. leiognathi* luciferase. Moreover we have found nine amino acid substitutions which could be specific for observed temperature effects. All of them are located in highly rigid regions of both enzymes except for one found within functional mobile loop mentioned above. These residues could be responsible for compensatory mechanism required to maintain a balance between flexibility and stability of the enzyme.

*Conclusion:* *V. harveyi* and *P. leiognathi* luciferases are characterized not only by different kinetics of light emission (“fast” and “slow”), but also by structural characteristics which could be responsible for different dependence of the enzymes function on temperature. The obtained results are important for development of enzyme based bioassays that are used for ecological monitoring.

*Acknowledgements:* The research was partially funded by the Ministry of Science and Higher Education of the Russian Federation (project No. FSRZ-2020-0006) and by RFBR, Krasnoyarsk Territory and Krasnoyarsk Regional Fund of Science (project No. 20-44-243002).

### *References*

1. Dunlap P. Biochemistry and genetics of bacterial bioluminescence. *Adv Biochem Eng Biotechnol.* 2014;144:37-64.
2. Esimbekova E.N. et al. Enzymatic biotesting: scientific basis and application. *Contemp Probl Ecol.* 2021;14(3):290-304.
3. Syed A.J., Anderson J.C. Applications of bioluminescence in biotechnology and beyond. *Chem Soc Rev.* 2021;50(9):5668-5705.

## Protein design strategy for structure determination of a peptide GPCR

Dmitrieva D.\*, Smirnova E., Sadova A., Mishin A.

*Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

\* [dmitrieva.da@phystech.edu](mailto:dmitrieva.da@phystech.edu)

**Key words:** gene modifications, stabilization, GPCR

*Motivation and Aim:* G protein-coupled receptors (GPCRs) are the largest and most diverse group of membrane receptors in eukaryotic organisms. These receptors are involved in a wide range of physiological processes and are targeted by roughly a third of all existing drugs, and studying them is without a doubt a real necessity. In this work we study a peptide GPCR associated with inflammation in autoimmune diseases. Despite the rapid advances in structural biology in recent years, obtaining structural data for GPCRs is still very difficult, mainly because they tend to be thermally unstable and have poor crystal quality and low yields. Success in obtaining the structure of membrane proteins requires the achievement of several objectives: produce monodisperse samples with improved thermo- and conformational stability, increased yield of pure protein (over 0.2–0.5 mg of purified protein per 1 L of cell culture), crystal size and diffraction quality [1].

*Methods and Algorithms:* For expression target protein we used baculovirus expression system in Sf9 cell line. This system is one of the most successful expression systems for obtaining properly folded and functional integral membrane proteins [2]. To obtain target protein suitable for future crystallization trials we developed different genetically engineered constructs with small fusion partner proteins. Since these fusion partners increase protein quality and improve possible crystal contacts [3]. To increase the protein expression level, yield and thermal stability, we introduced several point mutations taken from accumulated structural and biophysical data (mutations predicted by machine-learning-approach based on sequence and structure GPCRs data) [4]. Additionally, we introduced a hemagglutinin signal sequence to facilitate N-tail translocation at the endoplasmic reticulum membrane, a Flag tag for detection of the receptor by antibodies, and 10 × His tag for purification on affinity resin.

*Results:* The best constructs showed significantly increase in the receptor's surface expression levels, as well as improve its yield in comparison with the wild type, good size exclusion chromatography, increased thermal stability with antagonist molecules.

*Conclusion:* The development of genetic constructs is one of the most important parts of structural biology of GPCRs and requires a lot of search effort and characterization of a large number of constructs.

*Acknowledgements:* The work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-00337-20-03, project FSMG-2020-0003).

## References

1. Ishchenko A. et al. Crystallization of Membrane Proteins: An Overview. In: Wlodawer A., Dauter Z., Jaskolski M. (Eds.). *Protein Crystallography. Methods in Molecular Biology*. New York: Springer Science, 2017;1607:117-141.
2. Contreras-Gómez A. et al. Protein production using the baculovirus-insect cell expression system. *Biotechnol Prog*. 2014;30(1):1-18.
3. Chun E. et al. Fusion partner toolchest for the stabilization and crystallization of G proteincoupled receptors. *Structure*. 2012;20(6):967-976.
4. Popov P. et al. Computational design of thermostabilizing point mutations for G protein-coupled receptors. *eLife*. 2018;7:e34729.

## “Dynamic molecular portraits” of biomembranes: a computational insight

Efremov R.

*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

*Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

*National Research University Higher School of Economics, Moscow, Russia*

*r-efremov@yandex.ru*

**Key words:** Computer simulations, dynamics of lipid membranes, lateral heterogeneity of membrane, membrane protein structure, mosaicity of membrane surface, physico-chemical properties of lipid bilayers, spontaneously formed nanodomains

*Motivation and Aim:* To date, it has been reliably shown that the lipid bilayer/water interface can be thoroughly characterized by a sophisticated so-called “dynamic molecular portrait”. The latter reflects a combination of time-dependent surface distributions of various physicochemical properties, inherent in both model lipid bilayers and natural multi-component cell membranes. One of the most important features of biomembranes is their mosaicity, which is expressed in the constant presence of lateral inhomogeneities, the sizes and lifetimes of which vary in a wide range—from 1 to 10<sup>3</sup> nm and from 0.1 ns to milliseconds. In addition to the relatively well-studied macroscopic domains (so-called “rafts”), the analysis of micro- and nanoclusters (or domains) that form an instantaneous picture of the distribution of structural, dynamic, hydrophobic, electrical, etc. properties at the membrane-water interface is attracting increasing interest. This is because such nanodomains (NDs) have been proven to be crucial for the proper membrane functioning in cells. Therefore, an understanding with atomistic details the phenomena associated with NDs is required.

*Results:* The present work describes the recent results of *in silico* studies of spontaneously formed NDs in lipid membranes. The main attention is paid to the methods of ND detection, characterization of their spatiotemporal parameters, the elucidation of the molecular mechanisms of their formation. Biological role of NDs in cell membranes is discussed. In particular, self-adaptation of “dynamic molecular portraits” of membrane peptides/proteins and lipids/water is considered in atomistic simulations.

*Conclusion:* Understanding such effects creates the basis for rational design of new prospective drugs, therapeutic approaches, and artificial membrane materials with specified properties.

*Acknowledgements:* This work was supported by the Russian Science Foundation (grant No. 18-14-00375).

## Sampling the parameter space for heterologous bacterial production of intracellular domain of TLR1 reveals multiple optima

Goncharuk M.V.<sup>1\*</sup>, Lushpa V.A.<sup>1,2</sup>, Goncharuk S.A.<sup>1</sup>, Arseniev A.S.<sup>1</sup>, Mineev K.S.<sup>1</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Russia

\* [m.s.goncharuk@gmail.com](mailto:m.s.goncharuk@gmail.com)

**Key words:** bacterial protein expression, optimization, protein solubility, TLR1, TIR, Triton X-100

*Motivation and Aim:* The T7 expression system is widely used for the bacterial production of recombinant proteins for structural and functional analysis and therapeutic applications. Despite that many successful approaches are known, high level expression of soluble target proteins in *Escherichia coli* cells often requires a systematic optimization of wide range of cultivation parameters, particularly if heterologous protein expression takes place. Here we investigate the effect of three key cultivation parameters (concentration of the chemical inductor, temperature and time of postinductional culturing) on the expression level of TIR domain of TLR1 (TIR<sub>TLR1</sub>) in a soluble form. We also study the role of mild detergent Triton X-100 and the ionic strength of the cell lysis solution in the target protein solubility and the final yield.

*Methods and Algorithms:* Basic protocols of genetic engineering, protein chemistry and *Escherichia coli* cells cultivation using M9 minimal salts and LB reach growth media were implemented. A small-scale expression in BL21(DE3) / BL21(DE3)pLysS bacterial strain cells followed by an SDS-PAGE was chosen as the main strategy to analyze the expression level of the TIR<sub>TLR1</sub> under various cultivation conditions. The correct folding of the heterologously expressed target protein was biophysically characterized by CD and NMR spectroscopies.

*Results:* The results of the extensive screening of bacterial cell cultivation and lysis parameters used for the production of TIR<sub>TLR1</sub> domain are presented. Fine-tuning of all the parameters yielded about 30 mg/l M9 of correctly folded soluble protein. We showed that similar maximal yields may be obtained through several various combinations of the parameters, allowing to select the appropriate time consumption and reproducibility of the protocol.

*Acknowledgements:* This study was supported by the Russian Science Foundation (grant No. 22-14-00020). <https://rscf.ru/project/22-14-00020/>

## Solving the structural puzzle of neurotrophin signaling

Goncharuk S.A.<sup>1\*</sup>, Artemieva L.E.<sup>1,2</sup>, Kot E.F.<sup>1,2</sup>, Shabalkina A.V.<sup>1,2</sup>,  
Vasilieva E.V.<sup>1,3</sup>, Nadezhdin K.D.<sup>1</sup>, Arseniev A.S.<sup>1</sup>, Mineev K.S.<sup>1</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology,  
Dolgoprudny, Russia

<sup>3</sup> Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

\* [ms.goncharuk@gmail.com](mailto:ms.goncharuk@gmail.com)

**Key words:** Trk, P75NTR, NMR, receptor, neurotrophins, membrane proteins

*Motivation and Aim:* Neurotrophins and their receptors are involved in many important processes in neurons like proliferation, differentiation, neuroplasticity, apoptosis, etc. Incorrect functioning of any element in these cascades leads to the development of various neurodegenerative and mental diseases, including Alzheimer's and Huntington's diseases, depression and schizophrenia. A detailed understanding of the mechanisms of neurotrophin receptor activation, their interaction with ligands and adapter proteins can explain the diversity of biological responses and give an approach for the creation of novel drugs based on the rational design. Here we present our structural findings in the big puzzle of neurotrophin signaling.

*Methods:* We used bacterial and cell-free expression systems to produce the transmembrane domains of P75NTR and TrkA neurotrophin receptors, full intracellular domain and death domain (DD) of P75NTR, and receptor P75NTR without extracellular region. To investigate the transmembrane-containing fragments we used different membrane-mimetic media (micelles, bicelles, nanodiscs and liposomes). To characterize the behavior, dimerization propensity and structure of the proteins we used site-directed mutagenesis, light scattering, and solution NMR spectroscopy.

*Results:* First, we showed that there is no interaction between DDs of P75NTR in any context. We proposed a novel model of P75NTR functioning, suggesting that P75NTR dimerization occurs in the complex of two P75NTR molecules and “helper” protein, and ligand binding leads to the “helper” release, which restores the ability of P75NTR dimer to interact with adapter proteins. Second, we identified and characterized the direct interaction between p75NTR and TrkA through their transmembrane domains. Third, we investigated the dimerization of transmembrane domain of the TrkA and found two dimer interfaces that we believe correspond to active and inactive states of the receptor. Finally, we demonstrate that eJTM region of TrkA are intrinsically disordered and most likely that signal transduction occurs after direct interaction between ligand and this domain.

*Conclusion:* Membrane protein investigation is a big challenge for structural biology. To date, there are no experimental methods to investigate the structure of full-length receptor, like Trk or P75, therefore the researchers are trying to understand the activation mechanisms based on a lot of different indirect data. All our findings presented here open an additional part of the puzzle, but the other studies are required to see the full picture.

*Acknowledgements:* This study was supported by the Russian Science Foundation (grant No. 22-14-00130), <https://rscf.ru/project/22-14-00130>.

## Statistical approach to regulation of membrane-permeabilizing activity of antimicrobial peptides

Kondrashov O.V., Akimov S.A.\*

*A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, RAS, Moscow, Russia*

\* *akimov\_sergey@mail.ru*

**Key words:** lipid membrane, theory of elasticity, antimicrobial peptide, partition function, Mayer cluster expansion

*Motivation and Aim:* Antimicrobial peptides (AMP) are considered promising antibiotics with an extremely low predicted probability of drug resistance appearance. The development of new antimicrobial drugs is currently especially urgent: the problem of antibiotic resistance, according to the World Health Organization, is one of the most serious threats to human health, food security and development of the entire world. Purposeful rational development of antimicrobial drugs is impossible without a detailed understanding of the fundamental molecular mechanisms of their action. The action of some types of AMP is based on an increase in the permeability of the membranes of pathogenic cells, leading to their death. The formation of a through pore in the membrane is a cooperative process that requires the participation of several peptides. Permeabilization of membranes by AMPs at a certain surface concentration is a stochastic process, for understanding and adequate description of the mechanisms of which it is necessary to use statistical approaches.

*Methods and Algorithms:* We developed the statistical model that allows describing the effect of membrane-deforming inclusions on the equilibrium between AMP monomers and their cooperative membrane-permeabilizing structures. Permeabilization of the membrane via formation of pores or ion channels is accompanied by elastic deformations of the membrane, and these deformations contribute to characteristics of the membrane-permeabilizing structures. Thus, the average conductivity of the membrane should depend on the spatial structure of its deformations. Rigid membrane inclusions that purposefully modify the deformation fields of the membrane can potentially regulate the activity of AMPs. The pairwise interaction potentials of peptide molecules, membrane-permeabilizing structures, peptides in transient configurations, and rigid membrane inclusions regulating the deformation fields were calculated in the framework of theory of elasticity of lipid membranes. The obtained energies were then used to obtain the partition function of the system in the framework of the Mayer cluster expansion.

*Results:* Short transmembrane peptides were shown to strongly favor the membrane-permeabilizing states of channel-forming antimicrobial peptides (CFAP). Small amphiphathic peptides were also shown to increase the average conductivity of the membrane due to CFAP, although their working concentrations should be about two orders of magnitude higher as compared to short transmembrane peptides. Helical amphiphathic peptides favored formation of hydrophobic defects in the membrane, but they hindered formation of small hydrophilic pores.

*Conclusion:* The results obtained may be utilized for analysis of amplifying or weakening the effect of various membrane-deforming inclusions with respect to different types of AMPs. The only requirement is that at the sequential stages of the



action, the AMP should induce different deformations of the membrane. In multiple systems, where membrane-deforming particles (monomers) can form higher-order aggregates (dimers, trimers, etc.) an addition of deformation-inducing inclusions can change the equilibrium distribution of the aggregates. In such reactions, the inclusions can strongly favor or stabilize a particular state or aggregate thus shifting the equilibrium towards that state. The model predicts that amphipathic peptides and short transmembrane peptides playing the role of the membrane-deforming inclusions, even in low concentration can substantially increase the average conductivity of membranes. *Acknowledgements:* The study was supported by the Russian Science Foundation (grant No. 22-24-00834).

# Transfer learning for protein stability prediction

Konstantinova A.V.<sup>1,2,3\*</sup>, Kardymon O.L.<sup>2</sup>, Ivanisenko N.V.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Artificial Intelligence Research Institute, Moscow, Russia*

<sup>3</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *konstantaalina@gmail.com*

**Key words:** protein stability, structure prediction

*Motivation and Aim:* Protein stability is an important biologically significant characteristic. It is, for example, crucial in protein design or has to be considered when studying protein variants and effects of mutations. Deep learning approaches, including AlphaFold2 [2], currently demonstrate highest accuracy in protein structure prediction. In this work we aimed to investigate the accuracy of protein stability changes prediction using amino acid residues representations extracted from pretrained deep learning models including ESM-1b [1] and AlphaFold2 [2].

*Methods and Algorithms:* Benchmark included publicly available datasets of experimentally studied proteins with known stability characteristics including melting temperature or free energy changes upon amino acid substitutions. The amino acid representations were extracted from the hidden layers of pretrained models, including ESM-1b and AlphaFold2. Extracted representation were used to train the model to predict stability changes using boosting algorithm.

*Results:* The model based on AlphaFold2 and protein language models was developed, trained and demonstrated comparable accuracy to peer-reviewed methods. The obtained model was validated to estimate the effects of known SARS-CoV-2 variants on S-protein structure and stability.

*Conclusion:* Transfer learning approach applied in the current work allows to efficiently predict protein stability changes and can be applied for a wide range of protein design tasks.

*Acknowledgements:* Work 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.

## References

1. Rives A., Meier J., Sercu T., Goyal S., Lin Z., Liu J., ..., Fergus R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*. 2021;118(15):e2016239118.
2. Jumper J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.

## Activation of a Trk receptor is mediated by a disordered extracellular juxtamembrane region

Kot E.F.<sup>1,2\*</sup>, Franco M.L.<sup>3</sup>, Vasilieva E.V.<sup>1</sup>, Shabalkina A.V.<sup>1,2</sup>, Arseniev A.S.<sup>1,2</sup>, Vilar M.<sup>3</sup>, Mineev K.S.<sup>1,2\*\*</sup>, Goncharuk S.A.<sup>1,2</sup>

<sup>1</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

<sup>2</sup> *Moscow Institute of Physics and Technology, Dolgoprudnyi, Russia*

<sup>3</sup> *Molecular Basis of Neurodegeneration Unit. Institute of Biomedicine of València (IBV-CSIC), València, Spain*

\* [kot@phystech.edu](mailto:kot@phystech.edu)

\*\* [mineev@nmr.ru](mailto:mineev@nmr.ru)

**Key words:** Trk, receptor, neurotrophin, NMR, transmembrane domain, membrane protein, bicelle, micelle

*Motivation and Aim:* Tropomyosin receptor kinases (Trks) include TrkA, TrkB and TrkC receptors. They specifically bind neurotrophins to regulate neuronal differentiation, survival and growth. Trks are a prospective target to treat various neurodegenerative and neurological diseases. Recently, allosteric regulation of the TrkB receptor by antidepressant drugs was observed [1]. Further search of potential agonists and antagonists of Trks requires the knowledge of the spatial structure of these receptors and mechanism of activation.

To date, multiple high-resolution NMR and X-ray structures of separate extracellular and intracellular domains of Trk receptors have been obtained, including complexes with ligands. Structure of the dimeric transmembrane domain of human TrkA was resolved. It is commonly considered that Trk receptors dimerize and transduce signal upon ligand binding, but recently it was reported that TrkA additionally requires the mutual rotation of transmembrane domains in the dimer to be activated [2]. The ligand-binding domain is connected to the transmembrane helix through an extracellular juxtamembrane region (eJTM) with a low propensity to form a stable secondary structure, which has not yet been examined. Thus, ligand binding should be somehow conformationally coupled to the transmembrane domain by the eJTM, but the mechanics of this coupling remain obscure. In the present work we aimed to make a structural investigation of the eJTM region and to test the effect of its mutations on the behavior of the TrkA receptor.

*Methods and Algorithms:* For cellular assays HeLa, PC12 and PC12nr5 cell lines were used. For NMR studies we produced the construct containing the juxtamembrane region and transmembrane domain of human TrkA (TrkA-eJTM-TMD) using a cell-free expression system based on the E.coli lysate. To assign the protein backbone signals in NMR spectra we employed the standard set of triple resonance experiments along with relaxation measurements to assess the mobility and ordering and diffusion spectroscopy to evaluate the particle radii.

*Results:* We demonstrated that eJTM is intrinsically disordered and interacts with neither lipid bilayer nor divalent cations. According to cellular assays, replacement of 5 conservative prolines to glycines strongly reduced the affinity of TrkA to its ligand. NMR spectroscopy showed that this replacement made the eJTM on average more compact and helical, but left it disordered. The activating mutation, K410C [2], led to

the formation of a transmembrane dimer different from the previously observed, which should correspond to the active state of TrkA.

*Conclusion:* Neurotrophin binding to Trk receptors is mechanically coupled to the transmembrane domain through an intrinsically disordered region. Together, our findings strongly support the hypothesis that in addition to the ligand-binding domain, neurotrophins directly bind the disordered eJTM region of Trk receptors.

*Acknowledgements:* This study was supported by the Russian Science Foundation (grant No. 22-14-00130), <https://rscf.ru/project/22-14-00130/>.

### *References*

1. Casarotto P.C. et al. Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. *Cell*. 2021;184(5):1299-1313.e19. doi: 10.1016/j.cell.2021.01.034.
2. Franco M.L., Nadezhdin K.D., Goncharuk S.A., Mineev K.S., Arseniev A.S., Vilar M. Structural basis of the transmembrane domain dimerization and rotation in the activation mechanism of the TRKA receptor by nerve growth factor. *J Biol Chem*. 2020;295(1):275-286. doi: 10.1074/jbc.RA119.011312.

## NMR study of lipid/protein interactions in trifluoroethanol-water mixtures proposes the strategy for the refolding of helical membrane proteins

Kot E.F.<sup>1,2\*</sup>, Motov V.V.<sup>1,2</sup>, Shabalkina A.V.<sup>1,2</sup>, Goncharuk S.A.<sup>1</sup>, Arseniev A.S.<sup>1,2</sup>, Mineev K.S.<sup>1,2\*\*</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Dolgoprudnyi, Russia

\* kot@phystech.edu, \*\* mineev@nmr.ru

**Key words:** refolding, membrane protein, trifluoroethanol, NMR, bicelles, lipid, detergent

*Motivation and Aim:* Membrane proteins are a common target for drugs, and they need thorough structural studies to further design the specific medicines. However, prior to structural investigation a synthesized membrane protein should be brought into its native state, and this process is usually hard and tricky. To date, no universal protocol for membrane protein refolding is established, and many structural studies of membrane proteins start from an empirical and costly search of working refolding conditions and an appropriate membrane-like environment. In this work we aimed to develop a rational approach to refolding recombinant helical membrane proteins from the trifluoroethanol (TFE)-water mixtures, subsequently embedding them into lipid bicelles.

*Methods and Algorithms:* Mixtures of lipid, detergent and model protein were solubilized in TFE-water mixtures of different ratios, the radii of aggregates and molecule interactions were examined by high-resolution nuclear magnetic resonance (NMR) spectroscopy and dynamic light scattering (DLS). DMPC lipid, DHPC and CHAPS detergents were used to form membrane-like environments; single-helical transmembrane domains of TrkA; TLR9 receptors and the voltage-sensing domain of KvAP channel played the role of model proteins.

*Results:* We studied the behavior of lipid-protein-detergent associates in TFE-water mixtures with contents, gradually altering from 100 % TFE to 100 % water and established the key TFE/water ratios interesting for the refolding. Thus, the points of lipid aggregation, lipid-protein and lipid-detergent-protein complex formation were determined. Based on this data we came up with the possible protocols of refolding, tested their efficiency on the model proteins and proposed two protocols which are a promising starting point for newly studied proteins. The folding of tested proteins was controlled by NMR.

*Conclusion:* Upon adding water to the TFE solution of lipid and detergent the membrane protein interacts with lipids and then gets incorporated into lipid particles; at high water contents the detergent also enters the particles. Nevertheless, bicelles are formed only at 100 % water and even 10 % of TFE alters their morphology. TFE-water ratio as well as the moment of adding the detergent to the mixture can strongly affect the efficiency of refolding; solubilization of a membrane protein in a lipid-detergent mixture at 20–30 % or 50 % water appears to be a preferable refolding protocol.

*Acknowledgements:* This study was supported by the Russian Science Foundation (grant No. 22-24-00593), <https://rscf.ru/project/22-24-00593/>.

## Protein design and crystallization trials for constitutively active CysLT2R with oncogenic L129Q mutation

Kotova T.<sup>1\*</sup>, Shevtsov M.<sup>1</sup>, Gusach A.<sup>1</sup>, Marin E.<sup>1</sup>, Luginina A.<sup>1</sup>, Borshchevskiy V.<sup>1</sup>, Mishin A.<sup>1</sup>, Cherezov V.<sup>1,2</sup>

<sup>1</sup> *Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

<sup>2</sup> *Bridge Institute, Department of Chemistry, University of Southern California, Los Angeles, USA*

\* *Kotova.TV@phystech.edu*

**Key words:** G protein-coupled receptors, cysteinyl leukotrienes, uveal melanoma

*Motivation and Aim:* CysLT1R and CysLT2R are G protein-coupled receptors (GPCRs) activated by cysteinyl leukotrienes. They are key inflammatory mediators in the human body and stimulate bronchial muscle constriction as well as immune cells migration, mucus secretion and edema formation, thus playing an important role in various inflammation-related disorders, such as asthma, allergic rhinitis and urticaria. Additionally, through the immune response mediation CysLT1-2Rs are involved in cardiovascular and neurodegenerative diseases and several types of cancer. L129Q mutation transforms CysLT2R to a constitutively active state and is proved to be a driver oncogenic mutation in uveal melanoma [1, 2], meningeal melanocytic tumour [3], leptomeningeal [4] and blue nevi melanocytic tumors [5].

While there are several diverse antagonist families for CysLT1-2Rs there are no inverse agonists targeting CysLT2R L129Q and their design is remaining to be a challenging task. We believe that novel antagonists may be developed into better substitution of existing drugs against asthma and inverse agonists may be developed into first in class drugs against uveal melanoma, meningeal melanocytic tumour, and leptomeningeal and blue nevi melanocytic tumors.

*Methods:* We optimized primary sequence of CysLT-L129Q mutant receptor with stabilizing point mutations, truncations and fusion tags. Sf9 insect cells were used for protein expression. Receptors were purified with metal-affinity chromatography and checked for its thermostability and homogeneity. Monodisperse protein is being crystallized in lipidic cubic mesophase.

*Results:* Previously our team has solved several high-resolution crystal structures of CysLT1 and CysLT2 receptors [6, 7] in complexes with antiinflammatory compounds Zafirlukast and Pranlukast for the first subtype, and 3 drug candidate compounds for CysLT2R (see publications 1 and 2). Given 3D-structures helped us to perform structure-based drug design with experimental characterization by cell-based functional assays. After extensive virtual ligand screening with ENAMINE we have obtained five novel antagonist chemotypes with relatively small MW about 400 Da and cLogP about 3.5 that gives the room for further optimization. Two of them show inverse agonism against CysLT2R-L129Q mutation confirmed by functional assays, but nevertheless we want to obtain more compounds with independent scaffolds and better potencies, while lacking of the off-target effects. In order to do this, we are aimed at structure determination for CysLT2R-L129Q mutant form (not a wild-type protein). We believe, that this structure

will serve as a better template for the structure based drug design and will give more explanation for the molecular mechanism of the CysLT2R-L129Q constitutive activity and corresponding inverse agonism. Here we present our protein design strategy for re-optimizing the mutant protein and results of our crystallization trials.

*Acknowledgements:* The work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-00337-20-03, project FSMG-2020-0003).

### *References*

1. Ceraudo E. et al. Uveal Melanoma Oncogene *CYSLTR2* Encodes a Constitutively Active GPCR Highly Biased Toward Gq Signaling. *bioRxiv*. 2019;663153. doi: 10.1101/663153.
2. Moore A.R. et al. Recurrent activating mutations of G-protein-coupled receptor *CYSLTR2* in uveal melanoma. *Nat Genet*. 2016;48:675-680. doi: 10.1038/ng.3549.
3. Küsters-Vandeveldel H.V.N. et al. Whole-exome sequencing of a meningeal melanocytic tumour reveals activating *CYSLTR2* and *EIF1AX* hotspot mutations and similarities to uveal melanoma. *Brain Tumor Pathol*. 2018;35:127-130. doi: 10.1007/s10014-018-0308-1.
4. van de Nes, Johannes A P et al. Activating *CYSLTR2* and *PLCB4* mutations in primary leptomeningeal melanocytic tumors. *J Invest Dermatol*. 2017;137(9):2033-2035. doi: 10.1016/j.jid.2017.04.022.
5. Möller I. et al. Activating cysteinyl leukotriene receptor 2 (*CYSLTR2*) mutations in blue nevi. *Mod Pathol*. 2017;30:350-356. doi: 10.1038/modpathol.2016.201.
6. Gusach A. et al. Structural basis of ligand selectivity and disease mutations in cysteinyl leukotriene receptors. *Nat Commun*. 2019;10:5573. doi: 10.1038/s41467-019-13348-2.
7. Luginina A. et al. Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs. *Sci Adv*. 2019;5:eaax2518. doi: 10.1126/sciadv.aax2518.
8. Sadybekov A.A. et al. Structure-based virtual screening of ultra-large library yields potent antagonists for a lipid GPCR. *Biomolecules*. 2020;10(12):1634. doi: 10.3390/biom10121634.

## Structure and dynamics of the SARS-CoV-2 envelope protein

Kuzmin A.\*, Gushchin I.

*Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

\* *kuzmin.as@phystech.edu*

**Key words:** SARS-CoV-2, envelope protein, membrane protein, membrane curvature, palmitoylation, glycosylation, molecular dynamics

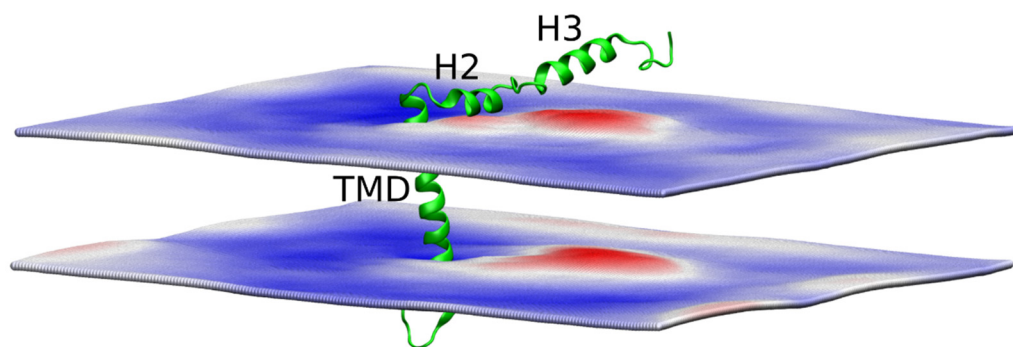
*Motivation and Aim:* SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) is a strain of coronavirus, which caused the COVID-19 (coronavirus disease 2019) pandemic. SARS-CoV-2 virus particle contains four main structural proteins: S (“Spike”) is responsible for attachment and fusion with the host cell; M (“Membrane”) plays the key role in assembly of virions; N (“Nucleocapsid”) packages RNA; E (“Envelope”) promotes assembly of virions, suppression of the stress response of the host cells, efficient transport of the virions along the secretory pathway and demonstrates ion channel activity. Coronaviral E proteins are small, integral membrane proteins of 75-109 amino acids, which have at least one helical transmembrane domain (TMD) and a long amphipathic region comprising one or two  $\alpha$ -helices at the C-terminus. SARS-CoV-2 E protein  $\alpha$  consists of 75 amino acids, and its sequence is 95 and 36 % identical to those of SARS-CoV and MERS-CoV E proteins respectively. It’s also known that E proteins may undergo palmitoylation and glycosylation, although the role of these modifications are not quite clear. In infected cells, E protein is localized at the intracellular trafficking endoplasmic reticulum (ER)-Golgi region and mostly at the intermediate compartment between ER and ER-Golgi intermediate compartment (ERGIC) [1]. Currently, the experimental structure of the full-length wild-type SARS-CoV-2 E protein is not available, and the influence of above-mentioned modifications on it has not been studied. Thus, the goal of the present study was to characterize the structure and dynamics of the SARS-CoV-2 envelope protein in the membrane depending on its contents and post-translational modifications.

*Methods and Algorithms:* The protein behavior was simulated using molecular dynamics. The Martini force field [2] was used for coarse-grained simulations, and the CHARMM36m force field [3] was used for all-atom simulations. All MD simulations were run with GROMACS [4].

*Results:* We have extensively characterized the structure and dynamics of the E protein in monomeric form, and studied its behavior in the pentameric assembly. In particular, we show that TMD has a preferable orientation in the membrane, while H2 and H3 reside at the membrane surface (Fig. 1). Orientation of H2 is strongly influenced by palmitoylation of cysteines Cys40, Cys43, and Cys44. Glycosylation of Asn66 affects the orientation of H3. We also observe that the E protein both generates and senses the membrane curvature, preferably localizing with the C-terminus at the convex regions of the membrane.

*Conclusion:* Our results highlight the role of the E protein in the infection process and explain the low abundance of the E protein in assembled virions. Localization of the protein to curved regions may be favorable for assembly of the E protein oligomers,





**Fig. 1.** Induction of membrane curvature by the SARS-CoV-2 envelope protein. Time-averaged position of the bilayer boundaries is shown. The white color of the bilayer corresponds to the unperturbed position of the membrane, blue – lowered, red – elevated

whereas induction of curvature may facilitate the budding of the viral particles. Flexibility of the C-terminus may be needed for interactions with the host proteins. Overall, these findings are in accordance with experimental observations, while providing a detailed description of the E protein structure and paving the way for development of therapies targeting the E protein [5].

*Acknowledgements:* We are grateful for the support to the Ministry of Science and Higher Education of the Russian Federation (agreement 075-03-2022-107, project FSMG-2021-0002).

### References

1. Ruch T.R., Machamer C.E. The coronavirus E protein: assembly and beyond. *Viruses*. 2012;4(3):363-382.
2. Marrink S.J. et al. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B*. 2007;111(27):7812-7824.
3. Huang J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*. 2017;14(1):71-73.
4. Abraham M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25.
5. Kuzmin A. et al. Structure and dynamics of the SARS-CoV-2 envelope protein monomer. *Proteins*. 2022;90(5):1102-1114.

## Study of membrane permeability for various peptides by steered molecular dynamics

Likhachev I.V.<sup>1,2</sup>, Balabaev N.K.<sup>2</sup>, Galzitskaya O.V.<sup>1,3\*</sup>

<sup>1</sup> *Institute of Protein Research, RAS, Pushchino, Russia*

<sup>2</sup> *Institute of Mathematical Problems of Biology, RAS, Keldysh Institute of Applied Mathematics, RAS, Pushchino, Russia*

<sup>3</sup> *Institute of Theoretical and Experimental Biophysics, RAS, Pushchino, Russia*

\* [ogalzit@vega.protres.ru](mailto:ogalzit@vega.protres.ru)

**Key words:** amyloidogenic peptides, cell-penetrating peptide, centre of mass, force, bilayer, repetition

*Motivation and Aim:* Natural membranes are involved in the regulation of many processes, so the composition and ratio of their molecular components are diverse. The membranes are based on phospholipid molecules, and one of the most common components is cholesterol, which has a significant effect on the physicochemical properties of the lipid matrix. Membrane active peptides serve many important functions in nature as they act as antimicrobials, channels, transporters or hormones. There is still no answer to the mechanism of penetration of antimicrobial peptides through the membrane bilayer. Several mechanisms for such a process have been proposed. It is necessary to understand whether it is possible by the steered molecular dynamics method to determine the ability of peptides of different composition and length to pass through a membrane bilayer. The purpose of the work is to find out which peptides most easily penetrate the cell membrane.

*Methods and Algorithms:* The PUMA [1-3] and PUMA-CUDA software packages were used as a molecular dynamics simulation program. AMBER force field was used as well as the TIP3P water model. Due to the support of periodic boundary conditions, an infinite cell membrane in an aqueous environment is simulated. The passage of peptides through the membrane at a constant rate of 0.1, 0.05 and 0.01 Å/ps was studied due to the force applied to the center of mass of the peptide or its terminal atom. The possibility of exerting force on two subsystems was added to the PUMA-CUDA software package. One subsystem, the cell membrane, was fixed at the center of mass. The second system, the peptide, was subjected to a force effect, leading to movement through the cell membrane at a constant speed. The force action was applied to the center of mass, as well as to the atom.

*Results:* In this work, we developed a technique for constructing bilayers of arbitrary dimensions, hydration under periodic boundary conditions, adding peptides of a given sequence to them, and force pulling the peptide through the resulting bilayer. 656 steered molecular dynamics calculations were carried out for pulling seven amyloidogenic peptides with antimicrobial potential and mono-peptides (homo-repeats consisting of 10 residues of the same amino acid: Poly(Ala), Poly(Leu), Poly(Met), Poly(Arg), Poly(Glu)) with various sequences through the membrane. Among the 15 studied peptides, the peptides exhibiting the least force resistance when passing through the bilayer were found, and the maximum reaction occurs at the boundary of the membrane bilayer entry.

*Conclusion:* The pulling force per atom is always lower than the force applied to the center of mass, due to the fact that when pulling out the atom, the peptide is pulled into a line, thereby the area of resistance to the peptide is smaller. Also, analysis of datasets in [3] would suggest that simultaneous duplication of several domains underlies the formation of repetitive regions, while duplication of a single domain is rare. This fact may explain the predominance of four- and six-domain ribosomal S1 proteins. At the same time, the parts of *rpsA* genes encoding adjacent domains are more identical what correlates with the data suggesting that duplication occurs mainly in the middle of the protein chain between other repeats. The maximum resistance force depends at least on two reasons: 1) partial charges of atoms; 2) amino acid resistance area. The best correlation between the maximum reaction force of the membrane and the calculated parameters corresponds to the instability index.

*Acknowledgements:* The study is supported by the Russian Science Foundation, grant No. 18-14-00321 (Galzitskaya O.).

### *References*

1. Lemak A.S., Balabaev N.K. A comparison between collisional dynamics and brownian dynamics. *Mol Simul.* 1995;15(4):223-231.
2. Glyakina A.V. et al. Determination of the Most Stable Packing of Peptides from Ribosomal S1 Protein, Protein Bgl2p, and A $\beta$  Peptide in  $\beta$ -Layers During Molecular Dynamics Simulations. *Methods Mol Biol.* 2022;2340:221-233.
3. Glyakina A.V. et al. Comparative mechanical unfolding studies of spectrin domains R15, R16 and R17. *J Struct Biol.* 2018;201:162-170.

## Changes in the structure and dynamics of the intracellular domain of Toll-like receptors 1 under the action of Zn<sup>2+</sup>

Lushpa V.A.<sup>1,2</sup>, Goncharuk M.V.<sup>2</sup>, Zalevsky A.O.<sup>2</sup>, Talyzina I.A.<sup>2,3</sup>, Luginina A.P.<sup>1</sup>, Vakhrameev D.D.<sup>1</sup>, Shevtsov M.B.<sup>1</sup>, Goncharuk S.A.<sup>2</sup>, Arseniev A.S.<sup>2</sup>, Borshchevskiy V.I.<sup>1</sup>, Mineev K.S.<sup>1,2</sup>

<sup>1</sup> *Moscow Institute of Physics and Technology (national research university), Moscow, Russia*

<sup>2</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

<sup>3</sup> *Center for Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia*

\* *lushpa@phystech.edu*

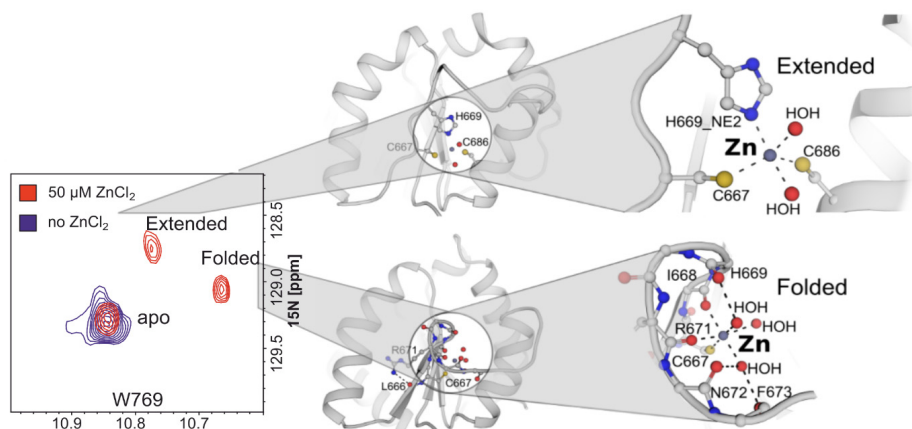
**Key words:** Toll-like receptors, Solution-state NMR, molecular modeling, spatial structure, Zn binding

*Motivation and Aim:* Toll-like receptors (TLRs) are key players in the innate immune response and may serve as targets against inflammatory, neurodegenerative, and autoimmune diseases [1]. The mammalian TLR family includes 13 members, 10 of which are found in humans. These receptors can recognize various molecular patterns associated with bacterial and viral pathogens [2]. The molecular mechanism of receptor activation is considered to be known: ligand binding causes the receptor dimerization, which activates the interaction of intracellular domains (TIR domains) with adapter proteins. In particular, TLR1 forms a complex with TLR2 when bound to lipopeptides or lipoteichoic acid.

Despite the large amount of data on TLR, many questions remain unresolved when considering the activation of the receptor from within the cell. Primarily, it is not known why the TIR domain interacts with adapter proteins exclusively in the dimeric state and how the interaction occurs, why TIR domains do not homodimerize *in vitro* (with the exception of TLR10). In present work the results of a study of the cytoplasmic domain of TLR1 in solution and crystalline form, and the effect of metal ions on the functioning of the receptor.

*Methods and Algorithms:* Methods for obtaining and purifying the protein are described in detail in the article [3]. Selected parameters were used for optimal protein production and subsequent two-stage purification using MCAC followed by sample concentration. Triple resonance spectra were used to assign the protein backbone and side chains by NMR. The obtained data were used to calculate the spatial structure in automatic mode with manual correction in regions with increased mobility of the protein chain. Metal binding experiments were analyzed by NMR spectroscopy. More detailed methods for studying the structure of TLR and the metal-binding activity of the protein are provided in the article [4].

*Results:* Data on the structure and dynamics of TIR TLR1 in solution were obtained. Studies of the TIR TLR1 metal-binding capacity revealed the formation of a TIR TLR1/Zn<sup>2+</sup> complex with the nanomolar dissociation constant mediated by interactions with residues C667 and C686. Possible structures of the complex were predicted *in silico* (Fig. 1). Functional assays for the heterodimeric TLR1/2 receptor showed an effect of zinc concentration on TLR1 activity. The introduction of the C667A mutation leads to a decrease in receptor activity [4].



**Fig. 1.** Overlay of fragments of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra corresponding to the region with signal W769 indole NH. Each zinc-bound state has its own mode of coordination (obtained *in silico*)

**Conclusion:** It was found that the spatial structure of TIR TLR1 in solution differs from the crystal structure in the BB-loop region due to non-native disulfide bonds that occur during crystallization. The resulting protein structure suggested the presence of a zinc-binding site, that was experimentally confirmed. The results obtained in this work suggest that the zinc-binding ability of the TLR1-TIR domain is crucial for receptor activation.

**Acknowledgements:** This study was supported by the Russian Science Foundation (grant No. 22-14-00020).

### References

1. Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol.* 2001;1:135-145.
2. Botos I., Segal D.M., Davies D.R. The structural biology of Toll-like receptors. *Structure.* 2011;19:447-459.
3. Goncharuk M., Lushpa V., Gocharuk S.A., Arseniev A.S., Mineev K.S. Sampling the cultivation parameter space for the bacterial production of TLR1 intracellular domain reveals the multiple optima. *Protein Expr Purif.* 2021;181:105832.
4. Lushpa V., Goncharuk M. et al. Modulation of Toll-like receptor 1 intracellular domain structure and activity by  $\text{Zn}^{2+}$  ions. *Commun Biol.* 2021;4:1003.

## Allosteric ligand subpocket of S1P<sub>5</sub> as a determinant of inverse agonism and ligand specificity

Lyapina E.<sup>1\*</sup>, Marin E.<sup>1</sup>, Gusach A.<sup>1</sup>, Orekhov Ph.<sup>1</sup>, Gerasimov A.<sup>2</sup>, Luginina A.<sup>1</sup>, Vakhrameev D.<sup>1</sup>, Kovaleva M.<sup>1</sup>, Khusainov G.<sup>1</sup>, Khorn P.<sup>1</sup>, Shevtsov M.<sup>1</sup>, Kovalev K.<sup>1</sup>, Okhrimenko I.<sup>1</sup>, Popov P.<sup>1</sup>, Gushchin I.<sup>1</sup>, Rogachev A.<sup>1,3</sup>, Gordeliy V.<sup>1,4</sup>, Borshchevskiy V.<sup>1</sup>, Mishin A.<sup>1</sup>, Cherezov V.<sup>1,5</sup>

<sup>1</sup> *Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

<sup>2</sup> *Vyatka State University, Kirov, Russia*

<sup>3</sup> *Joint Institute for Nuclear Research, Dubna, Russia*

<sup>4</sup> *Institut de Biologie Structurale (IBS), Université Grenoble Alpes, CEA, CNRS, Grenoble, France*

<sup>5</sup> *Bridge Institute, Department of Chemistry, University of Southern California, Los Angeles, USA*

\* *elizaveta0503@gmail.com*

**Key words:** GPCR, S1P<sub>5</sub>, X-ray structure, inverse agonist

*Motivation and Aim:* S1P is a bioactive lysophospholipid targeting GPCRs. It acts via five different subtypes of S1P receptors (S1PR)—S1P<sub>1-5</sub>. Despite sequence similarity, these receptors differ in expression profiles: S1P<sub>1-3</sub> receptors are abundant, S1P<sub>4</sub> is limited to the immune system, and S1P<sub>5</sub> is mainly present in nervous and immune systems. The last one regulates cell migration for NK cells and oligodendrocyte progenitors, playing a role in autoimmune, neurodegenerative and oncogenic disorders. Although FDA approved multiple non-selective (fingolimod) or dual S1P<sub>1</sub>/S1P<sub>5</sub> agonists, few selective modulators were available until the last year. Lack of these modulators leads to constraints in S1P<sub>5</sub> signaling research. To speed up the design of S1P<sub>5</sub> ligands, an inactive structure of this receptor can contribute.

*Methods:* Optimized for insect cell expression, the human *S1PR5* gene was fused with a partner protein in the 3rd intracellular loop, expression tags on N-term, and a C-term cleavable decahistidine tag. This construct was expressed in Sf9 cells using the baculovirus expression system, purified with metal-affinity chromatography and co-crystallized in the lipid cubic phase with an inverse agonist. The crystals undergone room temperature serial femtosecond crystallography (SFX) data collection at Pohang Accelerator Laboratory X-Ray Free-Electron Laser (PAL-XFEL) in South Korea.

*Results:* The crystal structure was solved at a 2.2 Å resolution in the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> space group with identical unit cell constants as the single-crystal data. The receptor crystallized with two monomers per asymmetric unit, forming an antiparallel dimer through the TM4-TM4 interface. The structure shares the classical Edg-receptor architecture with seven transmembrane helices bundle, N-terminal and C-terminal helices. Receptor's microswitches conformations correspond to inactive position, indirectly locked by the ligand.

*Conclusions:* This inactive-state structure of S1P<sub>5</sub> in a complex with an inverse agonist. It reveals an uncommon allosteric binding mode of ligand targeting the variable subpocket of an S1PR and displays the way to design inverse agonists.

*Acknowledgements:* The work was supported by the Russian Ministry of Science and Higher Education grant No. 075-15-2021-1354.

## Structure of Toll-like receptors: the input from solution NMR spectroscopy

Mineev K.<sup>1\*</sup>, Goncharuk S.<sup>1</sup>, Goncharuk M.<sup>1</sup>, Lushpa V.<sup>1,2</sup>, Kornilov F.<sup>1,2</sup>, Shabalkina A.<sup>1,2</sup>, Talyzina I.<sup>1,3</sup>, Zalevsky A.<sup>1</sup>, Lin C.<sup>4</sup>, Wang X.<sup>4</sup>, Volynsky P.<sup>1</sup>, Arseniev A.<sup>1</sup>, Luginina A.<sup>2</sup>, Vakhrameev D.<sup>2</sup>, Borshchevskiy V.<sup>2</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Russia

<sup>3</sup> Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>4</sup> Laboratory of Chemical Biology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China

\* [konstantin.mineev@gmail.com](mailto:konstantin.mineev@gmail.com)

**Key words:** Toll-like receptors, transmembrane domains, TIR domains, juxtamembrane regions, membrane

*Motivation and Aim:* Toll-like receptors (TLRs) are the key players in the innate immune response. TLRs recognize various pathogen-associated molecular patterns and induce the inflammation. They are involved in severe diseases, including cancer, which makes them prospective targets for the drug design. Despite their obvious importance, the structure of TLRs is not well investigated, especially of the transmembrane and intracellular parts. In the present work, we utilized protein engineering, NMR spectroscopy, computer simulations and functional assays to study the spatial structures of TLR transmembrane domains, cytoplasmic juxtamembrane regions and intracellular TIR domains.

*Results:* All the TLRs include the portion of hydrophobic residues in their cytoplasmic juxtamembrane regions. We solved the structures of four protein constructs, which encompass the transmembrane and juxtamembrane domains of TLR2, TLR3, TLR5 and TLR9 in a membrane mimetic environment. According to our data, the representatives of all the TLR subfamilies except for TLR4 form a short  $\alpha$ -helix in cytoplasmic juxtamembrane regions. The bioinformatic analysis reveals that these juxtamembrane domains are more conservative than the transmembrane ones. Functional assays demonstrate that the correct amino acid sequences of juxtamembrane domains are essentially important for the receptor activation. The intracellular TIR domain is another important part of TLR, which is believed to launch the downstream signaling via the interaction with the adapter proteins. Here we solved the solution structure of TLR1 TIR domains and studied its interactions with TLR2. As we found out, TLR1 TIR domains are capable of binding the zinc ions with nanomolar affinity in two different modes. Using mutagenesis, we identified the key cysteine residues responsible for the zinc binding; the structure of zinc-bound TLR1 states were proposed by computer simulations. According to the experiments with living cells, one of the zinc binding sites is functionally relevant, its blocking completely inhibits the activity of full length TLR1/TLR2 receptor.

*Conclusion:* NMR spectroscopy is a potent technique that can provide the structural data on the organization of single-pass membrane proteins, which is impossible to obtain using the alternative approaches. Application of NMR to the TLR systems allowed identifying several key points, essential for the correct functioning of these receptors:

the proper sequence of cytoplasmic juxtamembrane regions, and zinc binding by the intracellular globular domains. The obtained structures provide an insight on how the active state of full-length TLR could look like and allows putting forward the hypotheses on the mechanics of the receptor activation.

*Acknowledgements:* K. Mineev thanks the grant MD-2834.2022.1.4 from the grant council of the President of Russian Federation for the support.



## Evaluation of the effect of site-directed mutagenesis on the affinity of anti-PD1 nanobody by *in silico* modeling

Mirzaei M.<sup>1</sup>, Mirhoseini S.<sup>1</sup>, Heidari M.M.<sup>1</sup>, Khatami M.<sup>1</sup>, Orlov Y.L.<sup>2,3</sup>

<sup>1</sup> Department of Biology, Faculty of Sciences, Yazd University, Yazd, Iran

<sup>2</sup> Institute of Digital Medicine I.M. Sechenov First Moscow State Medical University, Moscow, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

**Key words:** site-directed mutagenesis, affinity, Anti-PD1 nanobody, *in silico* modeling

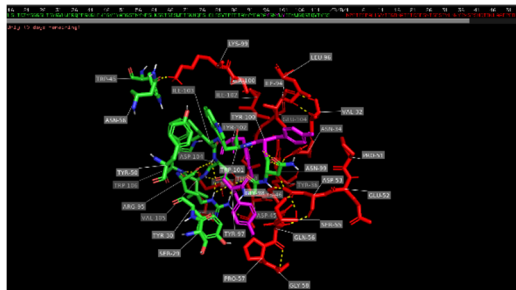
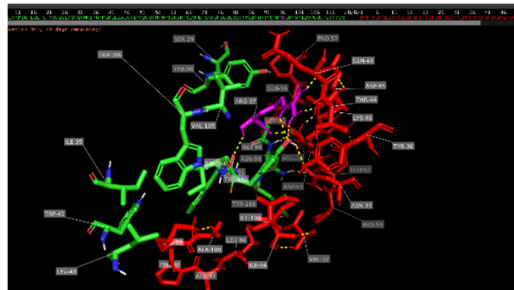
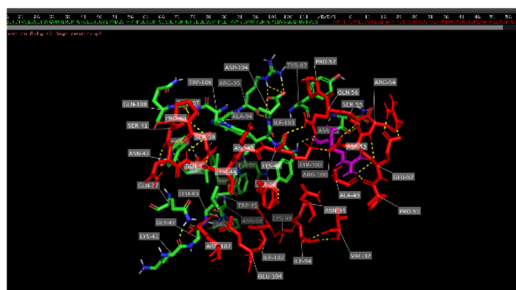
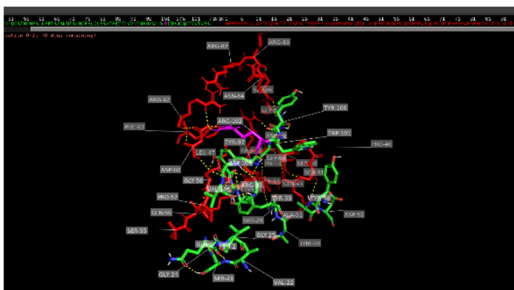
**Motivation and Aim:** Variable domain of heavy chain (VHH) or single domain antibody, also called nanobody (Nb), is the smallest fragment of an antibody (15 kDa) that can bind to antigen [1]. Nbs consist of three complementary determining regions (CDRs) and four framework regions that separate CDRs. CDRs are regions that bind to the antigen [2]. Low immunogenicity, high stability and solubility are some advantages of Nbs. Also, they are well expressed in microorganisms. Nbs have many applications, especially in medicine, which is utilized to diagnose and treat diseases [3,4]. A method for the production of Nb is grafting CDRs from non-camelid antibodies to Nb frameworks [5]. Programmable cell death 1 (PD-1) is an immune checkpoint protein on T cells that plays a key role in modulating immune responses. The interaction of PD-1 with its ligands (PD-L1 and PD-L2) induces common inhibitory signals which leads to the reduction of the activity of T cells. PD-L1 and PD-L2 have increased expression in tumor cells, enabling tumor cells to escape T cell-induced immune responses [6]. Blocking the interaction of PD-1 with its ligands by anti-PD1 Nbs can increase the activity of T cells, which is effective in treating cancer. The type of amino acids at the site of antigen and CDR interaction is important. CDR3 is a significant site of antigen binding. Substitution of the neutral or negative amino acid with a positive amino acid in CDR3 is one way to improve the affinity of CDR3 to the target antigen [7, 8].

**Methods and Algorithms:** In this study, CDR grafting method was utilized to produce anti-PD1 Nb. SWISS-MODEL server (<http://swissmodel.expasy.org>) was used for modeling the three-dimensional structure of anti-PD1 Nb, PD1 antigen and mutated anti-PD1 Nbs. Pymol software was applied to determine the Nb-antigen binding site. Then, HADDOCK server was used to investigate the interaction between Nbs and antigen. HADDOCK requires the atomic structures of each of the subunits of the complex in the PDB format which were obtained from the SWISS-MODEL. The HADDOCK score is a weighted sum of buried surface area and various energies, including van der Waals, electrostatic, desolvation, and restraint violation energy. Electrostatic and van der Waals forces play an important role in the interaction between two proteins. The electrostatic force represents the bond created between two charged amino acids of proteins. The van der Waals force is less powerful, but it is especially important because of the large number of links. The more negative score of these two energies and the more positive score of the buried surface area indicates the better affinity of the Nb.

**Results:** HADDOCK results showed that some of the mutated Nbs improved the affinity (Table 1). Evaluation of optimization strategies for the affinity of Nb indicated that conversion of three tyrosines of CDR3 to arginine improved affinity (Fig. 1–4).

**Table 1.** Comparison of affinity of Nbs according to HADDOCK

	Van der Waals energy	Electrostatic energy	Buried Surface Area
Anti-PD1 Nb	-44.7 +/- 2.8	-89.7 +/- 32.1	1249.6 +/- 98.4
Mutated Nb (1)	-45.1 +/- 0.6	-102.5 +/- 15.7	1298.9 +/- 126.2
Mutated Nb (2)	-45.8 +/- 3.6	-112.5 +/- 13.8	1388.6 +/- 86.0
Mutated Nb (3)	-52.6 +/- 4.5	-123.9 +/- 6.9	1434.9 +/- 65.6

**Fig. 1.** Overview of the interaction of anti-PD1 Nb and it's antigen in Pymol software**Fig. 2.** Overview of the interaction of mutated anti-PD1 Nb (1) and it's antigen in Pymol software**Fig. 3.** Overview of the interaction of mutated anti-PD1 Nb (2) and it's antigen in Pymol software**Fig. 4.** Overview of the interaction of mutated anti-PD1 Nb (3) and it's antigen in Pymol software

**Conclusion:** Nbs have many advantages and applications, especially in medicine. Optimization strategies provide the possibility of improvement of the Nbs. This study indicates that site-directed mutagenesis can improve the affinity and function of Nbs.

**Acknowledgements:** The study is supported by Yazd University.

### References

1. Sun S., Ding Z., Yang X. et al. Nanobody: a small antibody with big implications for tumor therapeutic strategy. *Int J Nanomedicine*. 2021;16:2337-2356.
2. Asaadi Y., Jouneghani F.F., Janani S., Rahbarizadeh F. A comprehensive comparison between camelid nanobodies and single chain variable fragments. *Biomark Res*. 2021;9(1):87.
3. Hu Y., Liu C., Muyldermans S. Nanobody-based delivery systems for diagnosis and targeted tumor therapy. *Front Immunol*. 2017;8:1442.
4. Deffar K., Shi H., Li L., Wang X., Zhu X. Nanobodies-the new concept in antibody engineering. *Afr J Biotechnol*. 2009;8(12):2645-2652.
5. Wagner H.J., Wehrle S., Weiss E., Cavallari M., Weber W. A two-step approach for the design and generation of nanobodies. *Int J Mol Sci*. 2018;19(11):3444.
6. Lee S.H., Lee H.T. et al. Crystal structure of PD-1 in complex with an antibody-drug tislelizumab used in tumor immune checkpoint therapy. *Biochem Biophys Res Commun*. 2020;527(1):226-231.
7. Muyldermans S. A guide to: generation and design of nanobodies. *FEBS J*. 2021;288(7):2084-2102.
8. Kim H.Y., Stojadinovic A., Izadjoo M.J. Affinity maturation of monoclonal antibodies by multi-site-directed mutagenesis. *Methods Mol Biol*. 2014;1131:407-420.

## Can Ca<sup>2+</sup>-regulated photoproteins perform different functions than bioluminescence?

Natashin P.V.\*, Vysotski E.S.

*Photobiology Laboratory, Institute of Biophysics, SB RAS, Federal Research Center "Krasnoyarsk Science Center, SB RAS", Krasnoyarsk, Russia*

\* [pavelnatashin@mail.ru](mailto:pavelnatashin@mail.ru)

**Key words:** bioluminescence, Ca<sup>2+</sup>-regulated photoproteins, coelenterazine

Ca<sup>2+</sup>-regulated photoproteins are single-chain polypeptide proteins (~22 kDa) that are responsible for the bioluminescence of marine coelenterates. The photoprotein represents an enzyme-substrate complex consisting of apoprotein with tightly but noncovalently bound preoxygenated coelenterazine derivative, 2-hydroperoxycoelenterazine. Blue light emission results from decarboxylation reaction of 2-hydroperoxycoelenterazine initiated by Ca<sup>2+</sup> binding to a protein. All Ca<sup>2+</sup>-regulated photoproteins contain three canonical EF-hand Ca<sup>2+</sup>-binding sites, each consisting of 12 contiguous residues. It allows assigning these proteins to a family of EF-hand Ca<sup>2+</sup>-binding proteins. Despite this fact, the photoproteins demonstrate no significant amino acid sequence homology with other proteins from this large family.

The <sup>15</sup>N-HSQC-NMR measurements carried out for obelin from *Obelia longissima* bound with different ligands offer five distinct conformational states of the photoproteins: (i) apophotoprotein; (ii) photoprotein bound with 2-hydroperoxycoelenterazine; (iii) Ca<sup>2+</sup>-discharged photoprotein, i.e. protein bound with the reaction product, coelenteramide, and Ca<sup>2+</sup>; (iv) Ca<sup>2+</sup>-discharged photoprotein without Ca<sup>2+</sup>; and (v) apophotoprotein bound with Ca<sup>2+</sup>. To date, the crystal structures of four conformations (states ii-v) have been determined. The structures of these ligand-dependent conformations reveal a similar globular scaffold with some distinctions caused by the bound ligand type. Despite the numerous attempts, the crystals of apophotoprotein were not obtained. The cause may be the absence of an ordered three-dimensional structure that is supported by NMR measurements [1] and denaturation studies [2]. Thus, apophotoprotein displays the features characteristic of intrinsically disordered proteins.

Presently, many disordered Ca<sup>2+</sup>-binding proteins have been discovered [3]. They assemble functionally active tertiary structure in response to calcium binding. Based on comparison of different conformational states of photoproteins with Ca<sup>2+</sup>-binding proteins, we here hypothesize that photoproteins may perform another function in host organisms alongside with bioluminescence. In total, five types of Ca<sup>2+</sup>-binding proteins were found in Protein Data Base which spatial structures reveal a high similarity with those of photoproteins. These are neuronal calcium-signaling protein calnexin from the squid *Loligo pealei* (PDB 2CCM), juvenile hormone diol kinase from the silk worm *Bombix mori* (PDB 6KTH, 7PJD), okadaic acid binding protein (OABP) from the marine sponge *Halichondria okadai* (PDB 4WRI), three sarcoplasmic Ca<sup>2+</sup>-binding proteins (SCBPs) from different species (PDB 2SAS, 2SCP, 3AKB) and Ca<sup>2+</sup>-dependent coelenterazine-binding protein (CBP) from *Renilla muelleri* (PDB 2HQ8, 2HPS). All these proteins have similar globular structures and contain four EF-hand motifs but only

three loops with canonical amino acid sequences ensuring  $\text{Ca}^{2+}$  binding. There is only one exception—okadaic acid binding protein does not contain loops capable of binding  $\text{Ca}^{2+}$  owing to replacement of the key residues that provide an oxygen atom for coordination of  $\text{Ca}^{2+}$ . The function of these proteins is diverse. For instance, SCBPs serve as calcium buffers while OABP binds okadaic acid, a marine polyether cytotoxin. JHDK is involved in metabolism of insects; this enzyme catalyzes phosphorylation of juvenile hormone using ATP or GTP but in contrast to photoproteins, calcium ions inhibit its activity. Neuronal calcium-signaling protein calyculin regulates different types of potassium channels and takes part in controlling the release of calcium ions from the endoplasmic reticulum by binding to the ryanodine receptor. After phosphorylation calyculin can be translocated to the cell membrane where the one affects the membrane excitability. In addition, the calyculin injection into photoreceptors of gastropods was capable to reproduce the electrophysiological effects of learning. In the last decade different groups showed that photoprotein aequorin may also administer the cognitive function [4]. Despite a very low sequence homology (18 %),  $\text{Ca}^{2+}$ -loaded conformational state of apo-aequorin and calyculin reveal a very high structural similarity of their main chains (RMSD 3.6 Å). Moreover, the sequence analysis of aequorin and other photoproteins using NetPhos 3.1 software predicts the presence of different sites of phosphorylation with high probability, i.e., similar to calyculin apo-aequorin can be also phosphorylated by protein kinase C. Earlier it was demonstrated that *Aequorea* jellyfish cannot synthesize coelenterazine; the substrate comes through diet [5]. At the same time, aequorin is expressed by jellyfish. Thus, in the absence of coelenterazine the aequorin should be in the apo-conformational state bound with either  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  or  $\text{Ca}^{2+}/\text{Mg}^{2+}$  depending on the affinity of its  $\text{Ca}^{2+}$ -binding sites and consequently can perform some other function, like, for example, calyculin. Although the above mentioned experimental and theoretical findings imply other functions of photoproteins alongside with bioluminescence, only revealing the molecular mechanism of regulation of cognitive processes e.g. by aequorin, can definitely confirm it.

*Acknowledgements:* The study was supported by grant No. 22-14-00125 of the Russian Science Foundation.

### References

1. Lee J. et al. Protein conformational changes in obelin shown by  $^{15}\text{N}$ -HSQC nuclear magnetic resonance. In: Case J.F., Herring P.J., Robison B.H., Haddock S.H.D., Kricka L.J., Stanley P.E. (Eds.). *Bioluminescence & Chemiluminescence 2000*. Singapore: World Scientific Publishing Co., 2001:99-102.
2. Ereemeeva E.V. et al. Ligand binding and conformational states of the photoprotein obelin. *FEBS Letters*. 2012;586:4173-4179.
3. Grzybowska E.A. Calcium-binding proteins with disordered structure and their role in secretion, storage, and cellular signaling. *Biomolecules*. 2018;8:42.
4. Ehlers V.L. et al. Apoaequorin differentially modulates fear memory in adult and aged rats. *Brain Behav*. 2020;10(11):e01832.
5. Haddock S.H.D. et al. Can coelenterates make coelenterazine? Dietary requirement for luciferin in cnidarian bioluminescence. *PNAS*. 2001;98:11148-11151.

## Microbial light-driven ion pumps and channels

Okhrimenko I.\*, Borshchevskiy V.\*

*Research Center for Molecular Mechanisms of Aging and Age-Related Diseases of Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*

\* [okhrimenko.is@mipt.ru](mailto:okhrimenko.is@mipt.ru); [borshchevskiy.vi@phystech.edu](mailto:borshchevskiy.vi@phystech.edu)

**Key words:** rhodopsins, optogenetics, X-ray crystallography, ion pumps, ion channels

*Motivation and Aim:* Rhodopsins are probably the most universal biological light energy transducers and abundant phototrophic mechanisms evolved on Earth [1]. They are found in all the domains of life and in viruses and have a remarkable diversity and potential for biotechnological applications. Channel rhodopsins, H<sup>+</sup> and Cl<sup>-</sup> pumps have become indispensable means of optogenetics and revolutionized neuroscience promising new approaches to the treatment of severe diseases. However, among thousands of identified rhodopsin genes only a few are characterized. This dramatically limits our knowledge of their functions, mechanisms and biotechnological applications. Moreover, high-resolution structures and molecular mechanisms of recently studied new rhodopsins are either not known or limited to non-active states. The amazing scientific and technological potential of rhodopsins is still to be explored.

*Results:* Since the discovery of the first microbial rhodopsin – a light-driven proton pump bacteriorhodopsin (BR), a seven alpha helical membrane protein comprising the retinal, in 1971 [2], BR was one of the major drivers of ground breaking developments in membrane protein research. However, before 2000 there were only several and only archaeal known rhodopsins, functioning as proton, chloride pumps and also sensory rhodopsins. At the end of the XX<sup>th</sup> century not many of us, if any at all, would imagine a new era of the rhodopsin world and their breaking contributions to the optogenetics applications in neuroscience.

*Conclusion:* In 2000 the first bacterial rhodopsin – a proton pump proteorhodopsin was discovered [3]. Then it appeared that there are unexpectedly many rhodopsins and they are present in all the domains of life [4] and even in viruses [5]. It was also unexpected that these ‘seemingly the same’ proteins can execute diverse functions. Ground breaking discoveries and characterization of channel rhodopsins in 2002/2017 [6, 7], sodium pump KR2 in 2013/2015/2019 [8, 9, 10] rhodopsin-guanylyl cyclase in 2014 [11], outward proton pumps xenorhodopsins in 2016/2017 [12, 13], pH [14] and divalent ion dependent rhodopsins support this conclusion.

*Acknowledgements:* The study was supported by the Russian Science Foundation (project No. 21-64-00018).

### References

1. Okhrimenko I.S. et al. Properties of new unexplored microbial rhodopsins. In: Chemistry, structure and function of biomolecules. VI<sup>th</sup> International Conference. Minsk, 2018;52.
2. Henderson R., Unwin P.N.T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*. 1975;257(5521):28-32.
3. Landau E.M., Rosenbusch J.P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *PNAS*. 1996;93(25):14532-14535.
4. Zabelskii D. et al. Structure-based insights into evolution of rhodopsins. *Commun Biol*. 2021;4(1):821.
5. Bratanov D. et al. Unique structure and function of viral rhodopsins. *Nat Commun*. 2019;10(1):4939.

6. Nagel G. et al. Channelrhodopsin-1: a light-gated proton channel in green algae. *Science*. 2002;296(5577):2395-2398.
7. Volkov O. et al. Structural insights into ion conduction by channelrhodopsin 2. *Science*. 2017;358(6366):eaan8862.
8. Inoue K. et al. A light-driven sodium ion pump in marine bacteria. *Nat Commun*. 2013;4:1678.
9. Gushchin I. et al. Crystal structure of a light-driven sodium pump. *Nat Struct Mol Biol*. 2015;22(5):390-395.
10. Kovalev K. et al. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci Adv*. 2019;5(4):eaav2671.
11. Avelar G.M. et al. A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. *Curr Biol*. 2014;24(11):1234-1240.
12. Inoue K. et al. A natural light-driven inward proton pump. *Nat Commun*. 2016;7:13415.
13. Shevchenko V. et al. Inward H<sup>+</sup> pump xenorhodopsin: Mechanism and alternative optogenetic approach. *Sci Adv*. 2017;3(9):e1603187.
14. Maliar N.L. et al. Novel pH-sensitive microbial rhodopsin from *Sphingomonas paucimobilis*. *Dokl Biochem Biophys*. 2020;495(1):342-346.

# All-D-enantiomeric peptide designed for Alzheimer's disease treatment dynamically interacts with amyloidogenic region of membrane-bound amyloid- $\beta$ peptide precursor

Okhrimenko I.<sup>1\*</sup>, Volynsky P.<sup>2</sup>, Zagryadskaya Y.<sup>1</sup>, Kamynina A.<sup>1,2</sup>, Kuzmichev P.<sup>1</sup>, Arseniev A.<sup>1,2</sup>, Efremov R.<sup>1,2,3</sup>, Bocharov E.<sup>1,2\*</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, MIPT, Dolgoprudny, Russia

<sup>3</sup> School of Applied Mathematics, Higher School of Economics, Moscow, Russia

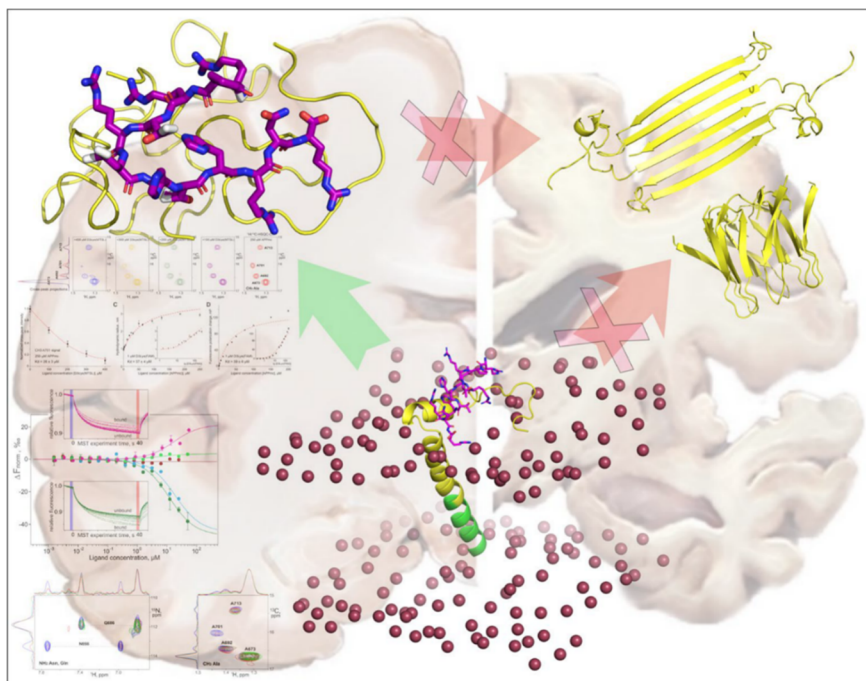
\* okhrimenko.is@mipt.ru; edvbon@mail.ru

**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 resulting in severe dementia. Detailed information on the structure, dynamics, and various intermolecular interactions for biomolecules directly involved in the development of AD is required for the rational development of new biologically active compounds and screening of existing ones to obtain the most effective candidates for medicines [1]. Genetic evidence strongly suggests that aberrant generation and/or clearance of the neurotoxic amyloid- $\beta$  peptide ( $A\beta$ ), being the products of sequential cleavage of amyloid precursor protein (APP), triggers the disease. All-D-enantiomeric peptide D3 and its derivatives were recently selected with the aid of phage display to directly destroy cytotoxic  $A\beta$  aggregates [2]. Currently, one of the D3-like compounds is about to undergo a phase II clinical trial, however, high resolution molecular details of its disease preventing or pharmacological action are not completely 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  $A\beta$  precursor, CTF $\beta$  transmembrane fragment of amyloid precursor protein (APP<sub>672-726</sub>,  $A\beta_{1-55}$  in amyloid- $\beta$  numeration). Namely, heteronuclear NMR spectroscopy showed that D3-peptide directly binds to the amphiphilic near-membrane JM region of  $A\beta_{17-26}$ , which, depending on external conditions, is capable of undergoing a conformational transition from the  $\alpha$ -helical conformation to the  $\beta$ -chain, which is involved in folding of  $\beta$ -amyloid into fibrils. The obtained structural data in agreement with ELISA and Western-blot analyzes, which reveal that D3-peptide dynamically binds in the vicinity of  $\alpha$ -secretase recognition site situated between the extracellular cation-binding domain and JM helix of APP. 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.



**Fig. 1.** Schematic representation using experimentally obtained structural data of the action of the D3-peptide (shown in purple, in the lower part of the figure) on the membrane-associated peptide APP<sub>672-726</sub> (A $\beta$ <sub>1-55</sub>), followed by the formation of D3/A $\beta$  complexes (in the upper left part of the figure) and inhibition formation of toxic A $\beta$  oligomers (in the upper right part of the figure)

**Conclusion:** The data suggest that D-enantiomeric peptide D3 recognizes the amyloidogenic region of APP also before its processing, restricting conformational diversity not compromising its  $\alpha$ -helicity and preventing intermolecular hydrogen bond formation, which would create prerequisites for inhibition of early steps of A $\beta$  conversion into  $\beta$ -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.

**Acknowledgements:** The study is supported by Russian Science Foundation (project No. 20-64-46027).

### References

1. Urban A.S. et al. Structural studies providing insights into production and conformational behavior of Amyloid- $\beta$  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 A $\beta$  Monomer Binding Specificity Enhances Their Potential to Eliminate Toxic A $\beta$  Oligomers. *ACS Chem 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- $\beta$  precursors. *J Med Chem*. 2021;64(22):16464-16479.



## Classification of families of DNA-recognizing protein domains based on structural features of DNA-protein complexes

Panova V.<sup>1\*</sup>, Baulin E.<sup>2</sup>, Karyagina A.<sup>3,4,5</sup>, Alexeevski A.<sup>3,6</sup>, Spirin S.<sup>3,6,7</sup>

<sup>1</sup> Faculty of Bioengineering and Bioinformatics of Moscow State University, Moscow, Russia

<sup>2</sup> Institute of Mathematical Problems in Biology, RAS, Pushchino, Moscow Region, Russia

<sup>3</sup> Belozersky Institute of Moscow State University, Moscow, Russia

<sup>4</sup> Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia

<sup>5</sup> All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

<sup>6</sup> Institute of System Studies, Moscow, Russia

<sup>7</sup> National Research University Higher School of Economics, Moscow, Russia

\* verapanova877@gmail.com

**Key words:** contact type, interaction class, interaction mode, miscellaneous

*Motivation and Aim:* DNA-protein interactions play a central role in such cellular processes as DNA replication, transcription, and repair [1]. Prediction of the interactions between DNA and protein is a significant but not trivial task [2]. Currently, 6637 DNA-protein 3D structures are available in open databases. Classification of structures of DNA-protein complexes makes it possible to effectively study the patterns of DNA-protein recognition. The protein domains are united in families, and domains of the particular family recognize DNA in a similar way.

This work aims to correct errors in the existing version of NPIDB ([3], <https://npidb.belozersky.msu.ru/>), to create a corrected classification of the structures of DNA-protein complexes, to describe structural features, and to make an internal classification of the most abundant families based on corrected data.

*Methods and Algorithms:* The classification is based on the principles from [5]. It classifies structures of protein domains with neighboring DNA double helix. Instead of domains allocated in protein chains according to the SCOP database whose support was discontinued in 2009, now the domains allocated according to the Pfam database [4] are classified. The classification is based on contacts between DNA and protein molecules. Hydrogen bonds, hydrophobic interactions, and water bridges are considered as contacts. The contact type is a pair of contacting structural elements. Protein structural elements are helices, beta-strands, and turn/unstructured segments, while DNA structural elements are the major groove, the minor groove, and the sugar-phosphate backbone. The interaction mode for the domain and domain structure is defined as a list of contact types, and the interaction class for a family of domains is defined as the intersection of interaction modes of all domains of the family. If the interaction class is empty, the family is called *miscellaneous*.

*Results and Conclusions:* A set of Python programs determining interaction modes and interaction classes was written. Only protein structures in complexes with a double DNA helix (with at least 6 complementary pairs), solved by X-ray and CryoEM analysis with a resolution  $< 3 \text{ \AA}$  were considered.

The data were statistically processed, the numbers of structures and domains with each interaction mode and the number of families with each interaction class were obtained. The classification of structures of DNA-protein complexes was significantly improved.

Several particular families (including all six miscellaneous families) have been described in detail. The internal classification of families, based on the superposition of DNA-binding parts of a protein in the domain, was described. For all six miscellaneous families, the reasons why the domains of these families do not have common contact types were found.

### *References*

1. Kulandaisamy A. et al. Dissecting and analyzing key residues in protein-DNA complexes. *J Mol Recognit.* 2018;31(4).
2. Gao M., Skolnick J. From nonspecific DNA-protein encounter complexes to the prediction of DNA-protein interactions. *PLoS Comput Biol.* 2009;5:e1000341.
3. Kirsanov D. et al. NPIDB: nucleic acid-protein interaction database. *Nucleic Acids Res.* 2013;41(D1):D517-D523.
4. Mistry J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419.
5. Zaneagina O. et al. An updated version of NPIDB includes new classifications of DNA-protein complexes and their families. *Nucleic Acids Res.* 2016;45(D1):D144-D153.

## Structure and dynamics of chromatin at nucleosome level: integrating experiments and simulations

Shaytan A.K.

*Department of Biology, Lomonosov Moscow State University, Moscow, Russia*

*shaytan\_ak@mail.bio.msu.ru*

**Key words:** nucleosomes, molecular dynamics simulations, structural bioinformatics, integrative modeling

Recent decades have witnessed tremendous progress in DNA sequencing and availability of complete genomic sequences for many living organisms. The digital-like discrete encoding of the genetic information within the primary sequence of the DNA opened many possibilities for bioinformatic analysis. However, understanding the functioning of the genomes ultimately requires the understanding of the 3D structural interactions between the DNA/RNA, proteins and their dynamics. In eukaryotic organisms, chromatin – the complex of DNA, RNA, histone and non-histone proteins – provides a complex framework for genome functioning and regulation. Deciphering the mechanisms of eukaryotic chromatin operation starts with understanding the structure and dynamics of the nucleosomes – the basic units of chromatin. Nucleosomes comprise an octamer of eight histone proteins and around 150 DNA base pairs wrapped around it. The canonical structure of the nucleosome was revealed by X-ray crystallography 25 years ago. It has since become apparent that at physiological conditions nucleosomes are highly dynamic and this dynamics is instrumental for genome functioning at epigenetic level [1]. However, despite recent progress in structural biology techniques we are still facing conceptual and methodological challenges in describing and understanding how large dynamic macromolecular such as nucleosomes function.

In this report I will showcase how combination of atomistic and coarse-grained supercomputer simulations supplemented with data from various biophysical and biochemical experiments can be used to study the structure and dynamics of chromatin at nucleosome and supranucleosome levels. Based on our works I will first discuss data analysis and modeling approaches that may be used to reconstruct the structures of arbitrary nucleosomes based on hydroxyl-radical footprinting experiments [2, 3]. In such experiments the DNA is cleaved within the nucleosome at solvent-accessible sites. Modeling of this process combined with pseudosymmetry of the nucleosome core allows for the reconstruction of the exact rotational and translational positioning of the DNA in nucleosomes. I will next discuss the application of atomistic supercomputer molecular dynamics (MD) simulations to decipher the functional motions of histones and DNA in nucleosomes [4]. Record long MD simulations at a timescale exceeding 10 microseconds allowed us to analyze the interplay between histone and DNA dynamics that facilitates nucleosome sliding along the DNA and regulates access of transcription machinery to the genetic information. I will conclude with showing how coarse-grained modeling and integrative modeling approaches may be used to model the structure of nucleosome complexes with chromatin proteins and the structure of chromatin fibrils.

*Acknowledgements:* The work was supported by Russian Science Foundation grant No. 18-74-10006.

## References

1. Armeev G.A. et al. Linking chromatin composition and structural dynamics at the nucleosome level. *Curr Opin Struct Biol.* 2019;56:46-55.
2. Shaytan A.K. et al. Hydroxyl-radical footprinting combined with molecular modeling identifies unique features of DNA conformation and nucleosome positioning. *Nucleic Acids Res.* 2017;45(16):9229-9243.
3. Shaytan A.K. et al. Structural interpretation of DNA–protein hydroxyl-radical footprinting experiments with high resolution using HYDROID. *Nat Protoc.* 2018;13(11):2535-2556.
4. Armeev G.A. et al. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nat Commun.* 2021;12:2387.

## Structure and dynamics of REGN10987 Fab interaction with S-proteins of Delta and Omicron SARS-CoV-2 variants

Shenkarev Z.O.<sup>1\*</sup>, Pichkur E.<sup>2</sup>, Nolde D.E.<sup>1</sup>, Varizhuk A.<sup>3</sup>, Kocharovskaya M.V.<sup>1,4</sup>, Kirpichnikov M.P.<sup>1,5</sup>, Lyukmanova E.N.<sup>1,5</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> National Research Center “Kurchatov Institute”, Moscow, Russia

<sup>3</sup> Federal Research and Clinical Center of Physical-Chemical Medicine, FMBA, Moscow, Russia

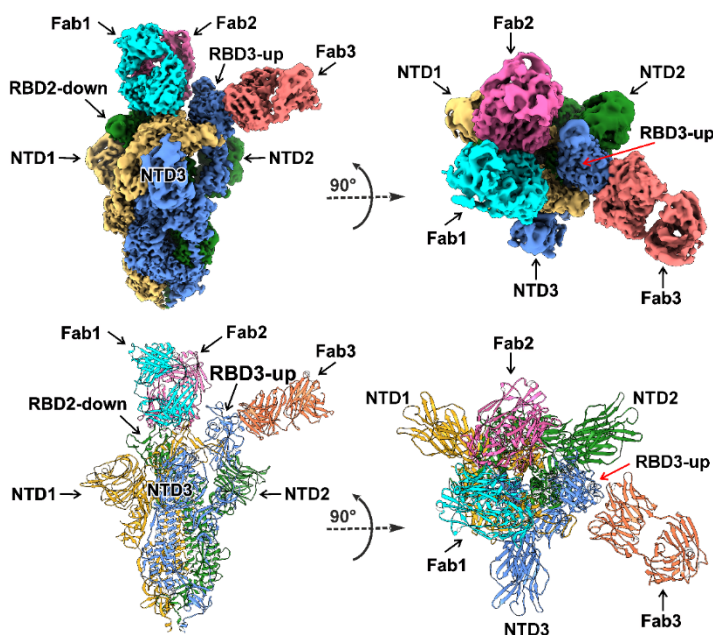
<sup>4</sup> Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia

<sup>5</sup> Interdisciplinary Scientific and Educational School of Moscow University “Molecular Technologies of the Living Systems and Synthetic Biology”, Biological Faculty, Lomonosov Moscow State University, Moscow, Russia

\* zakhar-shenkarev@yandex.ru

**Key words:** coronavirus; neutralizing antibodies; RBD; ACE-2; cryo-EM

Knowledge how neutralizing antibodies recognize various SARS-CoV-2 variants is critical for design of vaccines and antibody-based therapeutics. Here, we reported 2.3 Å cryo-EM structure of full-length trimeric S-protein of the SARS-CoV-2 Delta variant in a complex with recombinant analogue of REGN10987 antibody’s Fab (Fig. 1).



**Fig. 1.** Cryo-EM structure of the full-length Delta SARS-CoV-2 spike protein in a complex with REGN10987 Fab. Cryo-EM map of the S-protein/Fab complex refined to 2.3Å resolution and pseudoatomic model of the complex after local refinement of RBD/Fab regions. NTD – N-terminal domain; RBD – receptor-binding domain; Three protomers (1, 2, and 3) of S-protein are colored by yellow, green, and blue respectively. Corresponding Fabs are colored by cyan, magenta, and coral. RBD1/Fab1 and RBD2/Fab2 complexes are in ‘down’ conformation, while RBD3/Fab3 – in ‘up’

Receptor-binding domain (RBD)/Fab regions were locally refined to 3.2–3.4 Å. Two RBDs were in the ‘down’ state, while third RBD adopted the ‘up’ conformation. Fab interacted with RBDs in both conformations occupying a fragment of the receptor-binding motif and blocking ACE2 recognition. 3D variability analysis revealed high mobility of the RBD/Fab regions. Molecular dynamics simulations explained the differences observed in the Fab interaction with Delta and Omicron RBDs. Our study provides a structural insight into the role of all known to date significant RBD mutations of the Omicron and other SARS-CoV-2 variants in the REGN10987 evasion. Data obtained will be useful for development of new therapeutic antibodies.

*Acknowledgements:* This work was supported by the Russian Science Foundation (the project No. 19-74-30014).

## Membrane environment regulates interaction of three-finger toxins with nicotinic acetylcholine receptors

Shenkarev Z.O.<sup>1,2</sup>, Chesnokov Y.<sup>3</sup>, Zaigraev M.M.<sup>1,2</sup>, Chugunov A.O.<sup>1</sup>,  
Kulbatskii D.S.<sup>1</sup>, Kocharovskaya M.V.<sup>1,2</sup>, Paramonov A.S.<sup>1</sup>, Bychkov M.L.<sup>1</sup>,  
Shulepko M.A.<sup>1</sup>, Kirpichnikov M.P.<sup>1,4</sup>, Lyukmanova E.N.<sup>1,2,4\*</sup>

<sup>1</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*

<sup>2</sup> *Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow Region, Russia*

<sup>3</sup> *National Research Center “Kurchatov Institute”, Moscow, Russia*

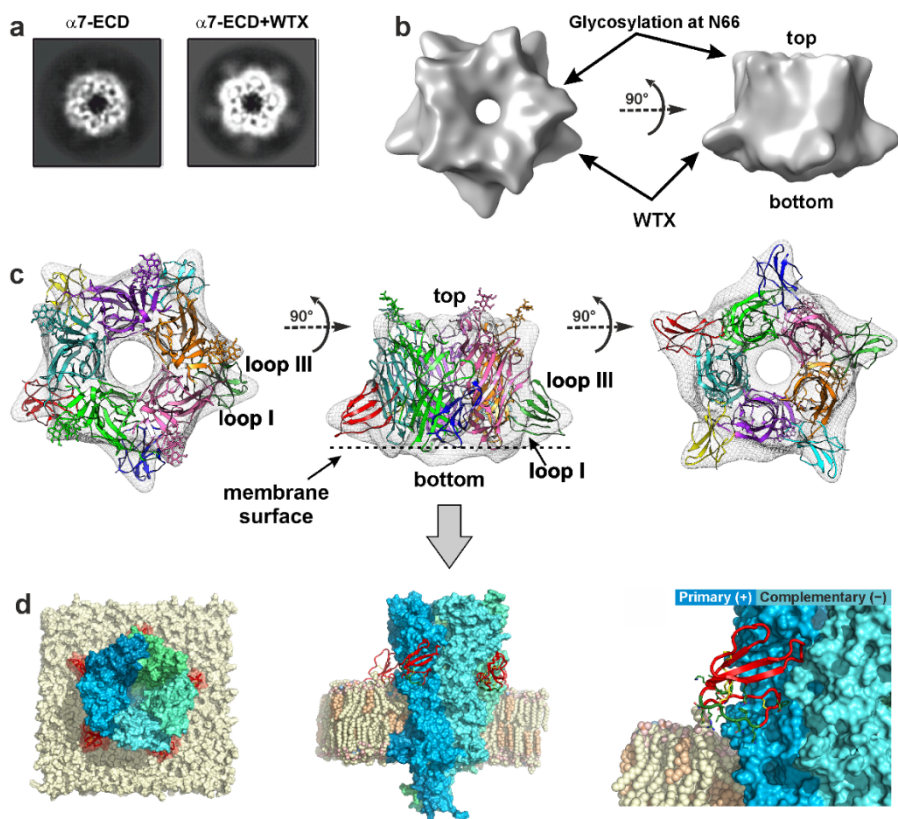
<sup>4</sup> *Interdisciplinary Scientific and Educational School of Moscow University “Molecular Technologies of the Living Systems and Synthetic Biology”, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

\* *ekaterina-lyukmanova@yandex.ru*

**Key words:** snake neurotoxin, nicotinic acetylcholine receptor, three-finger protein, Ly6/uPAR, cryo-EM, membrane catalysis, membrane binding site

Nicotinic acetylcholine receptor of  $\alpha 7$  type ( $\alpha 7$ -nAChR) presented in the nervous and immune systems and epithelium is a promising therapeutic target for cognitive disfunctions and cancer treatment. WTX is a non-conventional three-finger neurotoxin from *Naja kaouthia* venom, targeting  $\alpha 7$ -nAChR with weak affinity. There is no data on interaction mode of non-conventional neurotoxins with  $\alpha 7$ -nAChR. Using  $\alpha$ -bungarotoxin (classical three-finger neurotoxin with high affinity to  $\alpha 7$ -nAChR), we showed applicability of cryo-EM to study interactions of  $\alpha 7$ -nAChR extracellular ligand-binding domain ( $\alpha 7$ -ECD) with its ligands. Cryo-EM structure of the  $\alpha 7$ -ECD/WTX complex together with NMR data on membrane active site in the WTX molecule and mutagenesis data allowed to reconstruct the  $\alpha 7$ -nAChR/WTX structure in the membrane environment. WTX interacts at the entrance to the orthosteric site located at the receptor intersubunit interface and simultaneously forms the contacts with the membrane surface. WTX interaction mode with  $\alpha 7$ -nAChR significantly differs from  $\alpha$ -bungarotoxin's one, which do not contact the membrane. Our study provides evidence of the ‘membrane catalysis’ mechanism for non-conventional neurotoxins.

*Acknowledgements:* This study made under financial support of Russian Science Foundation (project No. 19-74-20163).



**Fig. 1.** 3D reconstruction of  $\alpha 7$ -nAChR/WTX complex determined by cryo-EM and computer modeling. **a** Comparison of the representative top-view 2D class averages for  $\alpha 7$ -ECD and  $\alpha 7$ -ECD/WTX complex. The fine features of pentameric domain complex and density corresponding to five WTX molecules are visible. **b** 3D density map of  $\alpha 7$ -ECD/WTX complex. The features corresponding to the linked glycan fragments and bound WTX molecules are indicated. **c** Model of  $\alpha 7$ -ECD/WTX complex fitted into experimental 3D density map. The position of the toxin's loops I and III and expected position of membrane interface in the full-length channel are shown. **d** Views of the  $\alpha 7$ -nAChR/WTX complex. nAChR subunits are individually colored in a blue-green spectrum; five WTX molecules at the interfaces of two adjacent receptor's subunits are shown by red. Carbon atoms of the membrane lipids are colored as follows: POPC and POPE, pale yellow; cholesterol, pale orange. Other lipid atoms: phosphorus, orange; oxygen, red; nitrogen, blue. Lipids of the proximal part of the membrane are removed for clarity. Membrane-interacting residues of WTX are colored by forest green. WTX interacts by the loops II and III with  $\alpha 7$ -nAChR and by loop I, head-1, and N- and C-termini with the membrane



## Structure features of novel D-amino acid transaminase from *Aminobacterium colombiense*

Shilova S.<sup>1\*</sup>, Nikolaeva A.<sup>2</sup>, Rakitina T.<sup>2</sup>, Matyuta I.<sup>1</sup>, Boyko K.<sup>1</sup>, Bezsudnova E.<sup>1</sup>

<sup>1</sup> *Bach Institute of Biochemistry, Research Centre of Biotechnology, RAS, Moscow, Russia*

<sup>2</sup> *Complex of NBICS Technologies, National Research Center “Kurchatov Institute”, Moscow, Russia*

\* *zavyalovasonya@yandex.ru*

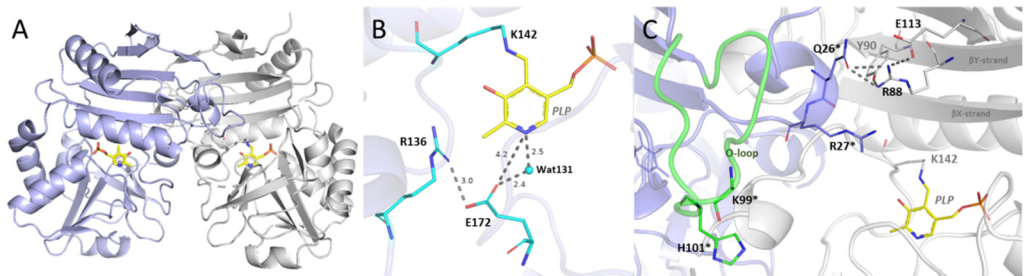
**Key words:** biocatalysts, transaminases, structure biology

**Motivation and Aim:** D-amino acid transaminases (DAATs) are pyridoxal-5'-phosphate (PLP)-dependent enzymes that catalyze reversible and stereoselective transfer of the amino group between D-amino acids and  $\alpha$ -keto acids. DAATs belong to fold type IV of the PLP-dependent enzymes in which the PLP forms a covalent bond with the catalytic Lys residue as well as a number of noncovalent conservative interactions among TAs of fold type IV. In the active site of known DAATs, two modes of the organization of residues responsible for the substrate binding are observed. The first type is observed in the DAAT from *Bacillus subtilis* (bsDAAT, PDB ID: 3DAA) [1], in which  $\alpha$ -carboxylic group is bound by a "carboxylate trap" formed by three residues – Y31 from the  $\beta$ X-strand and R98\*, H100\* from the O-loop. The second type is observed in TAs from *Curtobacterium pusillum* (CpuTA, PDB ID: 5K3W) [2] and *Haliscomenobacter hydrossis* (Halhy, PDB ID: 7P7X) [3]: here the  $\alpha$ -carboxylic group of substrate is bound by two residues, R51\*, K117 in CpuTA and R28\*, R90 in Halhy. The R51\* and R28\* residues are directed to the active site from the adjacent subunit, and the K117 and R90 residues are located on the  $\beta$ Y-strand. DAAT from *Aminobacterium colombiense* (Amico) in the active site possesses residues similar to the bsDAAT (K99\* and H101\* on the O-loop) and both CpuTA and Halhy (R27\* and R88). To understand the Amico's active site organization, structure of holo-form and the R88L variant were determined by the RSA.

**Methods and Algorithms:** The gene encoding Amico was optimized, cloned, and expressed with His-tag in *E. coli* cells. The recombinant form of the enzyme was isolated by Ni-IMAC chromatography, and then the His-tag was cut off using a TEV-protease. Subsequent purification of the enzyme included size-exclusion and anion exchange chromatography steps. Crystals of Amico and R88L variant were obtained by the “hanging drop” vapor diffusion method. The X-ray diffraction data were collected at the ID30A-3 beamline of the ESRF synchrotron (Grenoble, France). The data were indexed, integrated, and scaled using Dials [4]. The structures were analyzed in the PyMol data visualization program.

**Results:** Crystal structure of holoform of Amico was obtained at 1.9 Å resolutions, of R88L variant – at 1.8 Å. The asymmetric unit of enzyme contains two subunits and organization of a dimer is typical of TAs of PLP fold type IV (Fig. 1A). The PLP cofactor is covalently bound to K142 and divides the active site into two pockets: O-pocket (phenolic side) and P-pocket (phosphate side). Noncovalent interactions of PLP with the active site residues in Amico are similar known TAs of fold type IV, except for the E172 residue. Residue E172, which forms a hydrogen bond with the N1 atom of the pyridine ring of PLP, has a different orientation in the subunits: in subunit B it forms hydrogen

bond with N1 atom, however, in subunit A it coordinates N1 atom through the water molecule W131 (Fig. 1B). The O-loop is located outside the O-pocket in the Amico dimer (Fig. 1, C), otherways then in bsDAAT, CpuTA and Halhy. Residue R27\* is directed to the active site and does not form hydrogen bonds with any neighboring residues, while residue R88 forms a network of hydrogen bonds with three residues – Q26\*, Y90 and E113 (Fig. 1, C). In R88L variant the observed changes in the active site residue positions pointed to the disturbance of the active site integrity due to R88L substitution.



**Fig. 1.** The structure of Amico. (A) The active dimer of Amico, (B) binding of the N1 atom of PLP and the position of residue E172 in subunit A, (C) organization of the O-pocket. The PLP molecule is shown in yellow, the O-loop in green, the PLP-binding residues in cyan, the A subunit residues in gray, and the B subunit residues in purple

**Conclusion:** The structure of the Amico holo form and R88L variant were obtained. We found an unusual mobility of the E172 residue that coordinates the N1 atom of the PLP cofactor and a tightly anchored R88 residue in the active site. Thus, the R88 residue may be restricted in mobility during catalysis, unlike similar residues in CpuTA and Halhy, and bear both structural and catalytic function. The removal of the O-loop from the Amico's active site may indicate that K99\* and H101\* residues are not involved in binding of the  $\alpha$ -carboxylic group of substrates.

**Acknowledgements:** The study is supported by the Russian Science Foundation project 19-14-00164.

## References

- Höhne M., Schätzle S., Jochens H., Robins K., Bornscheuer U.T. Rational assignment of key motifs for function guides *in silico* enzyme identification. *Nat Chem Biol.* 2010;6:807-813. doi: 10.1038/nchembio.447.
- Pavkov-Keller T., Strohmeier G.A., Diepold M., Peeters W., Smeets N., Schürmann M. et al. Discovery and structural characterisation of new fold type IV-transaminases exemplify the diversity of this enzyme fold. *Sci Rep.* 2016;6:38183. doi: 10.1038/srep38183.
- Bakunova A.K., Nikolaeva A.Y., Rakitina T.V., Isaikina T.Y., Khrenova M.G., Boyko K.M. et al. The Uncommon Active Site of D-Amino Acid Transaminase from Haliscomenobacter hydrossis: Biochemical and Structural Insights into the New Enzyme. *Molecules.* 2021;26(16):5053. doi: 10.3390/MOLECULES26165053.
- Beilsten-Edmands J., Winter G., Gildea R., Parkhurst J., Waterman D., Evans G. Scaling diffraction data in the DIALS software package: Algorithms and new approaches for multicrystal scaling. *Acta Crystallogr Sect D Struct Biol.* 2020;76:385-399. doi: 10.1107/S2059798320003198.

## ***In silico* engineering towards enhancement of Anti-CD20 nanobody binding affinity**

Shirazi E.A.<sup>1</sup>, Cheraghi S.F.<sup>1</sup>, Heidari M.M.<sup>1</sup>, Khatami M.<sup>1</sup>, Orlov Y.L.<sup>2,3</sup>

<sup>1</sup> Department of Biology, Faculty of Science, Yazd University, Yazd, Iran

<sup>2</sup> Institute of Digital Medicine I.M. Sechenov First Moscow State Medical University, Moscow, Russia

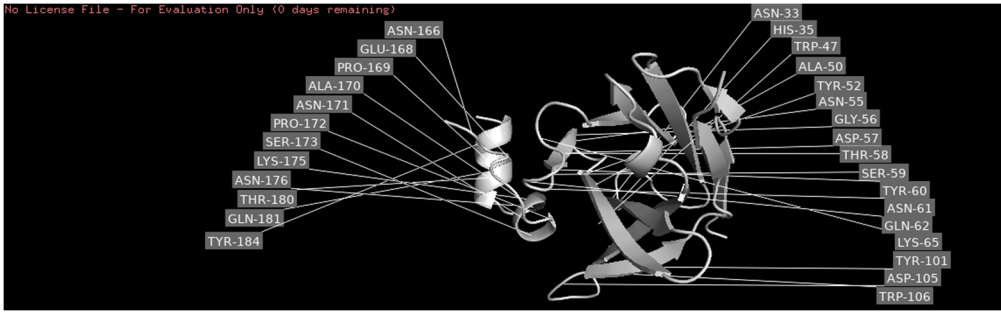
<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

**Key words:** CDR Grafting, CD20, cancer, Nanobody, Site Directed mutagenesis

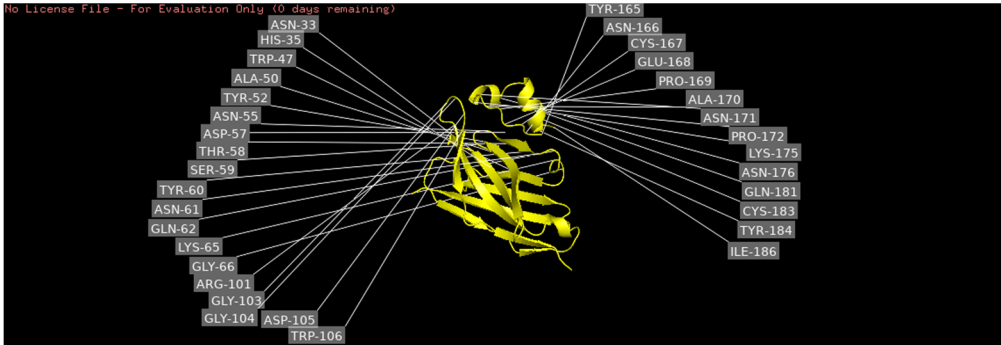
*Motivation and Aim:* CD20 is a membrane non-glycosylated protein (33-37 kDa) that expressed on the surface of normal and malignant B lymphocytes, and encoded by *MS4A1* Gene that is located on 11q12.2 (1). The efficacy of the anti-CD20 monoclonal Antibodies (like rituximab), in treating B-cell disorders such as cancer and autoimmunity diseases has been remarkable. Despite the clinical potential, became major detriments to their efficacy their immunogenicity and large size (~150 kDa). The discovery of heavy-chain only antibodies (HcAbs) in camelids was amazing. HcAbs consist of just two heavy chains, with a single variable domain (VHH) that is as known the antigen-binding region. VHHs or nanobodies (Nbs) could retain full antigen binding potential upon isolation, establishing them as the smallest, naturally-derived antigen binding fragment (2). The unique properties of Nbs are small size (~15kDa), affinity (in the nanomolar or picomolar range), high thermal and chemical stability, fast tissue penetration and rapid blood clearance, Recognition of hidden epitopes and sequences shared with human VH. Different approaches have been described for the enhancement binding affinity of Nbs. One of the ways to increase the efficiency of nanobodies is to use the Site-Directed Mutagenesis method.

*Methods and Algorithms:* in this study is used site-directed mutagenesis coupled with docking simulations for anti-CD20 nanobody to investigate how specific amino acid substitutions impact ligand-protein interaction. We created an anti-CD20 Nanobody with CDR grafting method and its modeling was performed by SWISS MODEL server. Then we introduced a point mutation on 101 residue of the nanobody that Tyrosine was converted to Arginine. In addition, we mutated the Tyrosine (non-polar hydrophilic aa) 101 and 107 to Arginine (polar hydrophilic aa) in the CDR3 region, the most involved region in binding to CD20, using PyMOL software and docked their to the CD20 using the HADDOCK program. PyMOL was also used to evaluate the amino acids involved in the interaction.

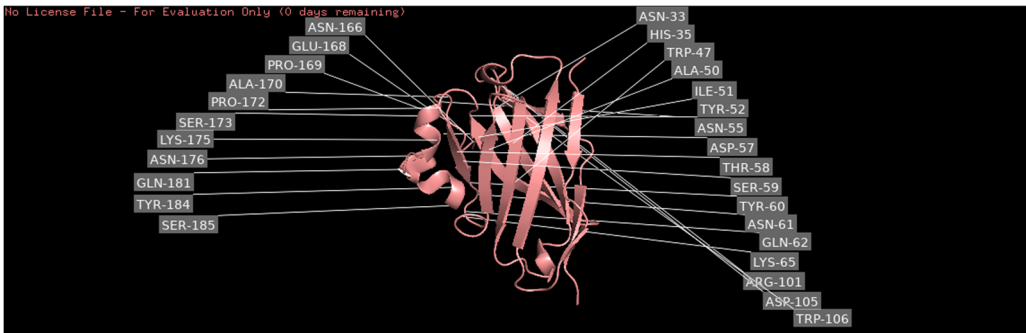
*Results:* By analyzing the structures, connectivity features were improved. The best structure designed in HADDOCK was selected based on the HADDOCK Score and energy level between normal and mutated nanobodies and the electrostatic and van der Waals energies between the CD20 and the nanobodies were calculated. The amount of van der Waals energy decreased (-47.1 (Normal) into -46.1(Y101R) and -38.9 (Y101.107R)) and the number of amino acids involved in the connection (Fig. 1-3) and the amount of electrostatic energy increased (-119.2 (Normal) into -180.3 (Y101R) and -187.3 (Y101.107R)).



**Fig. 1.** Normal nanobody residues interaction



**Fig. 2.** Mutated nanobody (Y101R) residues interaction



**Fig. 3.** Mutated nanobody (Y101.107R) residues interaction

**Conclusion:** Collectively, results obtained suggest that the binding of CD20 to the ligand (nanobody) was improved by converting non-polar hydrophilic amino acids into polar hydrophilic amino acids. Base of our results, this mutations increased the solubility and specificity, therefore, improved the binding affinity. However, further investigation is essential to clarify this binding.

### References

1. Pavlasova G., Mraz M. The regulation and function of CD20: an “enigma” of B-cell biology and targeted therapy. *Haematologica*. 2020;105(6):1494.
2. Yang E.Y., Shah K. Nanobodies: next generation of cancer diagnostics and therapeutics. *Front Oncol*. 2020;10:1182.

# Comparison of nucleosome unfolding by yeast and human FACT: electron microscopy analysis

Sivkina A.<sup>1\*</sup>, Volokh O.<sup>1</sup>, Karlova M.<sup>1</sup>, Sokolova O.S.<sup>1,2</sup>, Studitsky V.<sup>1,3</sup>

<sup>1</sup> *Biology Faculty, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *Department of Biology, Shenzhen MSU-BIT University, Shenzhen, China*

<sup>3</sup> *Fox Chase Cancer Center, Philadelphia, USA*

\* *anastasiia.sivkina@gmail.com*

**Key words:** FACT, Nhp6, SSRP1, SPT16, POB3, nucleosome, curaxins, nucleosome unfolding, transmission electron microscopy

*Motivation and Aim:* FACT (facilitates chromatin transcription) is a histone chaperone that participates in nucleosome removal and reassembly during transcription and replication. Previously we have shown that yeast FACT dramatically alters the nucleosome structure without ATP hydrolysis, but the extent of these alterations depends on the presence of HMGB domain-containing protein Nhp6 [1]. It was shown *in vitro* that hFACT drives nucleosome unfolding in presence DNA-intercalators curaxins; presumably curaxins destabilizes nucleosomes and thus facilitates FACT-nucleosome interaction [2]. Nevertheless, the detailed mechanism of this process is still unclear.

*Methods and Algorithms:* We used mononucleosomes assembled on the 603 Widom positioning sequence. Nucleosomes were assembled by histone octamer transfer to DNA templates after dialysis from 1M NaCl to 0.01M NaCl. SPT16/SSRP1 and SPT16/POB3 were prepared at concentration of 0.05  $\mu$ M. Complexes of FACT with the nucleosome were formed in the presence of 0.1  $\mu$ M FACT, 0.1  $\mu$ M core chicken nucleosomes, 0.5 nM fluorescently labeled core nucleosomes N35/112 and 5  $\mu$ M CBL0137 (for human FACT) or 10  $\mu$ M Nhp6 protein (for yeast FACT). For transmission electron microscopy (TEM) analysis samples were applied to the carbon-coated glow-discharged copper grid (Ted Pella, USA), subjected to glow-discharge using Emitech K100X device (Emitech Ltd., UK), stained for 30 sec with 1 % uranyl acetate, and air dried. Grids were studied in JEOL 2100 TEM (JEOL) microscope operated at 200 kV at low-dose conditions. Micrographs were captured by the Gatan Ultrascan camera with magnification x25,000, no tilt, with 4.1  $\text{\AA}$  pixel size using SerialEM software. Single particle images of FACT, complexes of FACT with the nucleosome and complexes of FACT with the nucleosome formed in the presence of CBL0137 were collected from the micrographs using a neural network provided by EMAN2.3 software. Single particles coordinates collected by the neural network were imported in RELION2.1 software; all further 2D-processing, analysis and CTF-correction were performed using RELION2.1 software. Extracted particles were used for iterative 2D-classification followed by the elimination of bad classes. Linear dimensions of the 2D-classes were measured with ImageJ.

*Results:* Here using TEM we studied human FACT (hFACT) and yeast FACT (yFACT) flexibility alone and in complexes with nucleosomes [3]. All studied complexes are highly flexible and adopt broad ranges of configurations. DNA-binding protein Nhp6 binds to the C-terminal tails of both yFACT subunits and induces formation of more open FACT complexes, thus altering the structures of FACT and the nucleosomes and facilitating nucleosome unfolding. Multiple closed and open conformations were also

demonstrated for nucleosome-free hFACT. The open conformations of FACT become predominant during curaxin-induced nucleosome unfolding involving multiple intermediates. We demonstrated that both yFACT and hFACT flexibility facilitates FACT-dependent nucleosome unfolding that occurs similarly for yFACT and hFACT, resulting in formation of nearly linear, extensively unfolded structure.

*Conclusion:* The data suggest that the process proceeds through a series of energetically similar intermediate structures, ultimately leading to an extensively unfolded form. We proposed FACT-dependent nucleosome unfolding pathway based on a large number of potential intermediates revealed by electron microscopy.

*Acknowledgements:* This work was supported by the Russian Science Foundation (No. 19-74-30003). Electron microscopy was performed on the Unique equipment setup “3D-EMC” of Moscow State University, Department of Biology.

### *References*

1. Valieva M.E. et al. Large-scale ATP-independent nucleosome unfolding by a histone chaperone. *Nat Struct Mol Biol.* 2016;23(12):1111-1116.
2. Chang H.W. et al. Histone Chaperone FACT and Curaxins: Effects on Genome Structure and Function. *J Cancer Metastasis Treat.* 2019;5:78.
3. Sivkina A.L. et al. Electron microscopy analysis of ATP-independent nucleosome unfolding by FACT. *Commun Biol.* 2022;5(1):2.

## Use of lipodiscs in structural studies of ion channels

Sokolova O.S.<sup>1,2\*</sup>, Kravchuk E.<sup>1</sup>, Kacher Yu.<sup>1,3</sup>, Gibadullina S.<sup>1</sup>, Karlova M.G.<sup>1</sup>,  
Glukhov G.S.<sup>1,2</sup>, Bagrov D.V.<sup>1</sup>

<sup>1</sup> *Moscow Lomonosov State University, Moscow, Russia*

<sup>2</sup> *MSU–BIT University, Shenzhen, China*

<sup>3</sup> *Lorraine University, Nancy, France*

\* *sokolova@mail.bio.msu.ru*

**Key words:** SMALP, HERG, KCNQ, structure, purification

*Motivation and Aim:* One of the modern methods for isolating membrane proteins is the use of a styrene-maleic acid (SMA) copolymer. This amphipathic copolymer can integrate into biological membranes and easily destroy them. As a result, discoid membrane fragments with a size of 10–40 nm are formed, surrounded by a copolymer belt [1]. Such particles are known as SMALPs (SMA lipid particles) or lipodiscs. The polymer has no affinity to any specific lipids, and, in SMALPs, the ratio of lipids remains the same as it was in the original membrane [2]. If the membrane contains proteins, they are enclosed into the forming lipodiscs, upon the addition of the copolymer. The ability to extract proteins with SMA has been demonstrated for liposomes [3], cell membrane fractions [4], and whole cells [5]. The SMA-extracted membrane proteins are quite stable and can be purified and further analyzed by various biochemical methods [4, 5]. The benefit of SMA-extraction is the possibility of completely avoiding detergents in the protein purification process. It means better preservation of the native conformation and lipid microenvironment of the proteins. Here, we used SMA purification for isolate the ion channels of interest expressed in COS-1 cells, for structural studies.

*Methods:* The COS-1 and CHO cell lines were maintained in Dulbecco's modified Eagle's medium (PanEco, Russia), supplemented with 10 % of fetal bovine serum (HyClone, USA). COS-7 cells (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (Invitrogen), supplemented with 10 % fetal calf serum (Eurobio) and antibiotics (100 IU/ml penicillin and 100 µg/ml streptomycin; Gibco). All cell lines were cultured at 5 % CO<sub>2</sub> and 37 °C in a humidified incubator. Cells were transiently transfected with plasmids pIRES2-EGFP/hKCNQ1-1D4, pcDNA6-V5-HisA/hKCNE1-hKCNQ1 and pMT3-hKCNH5 using the Metafectene PRO (Biontex, Germany) for purification purposes and the Fugene 6 Transfection Reagent (Promega) for the electrophysiological experiments. In transfected mammalian cells, currents were recorded using the whole-cell configuration of the patch-clamp technique. For preparation of protein-containing SMALPs, COS-1 cells expressing ion channel proteins were resuspended in the buffer A (10mM Tris-HCl, 150mM NaCl, 2mM DTT, 1mM EDTA, protease inhibitor cocktail, pH 8) in the presence of a 2.5 % (w/v) SMA copolymer, incubated for 30 min at 4 °C with shaking, sonicated with an ultrasonic sonicator (Branson Ultrasonic Corporation, USA) for 15 s on ice and incubated for an additional 30 min at 4 °C. Suspensions were centrifuged for 15 min at 200,000 g. The pellet and supernatant were analysed by SDS-PAGE and immunoblotting. Supernatants were subsequently purified on affinity resin. Dynamic light scattering experiments were performed on a Brookhaven 90 Plus instrument

(Brookhaven Instruments Company, USA), in a thermostated cell at 20 °C. The buffer solution was filtered through 0.22 µm membrane filters. The scattered light was recorded at an angle of  $\theta = 90^\circ$ , the accumulation time of the signal was 1 min.

**Results:** We overexpress the Kv11.1 and Kv7.1 ion channels on CHO-1 cells and compared the effectiveness of their solubilization by SMA and detergent. We demonstrated that the SMA copolymer was more efficient at solubilization of the human KCNQ1 channels than CHAPS. The advantage of using SMALP is that the solubilized membrane proteins can be easily concentrated on Microcon concentrators without aggregation. A DLS experiment demonstrated that nanodiscs have the overall size of 15 nm, while electron microscopy revealed a four-fold symmetry within channel-containing SMALPs. Using mass-spectroscopy (LC-MS) and Raman spectroscopy to analyze the presence of lipids in the lipodisc stabilizing the recombinant human KCNQ1 protein. Raman spectroscopy detected several spectral peaks, which could be attributed to lipids which was confirmed by mass-spectroscopy.

**Conclusions:** using SMA copolymers, we tested a method of detergent-free solubilization of human ion channels, particularly, the cardiac and neuronal potassium voltage-dependent channels. SMALPs appear to develop into a convenient platform for studying the structure of human ion channels and their complexes (which are hard to crystallize) using not only cryo-EM, but also NMR methods, as well as other structural methods that require using the single particle mode (including XFEL). The study of the structural and functional properties of voltage-dependent potassium channels would help to clarify the mechanisms that cause malfunction of these channels in case of point mutations. Understanding these mechanisms, in its turn, would pave the way to methods of targeted correction of channel function.

**Acknowledgements:** The study is supported by RSF (22-14-00088). Yulia Kacher is a Vernadsky PhD fellowship scholar.

## References

1. Tonge S.R., Tighe B.J. Responsive hydrophobically associating polymers: a review of structure and properties. *Adv Drug Deliv Rev.* 2001;53(1):109-122. doi: 10.1016/s0169-409x(01)00223-x.
2. Dominguez Pardo J.J., Dörr J.M., Iyer A. et al. Solubilization of lipids and lipid phases by the styrene-maleic acid copolymer. *Eur Biophys J.* 2017;46(1):91-101. doi: 10.1007/s00249-016-1181-7.
3. Knowles T.J., Finka R., Smith C., Lin Y.P., Dafforn T., Overduin M. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *J Am Chem Soc.* 2009;131(22):7484-7485. doi: 10.1021/ja810046q.
4. Gulati S., Jamshad M., Knowles T.J. et al. Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochem J.* 2014;461(2):269-278. doi: 10.1042/BJ20131477.
5. Karlova M.G., Voskoboynikova N., Gluhov G.S. et al. Detergent-free solubilization of human Kv channels expressed in mammalian cells. *Chem Phys Lipids.* 2019;219:50-57. doi: 10.1016/j.chemphyslip.2019.01.013.



## Molecular modeling of osmolytes' effect on bacterial luciferase

Sukovatyi L.A.<sup>1\*</sup>, Lisitsa A.E.<sup>1</sup>, Deeva A.A.<sup>1</sup>, Kratasyuk V.A.<sup>1,2</sup>, Nemtseva E.V.<sup>1,2</sup>

<sup>1</sup> Siberian Federal University, Krasnoyarsk, Russia

<sup>2</sup> Institute of Biophysics, SB RAS, Krasnoyarsk, Russia

\* [lsukovatyi@sfu-kras.ru](mailto:lsukovatyi@sfu-kras.ru)

**Key words:** bioluminescence, luminous bacteria, bacterial luciferase, osmolyte, molecular dynamics

*Motivation and Aim:* Intracellular accumulation of organic solutes is one of the strategies used by living organisms to manage the variety of extreme conditions that they encounter in nature. This is also the case for luminous bacteria that inhabit the ocean and can adapt to the variety of temperatures, salinity, concentration of nutrients and toxins. The cells accumulate small molecules to counter the osmotic stress. These low molecular weight protective compounds called osmolytes include polyols, sugars, amino acid derivatives etc. [1]. They are assumed to increase protein stability and prevent protein aggregation. Nevertheless, small organic compounds can cause structural and dynamic changes of proteins and so influence its function, i.e. catalytic activity. For *Photobacterium leiognathi* luciferase it was revealed that integral intensity of the reaction changes in the presence of different cosolvents as follows: sucrose > sorbitol > glucose > glycerol > ethylene glycol that correlates with the relation of hydrodynamic radii of the cosolvent molecules: 5.2 > 3.9 > 3.6 > 3.1 > 2.6 (Å) [2]. Moreover, sucrose increases catalytic constant of excited state intermediate formation up to 5-fold unlike other cosolvents. Atomic-level molecular dynamics (MD) is useful tool to examine effects of osmolytes on *P. leiognathi* structure. Thus, the aim of this study was to reveal the structural and dynamic effects caused by osmolytes on the bacterial luciferase using molecular modelling approach.

*Methods and Algorithms:* Three-dimension structure of luciferase from *P. leiognathi* was obtained by homology modelling using Swiss-model server. Crystal structure of luciferase from *Vibrio harveyi* (PDB ID: 3FGC) was used as a template. MD simulation of structure of *P. leiognathi* luciferase in water and in water-cosolvent mixtures (40 w/w %) was performed with GROMACS 2020.4 software. Sucrose, sorbitol, glucose, glycerol and ethylene glycol molecules were used as cosolvents. MD of each modeled system was calculated for 100 ns with three independent runs.

*Results:* To estimate the penetration of cosolvents into active site gorge of *P. leiognathi* luciferase a radial distribution function (RDF) was calculated between centre of mass of  $\alpha$ GLU43 and cosolvent molecules.  $\alpha$ GLU43 was chosen as reference point because it locates in the deep of luciferase active site. RDF demonstrates that ethylene glycol and glycerol molecules penetrate deeper the active site of luciferase than other cosolvents that may explain the decrease of integral intensity in the presence of these two osmolytes. Minimum distance of RDF for ethylene glycol is 1,5 Å and for glycerol – 1,6 Å, which is close to the value of water (1,45 Å). Sorbitol molecules can be found at 1,7 Å, but this case is extremely rare and the RDF remains close to zero up to the distance of 8 Å relative to the reference point. For glucose and sucrose the RDF function starts to rise from the distance of 4,3 and 5,5 Å respectively, and reaches significant values at >10 Å, i.e. closer to the entrance of the active site gorge.

Calculated gyration radius and solvent accessible surface area of the luciferase in presence of cosolvents indicate that ethylene glycol changes the structure of the protein. These both parameters were found to be significantly higher for this cosolvent. This result is in good agreement with the experimental data demonstrating that only ethylene glycol causes a decrease of the integral intensity of bioluminescent reaction of *P. leiognathi* luciferase.

Analysis of distribution of side chain rotamers of amino acids was carry out to study local conformational transitions in active site of the luciferase. It was found that in presence of sucrose the  $\alpha$ Glu175 side chain takes on the only conformation (mt-10, according to the S. Lovell rotamer library), which necessary for binding of luciferase with flavin [3]. However, in aqueous solution and in presence of the other cosolvents it could take additionally on second (known from crystal structure [4]) conformation, which makes impossible binding FMN in the active center. It can be assumed that the presence of sucrose causes local conformation changes in an active site with improved catalytic properties. This fact is in a good agreement with the experimentally observed increase of catalytic constant of the luciferase in presence of sucrose [2].

*Conclusion:* Thus, the effects of osmolytes on the bacterial bioluminescent reaction were explained by: (i) change in local dynamics of functionally significant regions of luciferase, (ii) global structural alteration of the protein in the presence of ethylene glycol, (iii) different accessibility of the luciferase active center to osmolytes.

*Acknowledgements:* The study is supported by the Russian Foundation for Basic Research as part of a research project No. 20-34-90118 and the Russian Foundation for Basic Research and the administration of the Krasnoyarsk Territory and Krasnoyarsk Regional Fund of Science within the framework of the scientific project No. 20-44-243002.

### References

1. Dandapath I. et al. Bacterial osmolyte system and its physiological roles. In: Cellular Osmolytes. Singapore: Springer, 2017;229-249.
2. Lisitsa A.E. et al. Mechanisms of viscous media effects on elementary steps of bacterial bioluminescent reaction. *Int J Mol Sci.* 2021;22(16):8827-8846.
3. Lovell S.C. et al. The penultimate rotamer library. *Proteins.* 2000;40(3);389-408.
4. Sparks J.M., Thomas O.B. Functional implications of the unstructured loop in the ( $\beta/\alpha$ ) 8 barrel structure of the bacterial luciferase  $\alpha$  subunit. *Biochemistry.* 2001;40(50);15436-15443.

## Analysis of binding properties of influenza hemagglutinins and human receptor analog

Unguryan V.<sup>1,2\*</sup>, Bakulina A.<sup>1</sup>, Danilenko A.<sup>1</sup>, Kolosova N.<sup>1</sup>, Ryzhikov A.<sup>1</sup>

<sup>1</sup>State Research Center of Virology and Biotechnology “Vector”, Novosibirsk, Russia

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

\* v.unguryan@g.nsu.ru

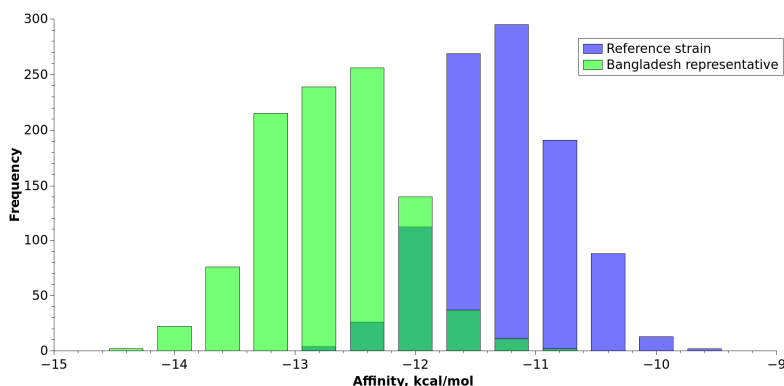
**Key words:** hemagglutinin, molecular docking, molecular dynamics, homology modelling, binding

*Motivation and Aim:* The surface protein of the influenza virus, hemagglutinin, plays a key role in the virus entry into the cell. When infecting a cell, the virus first binds to a receptor on the cell membrane surface, followed by the capture of the virus inside the cell. The influenza virus infectivity is defined by the duration of binding persistence, which is related to energy characteristics of the bonds.

The new group of A(H3N2) influenza virus serotype, the Bangladesh group, has spread in recent years in Russia, replacing other groups of this serotype. The aim of this work was to compare the binding properties of Bangladesh group representatives with the international reference strain of A(H3N2) subtype in the complex with the human receptor analog 6'-sialyl-N-acetylactosamine.

*Methods and Algorithms:* The initial structures of hemagglutinins under study were modelled with MODELLER with the template of the known structure H3N2 from the Protein Data Bank (sequence identity 95 %, code PDB 6bkt). The molecular dynamics simulations were performed using NAMD during 4 nanoseconds to determine more precisely the orientation of the model side chains and receptor position. The structures of the time interval from 3 to 4 nanoseconds with the time step of 1 picosecond were sampled for the subsequent docking analysis using the AutoDock Vina program.

*Results:* During the molecular dynamics simulations, the receptor analog was in the correct stable binding conformation. However, the root-mean-square deviation of heavy atoms was greater for the international reference strain. Docking analysis allowed a numerical estimation of the free energy of protein binding to the receptor analog. A histogram was built for the data obtained (Fig. 1).



**Fig. 1.** Distributions of AutoDock Vina energy for the strains studied

For a protein and the receptor analog prepared as rigid structures, the free energy estimate is 11.27 kcal/mol for the reference system and 12.7 kcal/mol for the Bangladesh group representative, with the standard deviation being 0.5 and 0.56 kcal/mol respectively.

*Conclusion:* The results demonstrate Bangladesh group influenza virus to have a higher affinity for the human receptor analog. The T-statistic being equal to 60 confirms the statistical validity of the results, suggesting the possibility of higher virulence of Bangladesh group viruses compared to other groups. The energy assessment and structural reasons for the higher affinity are likely to be important for the development of drugs against this group of the influenza virus.

# Identification of factors important for the translocation of water-soluble proteins across the membrane using computer simulations

Volynsky P.<sup>1\*</sup>, Maltseva D.<sup>2</sup>, Tabakmakher V.<sup>1,3</sup>, Bocharov E.<sup>1,4</sup>, Britikova E.<sup>5</sup>, Tonevitsky A.<sup>1,2</sup>, Efremov R.<sup>1,2,4</sup>

<sup>1</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>2</sup> National Research University Higher School of Economics, Moscow, Russia

<sup>3</sup> School of Biomedicine, Far Eastern Federal University, FEFU Campus, Moscow, Russia

<sup>4</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Russia

<sup>5</sup> Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus, Minsk, Belarus

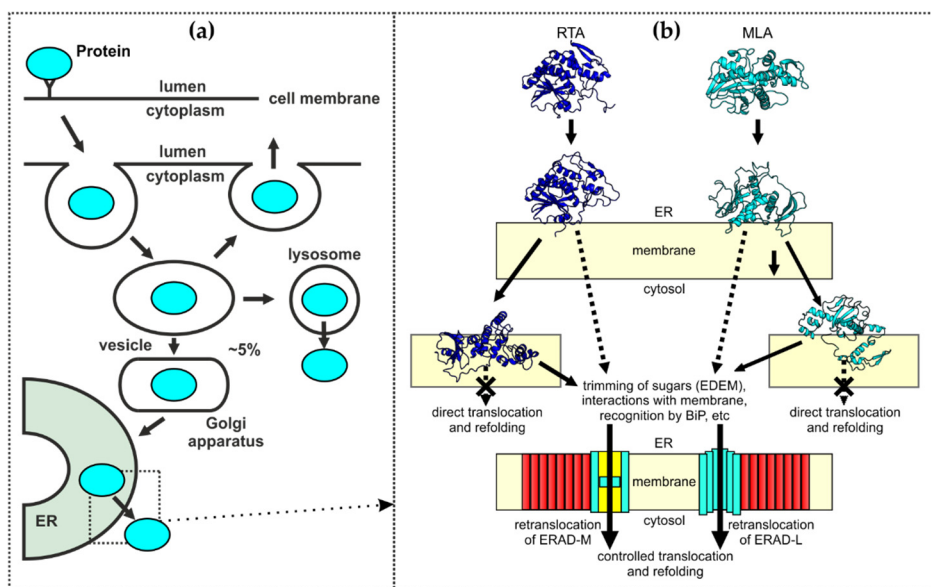
\* volynski@yandex.ru

**Key words:** membrane trafficking of proteins, medium effects on protein structure, protein-membrane interactions

*Motivation and Aim:* Some water-soluble proteins are able to penetrate the cell membrane. The principal pathways of their passage through the membrane are currently well known (Fig. 1, a). The first step is binding to the cell membrane with following endocytosis. Then part of the endosomes binds to the cell membrane and releases its contents into the lumen. The second part is transformed into lysosomes, where the proteins undergo proteolysis. In this pathway, some proteins cross the membrane by direct translocation after decreasing of the pH of environment. The third part of the endosomes binds to the Golgi apparatus and further to the endoplasmic reticulum, where proteins can cross the membrane using the endoplasmic-reticulum-associated protein degradation ERAD system. In this case, the similarity of the protein to the misfolded one is of great importance, for example, the presence of large non-polar regions on its surface or high internal mobility. One of the promising approaches to reveal these properties is molecular modeling. Molecular dynamics (MD) in homogeneous solvents of different polarity in the all-atom representation makes it possible to study the internal mobility of proteins. MD in the coarse-grained representation of the protein/membrane/water systems makes it possible to evaluate the efficiency of interaction with the membrane and its changes with decreasing pH. Identification of the key factors important for translocation across the membrane was carried out on a set of homologous ribosome-inactivating proteins viscumin (MLA), ricin (RTA), and trichobakin (TBK), the efficiency of membrane penetration of which was measured experimentally.

*Methods and Algorithms:* To study the internal mobility of proteins we used 5–10 mks MD in amber99sb.ildn force field using gromacs software. To study the role of the polarity of solvents we used water, 8M urea (high polarity) and 1/1 mixture of chloroform-methanol. To analyze the interactions of proteins with membrane we used coarse grained MD simulations with martini force field. The lipid bilayer was composed of 450 POPC, POPS and POPE molecules in a 1/1/1 ratio to mimic the composition of the ER membrane. The interactions of each protein with the membrane were studied in series of 6 calculations of 5  $\mu$ s each with different initial orientations of the proteins with respect to the membrane. In the simulation, the parameters recommended for the applied

force fields and system compositions were used. The analysis used the tools of the gromacs package and self-made programs.



**Fig. 1.** Potential pathways of proteins translocation from the lumen into the cytosol. (a) Network for the delivery of proteins from the lumen to the cytoplasm. (b) RTA and MLA in the conformations adapted to water, urea/water, and chloroform/methanol mixtures in the bulk water, in the strongly polar regions on the membrane surface and in contact with the membrane interface/hydrophobic core respectively. Schematic presentation of more hydrophilic ERAD-L and lipophilic ERAD-M retrotranslocation systems. The solid and dashed arrows show the most possible and unfavorable pathways of the toxin translocation respectively

*Results:* Analysis of the internal mobility of proteins showed that all proteins are relatively stable in water and urea. In a methanol-chloroform mixture, that mimics a membrane medium, all proteins lose their tertiary structure, retaining their secondary structure. The rate of structural transformations increased in the series TBK, MLA, RTA. Analysis of interactions of proteins with membranes have shown that effectiveness of interactions increased in the series RTA, MLA and TBK. After decreasing the pH (changing of the protonation state of several residues), the interaction with the membrane increased, especially in the case of TBK.

*Conclusion:* Computer simulations permits to delineate the factors important for translocation of proteins through the cell membrane. Namely: (1) The high internal mobility of proteins in a membrane-mimicking environment facilitates the translocation of proteins through the ER-associated pathway. (2) In the case of MLA, the decrease in the rate of internal rearrangements is partially compensated by an increase in the efficiency of interaction with membranes. (3) TBK possibly crosses the membrane along a direct translocation pathway.

*Acknowledgements:* The study was supported by the Russian Foundation for Basic Research (projects No. 20-04-00697 and No. 20-54-81015).

## Small-angle X-ray scattering study of histone-like protein from *Spiroplasma melliferum* in solution

Volynsky P.<sup>1\*</sup>, Gaponov Y.<sup>2</sup>, Altukhov D.<sup>2</sup>, Timofeeva V.<sup>2,3</sup>, Agapova Y.<sup>2</sup>,  
Bocharov E.<sup>1</sup>, Shtykova E.<sup>3</sup>, Rakitina T.<sup>1,2</sup>

<sup>1</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

<sup>2</sup> *National Research Center “Kurchatov Institute”, Moscow, Russia*

<sup>3</sup> *A.V. Shubnikov Institute of Crystallography, FSRC ‘Crystallography and Photonics’, RAS, Moscow, Russia*

\* [volynski@yandex.ru](mailto:volynski@yandex.ru)

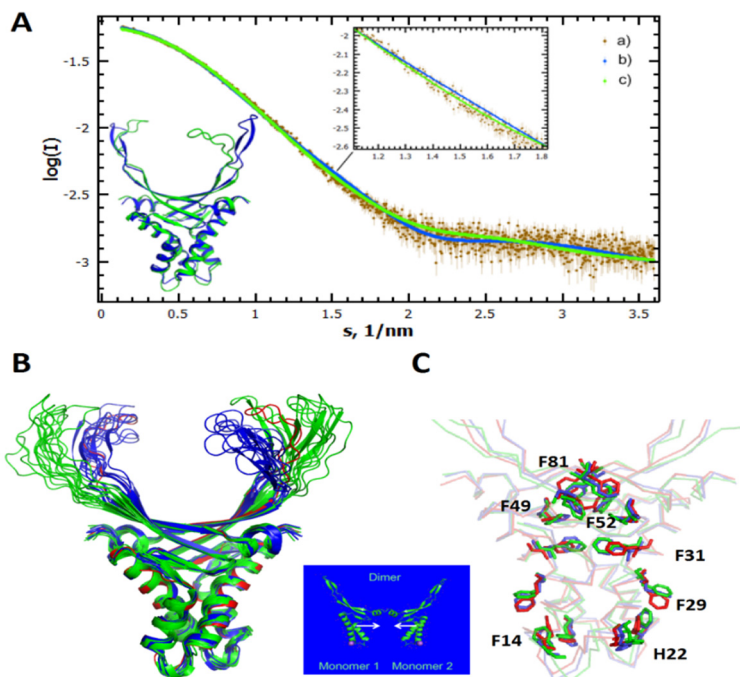
**Key words:** crystal structure, solution structure, histone-like protein, HU protein

**Motivation and Aim:** Histone-like proteins (HU) are the most abundant and ubiquitous nucleoid associated proteins in bacteria. These small dimeric proteins bind, bend, and wrap genomic DNA playing important roles in regulation of the major DNA-dependent transactions. Despite the well-known involvement of HU in the regulation of viability and virulence of pathogenic bacteria, only few chemical inhibitors of HU with antibacterial effects were reported until now. One of the numerous putative reasons preventing the successful development of such inhibitors is the inconsistency between the crystal structures of HU used for structure-based drug design and the shape, size and folding of the protein and in a solution.

**Methods and Algorithms:** In this study, a solution of HU from *Spiroplasma melliferum* (HUSpm) was studied by small-angle X-ray scattering (SAXS). The experimental SAXS curves were compared with theoretical curves calculated for both the HUSpm crystal structure (PDB ID 5L8Z) and dynamic models of solution structure obtained by the combination of the NMR spectroscopy and molecular dynamics (PDB ID 5OGU). The values of following parameters were analyzed: *hi2* -reduced chi-squared,  $R_g$  – Guinier radii,  $D_{max}$  – maximum distances in the molecule calculated using coordinates of the atoms,  $D_{mon}$  – distance between mass centers of the monomers in the dimer,  $MW$  – molecule weight.

**Results:** In all cases, both  $R_g$  and  $MW$  confirmed dimeric conformation of HUSpm. The 15 models from the NMR structure 5OGU were grouped into 2 clusters. All structures from the cluster-1 have smaller *hi2* (better agreement with experimental SAXS curve) and  $D_{mon}$  than those from cluster-2.  $R_g$  and  $D_{max}$  of the cluster-1 structures were also smaller. One can conclude that shorter  $D_{mon}$  corresponds smaller values of  $R_g$  and  $D_{max}$ . The best fitting to the experimental SAXS curve was obtained for structure 2 from cluster-1 with *hi2* = 1.28, while the crystalline structure fitted to the experimental curve with *hi2* = 1.36 (Fig. 1A). Values of  $R_g$ ,  $D_{max}$ ,  $MW$  of the best theoretical curve corresponds to the parameters of the experimental SAXS curve within the experimental errors. Since, the best structure from cluster-1 has smaller  $D_{mon}$  than that of the crystal structure, one can conclude that in solution monomers of the HUSpm dimer probably are closer to each other compare to the crystalline state. Analysis of the interface between the monomers in the HUSpm dimer revealed two aromatic clusters of stacking interactions, which, as NMR-data analysis can suggest, are responsible for structural

plasticity of the molecule and can provide a possibility of mutual sliding of the monomers toward each other (Fig. 1, B and C) [1].



**Fig. 1.** Variations in the HUSpm folding according to X-Ray, NMR and SAXS studies. A – An experimental SAXS curve (a) and its fitting with calculated SAXS curves for the 5L8Z crystal structure (b) and the best model from the 5OGU NMR structure (c). Superposition of the corresponding structures is shown at the bottom. B – Superposition of two clusters of the 5OGU structure (green and blue) and the 5L8Z crystal structure (red). C – Two aromatic centers of staking interactions in the interface between monomers of HUSpm. The color-coding is similar to B

*Conclusion:* Upon comparative structure analysis of the HUSpm dimer in solution and crystalline state and the analysis of calculated and experimental SAXS curves, we conclude that the distance between mass centers of the HUSpm monomers in solution is shorter compare to the distance in crystalline state; the difference of distances is about 0.1 nm.

*Acknowledgements:* This work was supported by RFBR (project No. 20-04-01001).

### References

1. Timofeev V.I. et al. Structural plasticity and thermal stability of the histone-like protein from *Spiroplasma melliferum* are due to phenylalanine insertions into the conservative scaffold. *J Biomol Struct Dyn.* 2018;36(16):4392-4404. doi: 10.1080/07391102.2017.1417162.



## The study of localization of beta-amyloid peptide monomers in cells by fluorescence microscopy

Zagryadskaya Yu.<sup>1\*</sup>, Burkatovskii D.<sup>1</sup>, Kamynina A.<sup>1,2</sup>, Dencher N.A.<sup>1</sup>, Borshchevskiy V.<sup>1</sup>, Okhrimenko I.<sup>1\*</sup>

<sup>1</sup> *Research Center for Molecular Mechanisms of Aging and Age-Related Diseases of Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*

<sup>2</sup> *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

\* *1989july@mail.ru; okhrimenko.is@mipt.ru*

**Key words:** Alzheimer's disease, beta-amyloid, fluorescence confocal microscopy, endocytosis

*Motivation and Aim:* Alzheimer's disease (AD) is the most common form of dementia in the elderly (60–80 % of all known dementia) and one of the socially significant age-related diseases. Currently, this neurodegenerative disease affects more than 50 million people worldwide and is the fifth most common cause of death for people over 65 in developed countries. This number is expected to triple by 2050 unless significant progress is made in research for early diagnosis and new effective therapies [1]. In Russia, the number of people who sufferer from Alzheimer's disease was estimated at more than 1 million people, the majority of patients are women (80 %), about 85 % of patients were aged 70–89 years. Previously, a significant increase in cases of Alzheimer's disease was predicted to 1.3 million people by 2020 [2] only due to an increase in the average age of Russian citizens. However, complications caused by viral diseases and stress will definitely lead to an additional increase in cases of neurodegeneration, since modern studies show an unambiguous effect of viral diseases on the risk of neurodegeneration due to the involvement of common molecular immunological mechanisms [3, 4]. At the same time, at the moment, the proportion of official mortality of Russian citizens from neurodegenerative (mainly Alzheimer's) diseases is significantly less than in other countries. Russian researchers explain this by the presence of a significant number of not diagnosed cases, due to the clear age dependence of neurodegeneration and the similarity of the age-and-sex pyramids of Russia and other countries. Researchers also pay attention to the lack of modern domestic approaches to diagnosis [5]. It is important that the indicators of the lost years of life of citizens at an economically active age are socially and economically significant [6]. Patients initially require general assistance at home, then there is a need for qualified assistance, nursing care and nurse services, as well as the placement of the patient in a hospital. This involves a significant amount of resources of citizens and the state. Improving the diagnosis of Alzheimer's disease at an early stage, the emergence of new diagnostic approaches, the use of new cost-effective drugs will reduce the social and economical burden imposed on relatives of patients and the healthcare system [5].

Despite the enormous progress in science and medicine, there are no effective methods of therapy and early diagnosis – more than 100 years after the disease was first described as a form of progressive, neurodegenerative and behavioral disorder by Alois Alzheimer in Germany [7]. From this fact, it can be concluded that the lack of a detailed description of the molecular and cellular mechanisms involved in the pathogenesis of Alzheimer's disease, and neurodegeneration in general, does not allow the development of effective

pharmacological drugs and early diagnosis. Increasing evidence indicates an important role that A $\beta$  peptides play in AD progression [8]. Despite extensive research in this area, there exists controversial uncertainty in aspects of A $\beta$  toxicity, mechanisms of A $\beta$  internalization and interaction with cell lipid membranes [9]. At the same time, it is worth paying attention to the presence of a small number of perfect new promising modern developments at the highest level of development of physico-chemical, including structural, methods [10–13]. Experimental information obtained in physiologically relevant conditions is required for understanding molecular mechanisms in detail, create therapeutic potential, and, further, for the rational development of new biologically active compounds and screening of existing ones to obtain the most effective candidates for medicines, as well as the development of methods of early diagnosis.

*Results:* The modern development of optical fluorescence microscopy methods allow us to use of the modern a state of the art to understand the molecular mechanisms of intracellular transport and the localization of proteins and peptides in simulated pathophysiological processes occurring in the human body.

*Conclusion:* The time-dependent interaction of physiologically justified low concentration A $\beta$  with human neuroblastoma SH-SY5Y was studied. By means of confocal microscopy we showed that externally applied fluorescently labeled A $\beta$  primarily localized to cellular membrane, this is followed by the internalization of A $\beta$ , depending on the cells differentiation state. The colocalization with acidic vesicles such as endosomes and lysosomes produces measurable toxic effects on mitochondria.

*Acknowledgements:* The study was supported by the Russian Foundation of Basic Research (project No. 20-015-00526).

## References

1. Alzheimer's Disease International. World Alzheimer Report 2019: Attitudes to dementia. 2019. London: Alzheimer's Disease International. <https://www.alz.co.uk/research/world-report-2019>.
2. Belousov Yu.B. et al. Clinical and economic aspects of Alzheimer's disease therapy in Russia. *Good Clinical Practice*. 2009;S1:3-28.
3. Verkhatsky A. et al. Can COVID-19 pandemic boost the epidemic of neurodegenerative diseases? *Biology Direct*. 2020;15(1):28.
4. McAlpine L.S. et al. Coronavirus disease 2019 and neurodegenerative disease: what will the future bring? *Curr Opin Psychiatry*. 2021;34(2):177-185.
5. Vatolina M.A. et al. Lost years of life and mortality due to Alzheimer's disease in Russia. *Psychiatry (Moscow)*. 2014;1:49-53.
6. Koberskaya N., Kobierskij B. Russian and UK Dementia Action Plans: a comparative analysis. *Int J Hum Rights Healthc*. 2022. doi: 10.1108/IJHRH-09-2021-0166.
7. Alzheimer A. Über eine eigenartige Erkrankung der Hirnrinde. *Allg Zeitschr f Psych*. 1907;64:146-148. [et *Zentralbl f Nerven u Psych*. 1907;18:177-179].
8. Urban A.S. et al. Structural studies providing insights into production and conformational behavior of Amyloid- $\beta$  peptide associated with Alzheimer's disease development. *Molecules*. 2021;26(10):2897.
9. Bogorodskiy A. et al. Role of mitochondrial protein import in age-related neurodegenerative and cardiovascular diseases. *Cells*. 2021;10(12):3528.
10. Kozin S.A. et al. Anti-amyloid therapy of Alzheimer's disease: current state and prospects. *Biochemistry (Mosc.)*. 2018;83(9):1057-1067.
11. Kutzsche J. et al. Safety and pharmacokinetics of the orally available antiprionic compound PRI-002: A single and multiple ascending dose phase I study. *Alzheimers Dement*. 2020;6(1):e12001.
12. Bocharov E.V. et al. All-d-Enantiomeric peptide D3 designed for Alzheimer's disease treatment dynamically interacts with membrane-bound Amyloid- $\beta$  precursors. *J Med Chem*. 2021;64(22):16464-16479.
13. Zheng Q. et al. USP25 inhibition ameliorates Alzheimer's pathology through the regulation of APP processing and A $\beta$  generation. *J Clin Invest*. 2022;132(5):e152170.

# The insight into structural properties of unique tardigrade radioprotective damage suppressor protein (Dsup)

Zarubin M.<sup>1\*</sup>, Murugova T.<sup>2</sup>, Ivankov O.<sup>2,3</sup>, Kravchenko E.<sup>1</sup>

<sup>1</sup> Dzhelapov Laboratory of Nuclear Problems, Joint Institute for Nuclear Research, Dubna, Russia

<sup>2</sup> Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, Dubna, Russia

<sup>3</sup> Institute for Safety Problems of Nuclear Power Plants, NAS of Ukraine, Kyiv, Ukraine

\* mzarubin@jinr.ru

**Key words:** tardigrade, Dsup, radioprotection, IDP, SAXS

*Motivation and Aim:* Extraordinary stress-tolerance of the most radioresistant animal *Ramazzottius varieornatus* being promoted by the spread of specific families of tardigrade intrinsically disordered proteins (IDPs) as CAHS, MAHS, SAHS and Dsup. Active investigation of Dsup protein (UniProt P0DOW4) had already revealed its direct DNA protecting role, moreover this effect was transferred to transgenic model human cells HEK293 [1, 2]. At this point, the mechanism of stress tolerance remains uncertain, but currently is interpreted through formation of macromolecular aggregate of DNA with highly charged and putatively intrinsically disordered Dsup protein (445-residues, 42.8 kDa), which leads to the promotion of defense from reactive oxygen species (ROS). The goal of our work is an *in vitro* investigation of Dsup protein structure, that improves both an understanding of Dsup-involving mechanism of radioprotection and its applications.

*Methods:* Dsup protein was expressed in *E. coli* and purified by the size exclusion chromatography (SEC). Structural features and stability of Dsup, its self-assembling and complex formation were studied under native, chemically denatured and elevated temperature conditions by small angle X-ray (SAXS) and neutron (SANS) scattering techniques at FLNP JINR, MIPT and ESRF Grenoble. Secondary structure of Dsup protein and its dynamical properties were determined by combination of circular dichroism (CD) spectroscopy, SAXS and computational approaches.

*Results:* In PBS solutions Dsup protein in 1.4–2.5 mg ml<sup>-1</sup> concentrations is a flexible rod-like monomer ( $R_{flex} \sim 0.85$ ) of ~10 nm with presence of secondary structure in the middle region. Notably, the formation of globular and rod-like fiber oligomers of 4–6 molecules under native conditions was observed and we outline the structurization of oligomers is specific for Dsup protein. Additionally, chemical denaturation (6 M urea) reduced this process and the form of native oligomers differed from temperature-induced aggregates of Dsup protein (20–60 °C).

*Conclusion:* Obtained data evidences intrinsically disordered properties of Dsup protein and approves the presence of secondary structure, also we report the formation of higher-order globular and rod-like oligomers through self-assembling. We expect, this topic is of special interest for neurodegeneration studies [3] as gives a special outlook into aggregation and fibrillization processes of DNA-binding IDPs.

## References

1. Hashimoto T. et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat Commun.* 2016;7:12808. doi: 10.1038/ncomms12808.
2. Chavez C. et al. The tardigrade damage suppressor protein binds to nucleosomes and protects dna from hydroxyl radicals. *Elife.* 2019;8:e47682. doi: 10.7554/eLife.47682.
3. Uversky V.N. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front Aging Neurosci.* 2015;7:18. doi: 10.3389/fnagi.2015.00018.

### 3.3 Section “Chemoinformatics and chemical biology”



# Identification of virus-host interaction mechanisms based on text-mining approach

Biziukova N. \*, Ivanov S., Filimonov D., Tarasova O., Poroikov V.

*Institute of Biomedical Chemistry, Moscow, Russia*

\* [nad.smol@gmail.com](mailto:nad.smol@gmail.com)

**Key words:** viral diseases, virus-host interactions, molecular mechanisms, text-mining

*Motivation and Aim:* Infections cause significant proportion of deaths worldwide [1]. The COVID-19 pandemic has clearly shown the urgent need to turn efforts towards the development of effective antiviral drugs able to affect human proteins that interact with viruses directly or indirectly. Due to the high mutability of viruses, the development of antiviral drugs is a great challenge. Therefore, it is necessary to have a deep understanding of the molecular mechanisms of interaction between the virus and the host. In our study, we focused on development of a workflow that is intended to retrieve the latest information on experimentally proven interactions between the most common viruses and human leading to immune system reaction and identification of key targets involved in the mechanisms of viral infections affecting immune system, including SARS-CoV-2, herpesviruses, hepatitis B, C, influenza, and HIV [2–4].

*Methods and Algorithms:* The texts of scientific publications were automatically extracted from NCBI PubMed using Python 3.10 scripts and Bio library Entrez module. Protein names recognition was performed using our web-applications previously developed [5, 6]. Fig. 1 represents the interface of the tool.

Home About Sign-up Login

Please log into your account or create a new one

### Chemical and protein named entity recognition

1. Upload the set of texts

Enter publications identifiers

Выберите файл Файл не выбран

2. Select the type or types of named entities to be recognized

Chemical named entities

Protein and genes named entities

3. Confirm the correctness of filling in the fields above

Confirm

After you press the "Confirm" button, the algorithm will proceed to the recognition of the selected named entities in the uploaded texts. This may take some time. The result will appear below in the current window.

**Fig. 1.** Interface of web-application SigNER

Named entity recognition is based on combined utilization of conditional random fields (CRF) algorithm and Naïve Bayes classifier. We used a set of rules to implement relation extraction. We manually analyzed a set of texts in order to find patterns between related entities. To build patterns, we considered only phrases that are commonly used for indicating association between two or more terms. As a result, we identified a set of proteins involved in virus-host interactions and immune suppression mechanisms shared by the studied viruses.

Detected proteins/genes named were then matched with differentially expressed genes. We presume that changes in the expression of these genes may underlie different responses to viral infections [7]. Therefore, we collected freely available transcriptomic data and analyzed them. Analysis was performed using R and BioConductor package.

*Results:* We constructed requests to PubMed in order to compile text corpus dedicated to mechanisms of action between human immune system and the following viruses: SARS-CoV-2, herpesviruses, hepatitis B, C, influenza, and HIV. We performed named entity recognition using the developed workflow. Then, we created a set of rules that allows to indicate relationships between proteins/genes named entities within the text. For example, APOBEC3 family members and SAMHD1 were identified as restriction factors for DNA- and RNA-viruses [8, 9]. These relationships were used to identify mechanisms of host response to viruses and immune suppression, caused by the viruses. In order to test the performance accuracy of our method, identification of differentially expressed genes was also carried out by transcriptomic data analysis.

*Conclusion:* Presented approach allowed us to collect data on genes involved in immune system reactions on different viral infections. In addition to evidence based on literature sources, the workflow also provides experimental data. Our approach can be used for design of novel drugs inhibiting virus-mediated immune suppression, and it can open the ways for drug repurposing.

*Acknowledgements:* This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030), project No. 122030100170-6.

## References

1. Piret J., Boivin G. Pandemics throughout history. *Front Microbiol.* 2021;11:631736.
2. Fragkou P.C. et al. Host immune responses and possible therapeutic targets for viral respiratory tract infections in susceptible populations: a narrative review. *Clin Microbiol Infect.* 2022;S1198743X22001495.
3. Piconese S. et al. Viral hepatitis, inflammation, and cancer: A lesson for autoimmunity. *J Autoimmun.* 2018;95:58-68.
4. Danastas K. et al. Herpes Simplex Virus Type 1 Interactions with the Interferon System. *Int J Mol Sci.* 2020;21(14):5150.
5. SigNER URL [<http://www.way2drug.com/signer/>]. Accessed on 24.04.2022.
6. Naïve-Bayes CNER URL [<http://www.way2drug.com/cner/>]. Accessed on 24.04.2022.
7. Ivanov S. et al. A computational analysis of transcriptional profiles from CD8(+) T lymphocytes reveals potential mechanisms of HIV/AIDS control and progression. *Comput Struct Biotechnol J.* 2021;19:2447-2459.
8. Milewska A. et al. APOBEC3-mediated restriction of RNA virus replication. *Sci Rep.* 2018;8:5960.
9. Sze A. et al. SAMHD1 host restriction factor: a link with innate immune sensing of retrovirus infection. *J Mol Biol.* 2013;425:4981-4994.

## A non-redundant set of long-range RNA tertiary interactions and their classification

Bohdan D.<sup>1\*</sup>, Baulin E.<sup>2</sup>

<sup>1</sup> *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia*

<sup>2</sup> *International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland*

\* *bogdan.d@phystech.edu*

**Key words:** RNA structure, non-coding RNA, long-range interactions, tertiary motifs

*Motivation and Aim:* It is now widely known that non-coding RNA molecules are a key player of molecular machinery in living cells. Non-coding RNA functions are largely determined by their 3D structure, which is of modular architecture and is made of building blocks of various sizes called tertiary motifs. In the context of RNA secondary structure, all the motifs can be divided into two groups – local motifs and long-range motifs. Local motifs are relatively well described and are subject to a natural comprehensive classification based on five standard secondary structure elements: a hairpin, a bulge, an internal loop, a multiple junction, and a stem. In turn, long-range motifs are formed by interactions between distant secondary structure elements and, at the moment, lack a systematic description. Thus, the purpose of the work is to search for long-range motifs and determine their systemic characteristics.

*Methods and Algorithms:* In this work, we compiled a non-redundant set of 89 PDB entries with resolution better than 4.0 Å. The set included RNA 3D structure representatives of 97 non-coding RNA families from Rfam. The share of ribosomal RNA residues made up only 58 %, unlike up to 82 % in conventional non-redundant lists of RNA structures. Using the original description of RNA secondary structure elements, we detected 17272 nucleotide residue pairs forming long-range interactions which we defined as two nucleotides on a distance closer than 6Å that belong to loops and stems distant from each other. Such a simple definition allowed us to annotate all possible direct and mediated interactions without having to set any additional thresholds to define particular interaction types. Then we intersected our nucleotide pairs with the results of several automatic RNA tertiary motifs and interactions annotation tools. The annotations were improved by superposition and clustering procedures and used to build a comprehensive classification.

*Results:* The result of our work is a dataset of nucleotide pairs forming long-range interactions along with a comprehensive annotation of the interaction types and the structural context of the interacting residues. The dataset is highly heterogeneous, e.g. the adenine(loop)-cytosine(stem) and adenine(loop)-guanine(stem) pairs that correspond to the most common type of long-range RNA tertiary motifs, the A-minor motif, compose only 9 and 12 % of the dataset respectively.

*Conclusion:* The produced classification and the dataset improves our understanding of the diversity of long-range RNA tertiary interactions and will be a helpful contribution to the field of non-coding RNA 3D structure modeling.

## G-quadruplex structures in the genomes of Human Papillomaviruses 16, 18 and 58: Promising molecular targets

Kantún N.<sup>1</sup>, Ontiveros N.J.<sup>1,2\*</sup>, Montero J.<sup>3</sup>

<sup>1</sup> Centro de investigación Regional Dr. Hideyo Noguchi, Yucatan, Mexico

<sup>2</sup> Universidad Autónoma de Yucatán, Yucatan, Mexico

<sup>3</sup> Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Mérida, Yucatán, Mexico

\*josynan9@gmail.com

DNA structure is a field little addressed in Human papillomaviruses (HPV) that cause common reproductive tract infections and a variety of cancers. Among the diversity of Non-B DNA conformations can gain, the G-quadruplex (G4) structures that form in guanines (G)-rich regions stand out. They are abundant in living organisms and viruses and their formation is not random in the organism studied. In addition, they are involved in important cellular processes [1]. The structural diversity, folding topologies, and *in vitro* stability of G4s put them in the spotlight to be considered as a promising new drug target for small molecules or to exert antiviral activity. The aim of this work is the *in silico* characterization and modelling of canonical and typical G4 in the genomes of three high-risk HPV by using bioinformatics tools that facilitate this type of study in terms of cost and time. For this, we selected the most appropriate bioinformatic predictors based on established criteria and identified typical and atypical G4 motif-forming sequences in both DNA strands of HPV 16, 18 and 58, as well as in some of their variants. More than 100 G4 motifs were identified for each HPV, most of which correspond to atypical G4. These are characterised by a higher frequency of two consecutive guanines in each tetrad. The L1 gene region encoding an important structural protein was one of the most abundant areas for G4 motifs for all three HPVs. We have identified two G4 motifs that stand out for their conservation (identities of 83.3 and 80.76 %). These same motifs were also identified in the variants, so they could be potential targets for study.

### References

1. Spiegel J., Adhikari S., Balasubramanian S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* 2020;2(2):123-136.



## Predicting the protein-ligand interactions based on ligands' structures and proteins' sequences

Karasev D.<sup>1\*</sup>, Sobolev B.<sup>1</sup>, Lagunin A.<sup>1,2</sup>, Filimonov D.<sup>1</sup>, Poroikov V.<sup>1</sup>

<sup>1</sup> *Laboratory of Structure-Function Based Drug Design, Institute of Biomedical Chemistry, Moscow, Russia*

<sup>2</sup> *Department of Bioinformatics, Pirogov Russian National Research Medical University, Moscow, Russia*

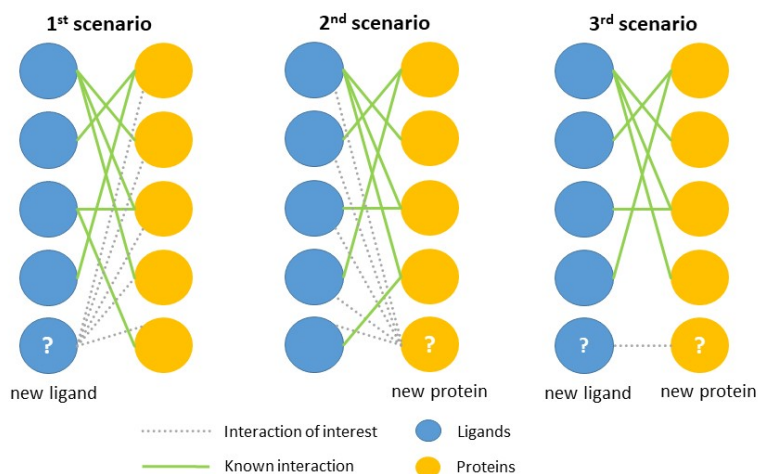
\* *w.dmitrykarasev@gmail.com*

**Key words:** proteochemometrics, protein-ligand interaction, drug design

*Motivation and Aim:* Use of computational methods predicting protein-ligand interactions significantly facilitate finding of biologically active compounds and identification of their molecular targets. Existing ligand-based drug design methods are limited due to the incompleteness of data about ligand-target relationships. Overcoming this limitation, the so-called proteochemometrics methods implement the predictive models including structural features of both the ligands and targets. Currently available methods have a limited applicability domain and are commonly not available online. We propose the proteochemometrics method, providing the wider applicability, and present it online as a web-service.

*Methods and Algorithms:* Information about protein-ligand interaction was extracted from ChEMBL'28 database. An approach was developed to combine data from heterogeneous experimental sources and evaluate their reliability. We provided prediction options for the most popular scenarios that arise in the computer-aided evaluation of protein-ligand interactions (Fig. 1). The first scenario is designed to assess the ligand interaction with protein by comparing the ligand structures with the PASS approach. In PASS (Prediction of Activity Spectra for Substances), the MNA descriptors (multilevel neighborhoods of atoms) are used as an input for the predictor based on a Naive Bayes classifier [1]. In the second scenario, the protein interaction with ligand is assessed by comparing the amino acid sequences with SPrOS method using the local sequence similarity. The use of the third scenario is required if the interaction spectra are not available for both the ligand and target. In this case, the prediction is based on handling both the ligand structures and protein sequences [2].

*Results:* We collected training sets for various groups of targets differed in protein target heterogeneity. Two sets included the data on families of homologous targets (protein kinases and G-protein-coupled receptors) and their ligands. The third set united data irrelative to protein families. Testing under the first scenario showed a high prediction accuracy of 0.96–0.99. Testing under the second scenario showed an accuracy of 0.97–0.99 for prediction within specific protein families and sets disordered by protein homology. The third scenario showed the lower but acceptable accuracy of 0.86–0.94. The developed web service allows the user to perform the prediction under the specified scenarios, thereby covering the needs of researchers when planning experimental studies [3].



**Fig. 1.** The most popular scenarios that arise in the computer-aided evaluation of protein-ligand interactions

Web service is available online at the web-site: <http://way2drug.com/proteochemometrics/>.

**Conclusion:** The developed freely available web service provides the opportunity to use the proteochemometrics technology to a wide range of users. Researchers can use both classical ligand structure prediction and look for likely ligands for a chosen target. The method makes it possible to predict protein-ligand interactions for popular protein families representing promising drug targets.

**Acknowledgements:** This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030), project No. 122030100170-5.

### References

1. Filimonov D.A. et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. *Biomed Chem Res Method.* 2018;1(1):e00004.
2. Karasev D.A. et al. Prediction of protein-ligand interaction based on sequence similarity and ligand structural features. *Int J Mol Sci.* 2020;21:8152.
3. Karasev D.A. et al. The method predicting interaction between protein targets and small-molecular ligands with the wide applicability domain. *Comp Biol Chem.* 2022;98:107674.

# Developing an accurate force field for simulating modified RNA

Lahiri A.

University of Calcutta, Kolkata, India  
albmbg@caluniv.ac.in

**Key words:** post-transcriptional modifications, molecular dynamics, glycosidic torsion parameter

*Motivation and Aim:* Post-transcriptional modifications occur extensively in RNA and provide an additional layer of chemical coding that presumably alters the chemistry, structure and dynamics of the molecules. Molecular dynamics simulations offer a convenient tool to elucidate the mechanisms underlying the roles played by such modifications. However, since the reliability of the simulation results depend on the accuracy of the force field employed, we tested some of the available force field parameters for modified RNA residues for the AMBER molecular modeling suite of programs. The results from our work revealed that the available parameters from the AMBER distribution were inadequate in reproducing the experimentally observed conformational distributions [1].

*Methods and Algorithms:* We earlier adopted a procedure for reoptimizing the dihedral parameters of a few modified uridines using high-level ab initio calculations [2]. Subsequently, we have revised the method to incorporate the recalculation of partial charges and the reoptimization of selected dihedrals. The revised parameters are currently applicable to a much larger set of modified RNA residues.

*Results:* The reoptimized parameter set has been validated against experimentally observed conformational distributions for a considerable number of modified residues. We have also shown that for some of the modifications, like 2-thiolation of uridines and pseudouridylation, the modifications stabilize the RNA structure. A major role in such stabilization is played by the enhancement in stacking interaction.

*Conclusion:* The results obtained suggest that the combination of AMBER force field for RNA with our reoptimized set of dihedral parameters and the recalculated partial charges provide a better agreement with experimental data and as such should be more useful in producing a more accurate model of the structure and function of modified RNA.

*Acknowledgements:* The study is supported by the SERB grant (EMR/2016/007753) and the computational facility provided by the Poznan Supercomputing and Networking Center.

## References

1. Deb I. et al. Conformational preferences of modified uridines: comparison of AMBER derived force fields. *J Chem Inf Model.* 2014;54(4):1129-1142.
2. Deb I. et al. Reparameterizations of the  $\chi$  torsion and Lennard-Jones  $\sigma$  parameters improve the conformational characteristics of modified uridines. *J Comput Chem.* 2016;37(17):1576-1588.

## Detecting broad specter of viral nucleic acids via NGS: novel k-mer-based algorithm for multiplex degenerate primer sets design

Matsvay A.D.\*, Stetsenko I.F., Ayginin A.A., Mikhaylov I.M., Bogacheva A.I., Say A.V., Shipulin G.A.

*Federal State Budgetary Institution "Centre for Strategic Planning and Management of Biomedical Health Risks" of the Federal Medical Biological Agency, Moscow, Russia*

\* [AMatsvay@cspmz.ru](mailto:AMatsvay@cspmz.ru)

**Key words:** multiplex PCR, primer design, k-mer, algorithm

**Motivation and Aim:** Multiplex PCR is a widely used method of target enrichment prior to library preparation for next generation sequencing. Compared to other techniques, it is time-efficient, cheap and provides very high proportion of on-target reads. However, designing of primer panel for multiplex PCR is challenging. In case of applying multiplex PCR for detection of microorganism, a commonly used approach is constructing a multiple sequence alignment in order to identify sequence regions conserved between organisms of interest using, e.g., ClustalW [1]. However, if genomes of organisms of interest are diverged enough, which is a common case for taxonomically distant viruses, construction of MSA may not be possible at all, or the resulting alignment may be not suitable for primer design.

**Methods and Algorithms:** A newly developed algorithm avoids this issue and allows designing a primer panels for parallel detection of diverged viral taxonomic groups without MSA construction. It utilizes k-mer approach and automatically selects degenerate primers, considering their thermodynamic characteristics, representative database coverage, predicted sizes of amplified sequences, stability of homo- and hetero-dimeric structures and hairpins and the stability of 3' end annealing. Relevant degenerate positions are introduced during the early stages of the algorithm, which allows to calculate unbiased statistics at subsequent steps. Using GPU parallelization for matrix calculation allows processing large databases containing highly diverged genomes of phylogenetically distant groups of viruses, which was shown to be impossible with other existing k-mer-based primer design software [2].

**Results:** As a proof of concept, the algorithm was used to design a primer panel targeting all known viruses causing respiratory infection. As predicted by *in silico* modelling, panel provides 99.81 % coverage of all known representatives (16 genera of 10 different families). Compared to primer sets described earlier, the newly developed primer set provides a number of benefits, including smaller degenerate position count and broader viral genome coverage per single structure within primer pair. The primer was used to prepare NGS libraries from a number of control samples, containing human adenoviruses (types 1–7, 21), bocavirus, alpha- (229E) and betacoronaviruses (SARS-CoV-2, type 1), influenza B virus, measles virus, meta- and orthopneumovirus, para-influenza virus (types 1–3) and rhinovirus (types 13, 15, 17, 26, 29) nucleic acids, and the sequencing results were 100 % concordant with viral identification by real time PCR-based test systems.

### References

1. Chenna R. et al. *Nucleic Acids Res.* 2003;31(13):3497-3500.
2. Hysom D.A. et al. *PLoS One.* 2012;7(4):e34560.

## MnFe<sub>2</sub>O<sub>4</sub> nanoparticles invasion in nose-associated lymphoid tissue

Morozova K.<sup>1,2\*</sup>, Kiseleva E.<sup>1</sup>, Romashchenko A.<sup>1</sup>, Moshkin M.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* morozko@bionet.nsc.ru

**Key words:** magnetic nanoparticles, capture of nanoparticles by cells, lymphoid tissue, magnetic resonance imaging, transmission electron microscopy

*Motivation and Aim:* Over the past decades, nanomaterials have become widespread in medicine. For example, at the moment, magnetic nanoparticles (MNPs) are already used as contrast agents in magnetic resonance imaging (MRI), magnetic indicators in magnetic particle imaging (MPI) and alternating current bioimpedance (ACB), as well as in magnetic hyperthermia therapy [1]. In addition, MNPs are actively used as a platform for targeted delivery of therapeutic agents, including in the development of vaccines [2]. At the same time, recent data show that MNPs in vaccines can act as adjuvants with immunostimulatory properties. One of the most promising approaches to inducing the production of specific antibodies to various pathogens that affect the upper and lower respiratory tract are intranasal vaccines that stimulate local lymphoid tissue, which, both in humans and in other terrestrial mammalian species, including rodents, is located in nasal passages (nasal associated lymphoid tissue (NALT)). In this connection, the study of the factors causing the capture of nanosized objects by cells of nasal epithelium, including NALT, is of fundamental importance for understanding the mechanisms of interaction of the local immunity system with viral and bacterial pathogens, and practical, for increasing the effectiveness of vaccination.

*Methods and Algorithms:* Using transmission electron microscopy (TEM) and magnetic resonance imaging (MRI), we studied the capture of intranasally administered MnFe<sub>2</sub>O<sub>4</sub> nanoparticles (MnFe<sub>2</sub>O<sub>4</sub>-NPs) by nasal epithelial cells, including NALT cells. TEM and MRI contrast MnFe<sub>2</sub>O<sub>4</sub> NPs were administered to Balb/c mice (age: 12 weeks, Th2 immune response). An hour after the particle injection, the nasal epithelium was isolated and fixed for subsequent TEM.

*Results:* Using both MRI and TEM, we confirmed the accumulation of intranasally injected MnFe<sub>2</sub>O<sub>4</sub>-NPs in mouse NALT. In addition, we discovered, for the first time, that inorganic nanoparticles are able to penetrate the cilia membrane of epithelial cells and be transported to the cytoplasm via intraflagellar retrograde transport. Previously, such transport was described only for viral particles. We also demonstrated the endocytosis-independent nature of the uptake of MnFe<sub>2</sub>O<sub>4</sub>-NPs from the nasal cavity using MRI in experiments with preliminary administration of specific and nonspecific inhibitors of macropinocytosis, clathrin-, calveolin-, and dynamin-dependent types of endocytosis. Successive stages of NP invasion are described. In the cytoplasm, NPs accumulate inside the membrane vacuole.

*Conclusion:* The results obtained, on the one hand, demonstrate the possibility of using MNPs to increase the efficiency of antigen delivery to local immunity cells, and, on the

other hand, expand the understanding of the mechanisms of interaction between epithelial cells and nanosized objects.

*Acknowledgements:* The study is supported by budget project FWNR-2022-0015.

### *References*

1. Billings C. et al. Magnetic particle imaging: Current and future applications, magnetic nanoparticle synthesis methods and safety measures. *Int J Mol Sci.* 2021;22(14):7651.
2. Zhu M. et al. Applications of nanomaterials as vaccine adjuvants. *Hum Vaccines Immunother.* 2014;10(9):2761-2774.

## **Anthelmintic activity of antioxidants: *in vitro* effects on the liver fluke *Opisthorchis felineus***

Pakharukov Y.V.<sup>1\*</sup>, Ponomarev D.V.<sup>2</sup>, Pakharukova M.Y.<sup>2</sup>, Mordvinov V.A.<sup>2</sup>

<sup>1</sup> *Industrial University of Tyumen, Tyumen, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *pacharukovyu@yandex.ru*

**Key words:** antioxidant, flavonoid, liver flukes, praziquantel, coenzyme Q

*Background:* Currently, pharmacologists and medical parasitologists are searching for new agents against foodborne trematodiasis. Redox metabolism is important for parasites as far as long-lived adult parasites inside a mammalian host are exposed to redox challenges. Thus, parasite survival may depend on the ability to maintain the necessary balance between oxidation and antioxidation. The importance of redox metabolism for parasites is underscored by the number of relevant genes in helminth genomes, levels of expression of these genes, and the presence of relevant proteins in the secretome. Antioxidants have been poorly studied as anthelmintic agents, in particular against the foodborne trematodes.

*Methods and Algorithms:* Study of *in vitro* anthelmintic activity of nonenzymatic natural and synthetic antioxidants of various chemical structures was performed using standard motility and mortality assays against juvenile and adult *Opisthorchis felineus* worms. Kaplan–Meier survival curves were built using the 'survival' (v.2.38) R package. Four-parameter logistic regression was utilized to calculate IC<sub>50</sub> and standard error (SE) values ('drc 3.0-1' R package). The ANOVA lack-of-fit test was performed to assess the hypothesis that a proposed regression model fits the data well. Hematoxylin and eosin staining was performed to analyze the tegument structure of adult flukes after SkQ1 treatment.

*Results:* Promising agents have been found among both natural and synthetic compounds. The mitochondria-targeted antioxidant SkQ1 [10-(6'-plastoquinonyl) decyltriphenylphosphonium] in motility assays was as effective (half-maximal inhibitory concentration [IC<sub>50</sub>] 0.6–1.4 μM) as officinal therapeutic agent praziquantel (IC<sub>50</sub> 0.47–1.4 μM). Moreover, SkQ1 was significantly more effective than praziquantel in mortality assays. Furthermore, extensive tegument damage of the adult fluke was revealed after SkQ1 treatment. Flavonoids manifested potency too, with IC<sub>50</sub> values in a micromolar range (5.1–17.4 μM). Other natural and synthetic compounds tested against helminths were significantly less effective than praziquantel.

*Conclusion:* Results of our study indicate that SkQ1 and flavonoids have high anthelmintic activities against the liver flukes. We propose that structure–activity relationship research might be worthwhile based on the structures of the most effective substances.

*Acknowledgements:* This work was supported by the Russian Science Foundation (22-24-20010).

## Drug repurposing for COVID-19 therapy: challenges and opportunities

Poroikov V.

*Institute of Biomedical Chemistry, Moscow, Russia*

*vladimir.poroikov@ibmc.msk.ru*

**Key words:** SARS-CoV-2, COVID-19, virus-host interactions, drug repurposing, *in silico* prediction, *in vitro* validation, informational-computational platform

**Motivation and Aim:** Drug repositioning is the identification of new indications for the approved medicines [1]. The availability of information on pharmacological and toxicological characteristics provides conditions for rapid introduction of launched drugs for a new nosology. The need to respond quickly to the COVID-19 pandemic has spurred large-scale research in this direction. To identify new pharmacological effects of known drugs, *in silico* and *in vitro* studies are carried out [2]. Several large-scale experimental studies are devoted to *in vitro* screening for one or several targets from 1,400 to 12,000 drugs, making it possible to select some "candidates" for repositioning. In many cases, the results obtained for the same drugs in different test systems do not agree with each other. The observed contradictions could be explained by the absence of standardization of the assays developed by various researchers independently of each other and the lack of generally accepted reference drugs. Pre-selection of potentially active compounds based on *in silico* evaluations significantly increases the chances of success. The purpose of our research is the development of the Informational-Computational Platform (ICP) for drug repurposing aimed to COVID-19 treatment.

**Methods and Algorithms:** Informational modules of ICP integrates the current data about SARS-CoV-2 life cycle, virus-host interactions, mechanisms of pathological process at the early and later stages of the disease, most promising pharmacological targets, approved pharmaceuticals that may be analyzed as the candidates for repurposing. Computational modules of ICP includes the tools for (1) text mining to select the relevant information despite the "infodemic" [3], (2) network pharmacology to identify the promising molecular targets [4], (3) assessment of the chemical similarity using the distinct structures of the potent anticoronavirus agents as a query [5], (4) machine learning to predict wide biological activity profiles of the probable "candidates" [6]; (5) molecular modeling to estimate the plausibility of the "candidates" binding modes with the putative targets [7]. The general scheme of ICP is presented in Figure 1.

**Results:** The pilot ICP version is developed based on the information for the SARS-CoV-2 main protease (Mpro) inhibitors as the most studied anticoronavirus target. Currently it includes the available data on 1136 pharmaceutical substances considered as the potential Mpro inhibitors investigated *in silico* (1044 substances) and *in vitro* (304 substances). Database of World-Wide Approved Drugs including 4128 pharmaceutical substances is prepared for structural similarity assessment and application of machine learning methods. Structural similarity assessment is developed on the basis of MNA and QNA descriptors, which are successfully applied for analysis of structure-activity relationships with classification and regression approaches [5].



Prediction of anticoronavirus activity is carried out using machine learning method implemented in the computer program PASS [6]. Based on the available information, we selected the putative Mpro inhibitors for experimental testing, which could be presented in the following order: Disulfiram > Omeprazole > Silibinin > Saquinavir > Montelukast > Imatinib > Atazanavir > Dasatinib > Ciprofloxacin = Glycyrrhizic Acid > Dihydroquercetin > Nalraprevir = Teicoplanin = Bedaquiline = Doxazosin.

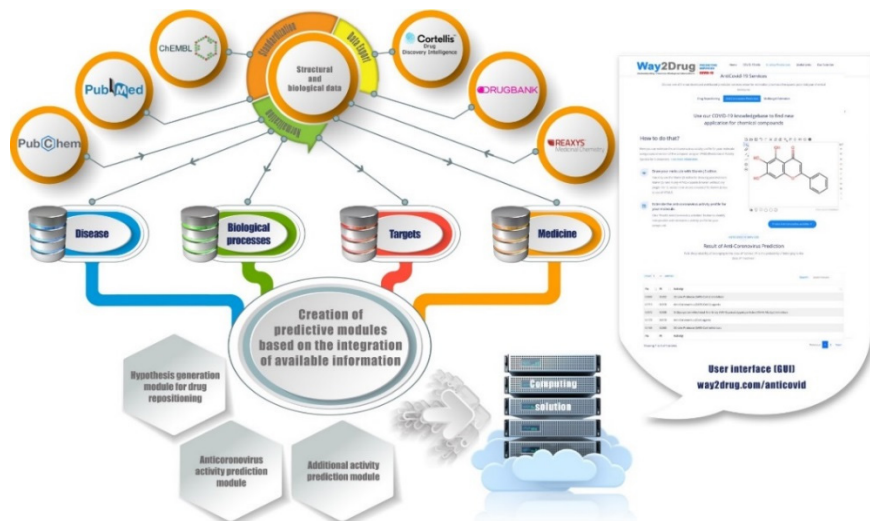


Fig. 1. Informational-computational platform for drug repurposing

**Conclusion:** We have developed a pilot version of ICP for drug repurposing to identify the most promising anticoronavirus drug-candidates and selected the most promising approved drugs as putative Mpro inhibitors. Some selected drug substances are currently tested by the Chumakov Federal Scientific Center for Research and Development of Immune-and-Biological Products of the Russian Academy of Sciences.

**Acknowledgements:** This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030), project No. 122030100170-6.

### References

1. Poroikov V., Druzhilovskiy D. Drug repositioning: new opportunities for older drugs. In: Roy K. (Ed.). *In Silico Drug Design, Repurposing Techniques and Methodologies*. Academic Press, 2019;3-17.
2. Muratov E.N. et al. A critical overview of computational approaches employed for COVID-19 drug discovery. *Chem Soc Rev*. 2021;50(16):9121-9151.
3. Biziukova N. et al. Automated extraction of information from texts of scientific publications: insights into HIV treatment strategies. *Front Genet*. 2020;11:618862.
4. Ivanov S. et al. Network-based analysis of OMICs data to understand the HIV-Host interaction. *Front Microbiol*. 2020;11:1314.
5. Druzhilovskiy D.S. et al. Computational approaches to identify a hidden pharmacological potential in large chemical libraries. *Supercomput Front Innov*. 2020;7(3):57-76.
6. Filimonov D.A. et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. *Biomed Chem Res Methods*. 2018;1(1):e00004.
7. Tarasova O. et al. Molecular docking studies of HIV-1 resistance to reverse transcriptase inhibitors: mini-review. *Molecules*. 2018;23:1233.

## A Bayesian approach to lipophilicity prediction

Pyakillya B.I.\*, Goncharov V.I.\*

*Tomsk Polytechnic University, School of Computer Science & Robotics, Tomsk, Russia*

\* *morphism@tpu.ru*

**Key words:** machine learning, Bayesian theorem, lipophilicity, drug design

*Motivation and Aim:* It is well-known that lipophilicity affects the absorption dynamics of drugs and chemical substances in total. Accurate prediction of lipophilicity is a key problem in drug discovery and other related biochemical or physicochemical problems [1]. Current rise in popularity of machine learning shed some light on this problem and gave state of the art models like graph neural networks or transformer-based. Unfortunately, most machine learning methods cannot afford full interpretation of its predictions to research community e.g., like models' uncertainty – which tends to be unknown. Overfitting is still a big issue in machine learning tasks, and it also must be taken into account based on models' uncertainty. Therefore, we proposed to use Bayesian approach to overcome this problem and consider the uncertainty of model's behavior [2].

*Methods and Algorithms:* We utilized the Bayesian approach in the neural networks design and used it in the problem of lipophilicity prediction. We used PubChem library as the training dataset and verified the Bayesian neural networks results in comparison to conventional multi-layer perceptrons. We proposed to use model's uncertainty as a standard deviation, which could be easily calculated in Bayesian sampling process.

*Results:* The connection between prediction error and uncertainty values were found, which gives a lot of information to computational chemistry groups and could be used as a guide to make the training process more valuable. In addition, complexity of Bayesian neural networks was found to be significantly lower than in the case of conventional neural networks. This complexity benefit comes without drop in prediction accuracy.

*Conclusion:* The results obtained suggest to use the Bayesian approach in the cheminformatics problems like property prediction tasks. The approach gives the ability to consider model's uncertainty and take into account how much effort a cheminformatics researcher should put to make the prediction accuracy higher. Moreover, lower amount of training parameters makes the training of Bayesian Neural networks more favorable in comparison to conventional models, which tend to overfit to the training data.

### References

1. Arnott J.A. et al The influence of lipophilicity in drug discovery and design. *Expert Opin Drug Discov.* 2012;10:863-875.
2. Kendall A. et al. What uncertainties do we need in Bayesian deep learning for computer vision? In: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017 Long Beach, California, USA. 2017:5580-5590.

## DNA-nanomachines for nucleic acid detection

Rubel M.<sup>1\*</sup>, Zablotskaya S.<sup>1</sup>, Pokatova O.<sup>1</sup>, El-Deeb A.<sup>1</sup>, Ateiah M.<sup>1</sup>, Gorbenko D.<sup>1,2</sup>, Shkodenko L.<sup>1</sup>, Kolpashchikov D.M.<sup>3</sup>

<sup>1</sup> ITMO University, St. Petersburg, Russia

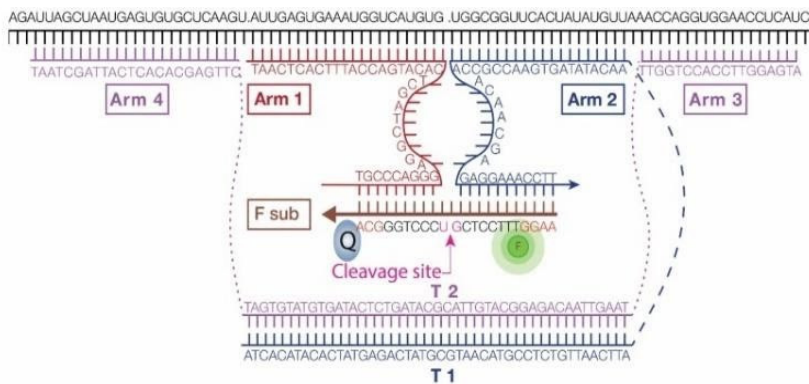
<sup>2</sup> Ioffe Institute, St. Petersburg, Russia

<sup>3</sup> University of Central Florida, Orlando, FL, USA

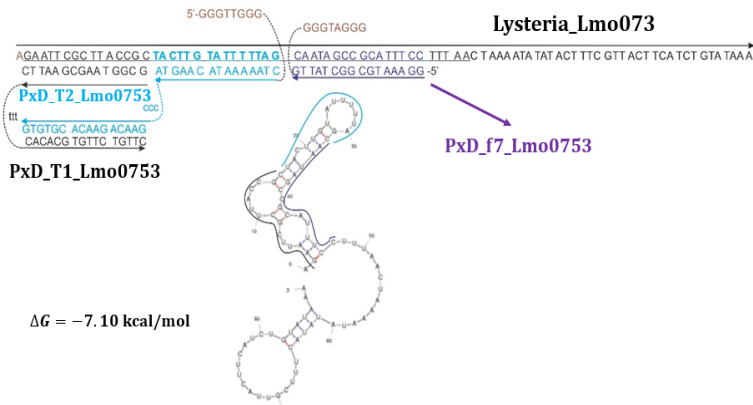
\* rubel@scamt-itmo.ru

**Key words:** DNA-nanomachines, DNAzyme, G-quadruplex, nucleic acid detection

**Motivation and Aim:** A DNA-nanomachine (DNM) is a powerful tool for nucleic acid detection. DNMs are oligonucleotide-based multifunctional nanostructures that can tightly bind nucleic acids, including folded single stranded (ss) RNA or double stranded (ds) DNA targets. DNMs can recognize single nucleotide substitutions (SNS) with a high selectivity and provide a readable output. DNMs are compatible with point-of-care (POC) diagnostics. We report here two DNM variants: (i) DNAzyme-based DNM with fluorescent output [1] (Fig. 1) and (ii) G-quadruplex based DNM with visual output [2] (Fig. 2).



**Fig. 1.** An example of a DNA-nanomachine for fluorescence detection of coronavirus



**Fig. 2.** An example of a DNA-nanomachine for colorimetric detection of *Listeria monocytogenes* and the fragment of interest

*Methods and Algorithms:* A series of DNM targeting sequence of 20 human pathogens were designed. We took in account predicted RNA secondary structure and optimized DNMs for differentiating SNS using UNAFold and NUPACK software. DNM were assembled from custom-made oligonucleotides and optimized for fluorescent or visual detection of synthetic ssDNA, dsDNA and RNA amplicons, unamplified bacterial rRNA or viral RNA of human pathogens.

*Results:* Optimized DNMs produced signals 1,5 to 4 above the background detect in all cases after 15 to 180 min assays. The sensitivity assays were down to 800 fM and in a micromolar range for fluorescent and visual outputs respectively. The sensitivity is dictated by the temperature of the environment, length of the recognition arms, its number, type of a core, and type of the nucleic acid. DNMs selectively recognized SNS under experimental conditions, including room temperature. The G-quadruplex-based DNM are suitable for rapid visual room temperature detection of dsDNA and RNA amplicons. The fluorescent DNMs can quantify nucleic acid without amplification, but require elevated temperature (55 °C) and prolonged incubation time.

*Conclusion:* The results confirm that DNM can be adapted for detection of dsDNA and folded ssRNA of various sequences. The DNM can be adjusted for SNS differentiation under broad experimental conditions. They are made of unmodified synthetic DNA and, therefore, are relatively inexpensive. DNM can be incorporated in POC diagnostic tests for visual or fluorescent detection of human pathogens.

*Acknowledgements:* The study is supported by the RSCF grant (22-24-00664).

### *References*

1. Cox A.J., Bengtson H.N., Rohde K.H., Kolpashchikov D.M. DNA nanotechnology for nucleic acid analysis: multifunctional molecular DNA machine for RNA detection. *Chem Commun (Camb)*. 2016;52:14318-14321.
2. Kovtunov E.A., Shkodenko L.A., Goncharova E.A., Nedorezova D.D., Sidorenko S.V., Koshel E.I., Kolpashchikov D.M. Towards point of care diagnostics: visual detection of meningitis pathogens directly from cerebrospinal fluid. *Chemistry Select*. 2020;5(46):14572-14577.

## Computational approach for searching bioactivity of natural products based on their 2D chemical structure

Sadovskaya A.\*, Gurevich A.

Center for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia

\* alexandrasad04@gmail.com

**Key words:** peptidic natural products, bioinformatics, bioactivity, chemical structure

*Motivation and Aim:* The object of this study is peptidic natural products (PNPs), such as ribosomal and nonribosomal peptides. PNPs are an important group of natural products that include many antibiotics, anticancer drugs, and other pharmaceuticals. Recent breakthroughs in mass spectrometry (MS) and genome sequencing enabled high-throughput PNP discovery. Emerging computational methods allowed identification of novel PNPs via modification-tolerant database search of MS data [1], *de novo* MS sequencing [2], and metabologenomics approaches [3]. These methods produce large amounts of *in silico* predicted PNPs that require further experimental analysis. Creating a robust strategy for selecting the most promising compounds out of many computationally predicted molecules remains an open problem.

*Methods and Algorithms:* We handle PNPs structures in standard chemical formats (currently SMILES, MDL MOL and SDF) and use the RDKit library to convert them into the SMILES chemical format. The resulting SMILES are queried in PubChem via the exact and similarity search modes [4]. We filter found molecules using the Tanimoto coefficient and the NP-likeness score to retain likely natural products having chemical structure similar to the query molecule. Key compound information, such as bioassay tests results and known producers, is summarized into a short report, while more detailed metadata for each identified molecule is saved into an extended report.

*Results:* In this work, we developed a computational pipeline for predicting biological activities of PNPs based on searching their chemical structure online in the PubChem database. Our tool is suitable for high-throughput analysis that distinguishes it from the alternatives: the PubChem information search and PASS, the leading bioactivity prediction software [5]. In addition, our method works with compounds absent in the PubChem database and also predicts taxonomic data missed by PASS. The created pipeline was tested on 550 PNPs from the MIBiG database. For 84 % of the PNPs, the pipeline found compounds identical in 2D structure in PubChem and generated a report with the biological activities described for them.

*Conclusion:* We develop a computational approach for predicting bioactivity of PNPs based on their 2D chemical structure. We plan to integrate the developing pipeline with the state-of-the-art approaches to PNPs discovery [1-3] to complement their output with the tentative biological activity of the identified compounds and likely taxonomy of their producers. We anticipate our pipeline will facilitate the search for new therapeutic agents by providing researchers with biologically relevant information. This data will help to prioritize *in silico* predicted PNPs for experimental validation and testing.

*Acknowledgements:* The study is supported by RSF (20-74-00032).

## References

1. Gurevich A. et al. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat Microbiol.* 2018;3:319-327.
2. Behsaz B., Mohimani H., Gurevich A., Prjibelski A., Fisher M., Vargas F. et al. De novo peptide sequencing reveals many cyclopeptides in the human gut and other environments. *Cell Syst.* 2020;10:99-108.
3. Cao L. et al. MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities. *Cell Syst.* 2019;9:600-608.
4. Kim S., Chen J., Cheng T., Gindulyte A., He J. et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2020;49:D1388-D1395.
5. Filimonov D.A. et al. Prediction of the biological activity spectra of organic compounds using the pass online web resource. *Chem Heterocycl Compd.* 2014;50:444-457.

## Development of a searching algorithm for new cell-penetrating peptides based on machine learning methods

Serebrennikova M.\*, Grafkskaia E., Latsis I., Lazarev V.

*Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*

\* *maria.serebrennikova.msu@gmail.com*

**Key words:** cell-penetrating peptides, machine learning algorithm, prediction

With the development of medicine, systems for the intracellular delivery of biologically active molecules become an increasingly urgent task. From this point of view, cell-penetrating peptides (CPPs) are attractive candidates for new approaches in various therapeutic areas, including cancer treatment. This is a class of peptides, from 5 to 60 amino acids in length, which are the most promising vectors for transporting various therapeutic agents. The presented study aimed to develop the predictive algorithm based on the ML techniques for the identification of new CPPs.

The first stage of the study was to create a sample of all currently known CPPs and non-CPPs sequences. Those that had a high level of homology or contained non-natural amino acid residues and synthetic modifications were excluded. Based on this, the training set consists of 2398 molecules ranging in length from 5 to 61 amino acid residues.

Further, molecular descriptors, that characterize the unique properties of each peptide, were determined for each sequence. Using dimensionality reduction methods, 20 optimal parameters were identified to optimize the predictive ability of the models.

At the following stage, using k-fold cross-validation and grid search, we determined the optimal values of incoming hyperparameters for each of the following algorithms: k-nearest neighbors, logistic regression, gradient boosting, random forest, and support vector model.

According to the results of evaluating the quality of the obtained predictive models searching for penetrating peptides, we choose the three best models based on k-nearest neighbors, gradient boosting, and random forest algorithms. Notice, that the resulting predictive method, in comparison with existing predictive models, demonstrates comparable results.

The predictive method was applied to search for new CPP using an independent test set of data obtained from open sources. A list of candidate CPPs was formed and their structural and functional characteristics were assessed.

*Acknowledgements:* The research was supported by the Russian Science Foundation (project No. 20–15–00270).

## Prediction of adverse effects of drug-drug interactions on the cardiovascular system based on the analysis of structure-activity relationships

Sukhachev V.\*, Ivanov S., Dmitriev A.

*Institute of Biomedical Chemistry, Moscow, Russia*

\* *withstanding@yandex.ru*

**Key words:** drug-drug interactions, SAR, adverse drug reactions, polypharmacy

*Motivation and Aim:* The combination of several drugs can lead to a number of diverse side effects, which may be followed by certain risks to life and health. The drug effects on the cardiovascular system are the most serious ones, since cardiovascular diseases are the leading cause of death worldwide [1]. Experimental determination of all possible drug interactions seems impracticable, given the many probable combinations of drugs. The aim of our study is to build a structure-activity relationships (SAR) models for predicting cardiovascular drug interactions.

*Methods and Algorithms:* Cardiovascular-related and non-cardiovascular data were extracted from the DrugBank [2] and TwoSides [3] databases. Overall, 283,439 drug-pairs were obtained, including 2,313 specific drugs causing 113 side effects, as well as 62,351 additional drug-pairs for negative examples which were opposed to pairs that cause a certain undesirable effect. Thus, we unambiguously categorize data into those that cause an effect and those that do not. The Pearson correlation coefficients (PCC) for each pair of side effects were calculated. Based on PCC values, building a model for blood pressure disorders, we excluded from negative examples drug-pairs related to blood pressure: hypotension, hypertension, vasodilation, vasoconstriction, etc.

The description of drug-pairs was performed using two types of descriptors: PoSMNA structural descriptors [4, 5]. It is a linear combination of structure fragments for two compounds pivoted in table when each row is a drug-pairs and each column is an appearance of concrete structures fragments in this pair. Totally, the data contains 345,790 rows of drug-pairs and 15,000 columns of structural elements present in each pair.

Linear combinations of estimates of the probability of biological activity obtained with PASS [6]. Totally, data contains 345,790 rows and 14032 columns for each activity. Also, the most optimal parameters of these descriptors were determined empirically, providing the greatest accuracy of prediction.

Classification models were built by Random Forest followed by 5-fold cross-validation for six most major side effects: arrhythmia, bradycardia, QT-interval prolongation, tachycardia, hypotension and hypertension. PASS descriptors corresponding to the activities that contributed most to the accuracy of the model were extracted to build a network of proteins that are involved in the mechanism of drug side effects.

*Results:* Accuracy parameters obtained for 12 models is given in the table below.



PoSMNA						
	Bradycardia	arrhythmia	QT	Tachycardia	Hypertension	Hypotension
AUC	0,91	0,88	0,88	0,93	0,94	0,89
SENS	0,71	0,69	0,69	0,73	0,75	0,67
SPEC	0,97	0,99	0,91	0,98	0,95	0,91

PASS						
	Bradycardia	arrhythmia	QT	Tachycardia	Hypertension	Hypotension
AUC	0,94	0,90	0,90	0,96	0,91	0,89
SENS	0,65	0,55	0,76	0,70	0,85	0,77
SPEC	0,97	0,96	0,86	0,97	0,82	0,85

The average estimates of the accuracy parameters are AUC: 91 %, sensibility: 71 %, specificity: 95 %. A high AUC value indicates that the models can produce a binary classification with high accuracy. Close values of sensitivity and specificity indicate that the model makes it possible to distinguish with almost equal accuracy a pair of drugs that cause this effect and those that do not.

Also, protein-protein interactions networks were built based on the descriptors that have the greatest contribution to the accuracy of the model.

*Conclusion:* The constructed SAR models make it possible to predict a number of undesirable drug interactions with high accuracy. The characteristics of the models are interpretable and may be used to uncover the mechanisms of drug-drug interactions. The accuracy parameters obtained by two different descriptors are very close. This justifies that the chosen mode of describing a pair of drugs make it possible to make a prediction.

*Acknowledgements:* The authors are grateful for the support by the Russian Science Foundation Grant (Grant No. 17-75-20250) of this study.

### References

1. The top 10 causes of death URL [<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>]. Accessed on 24.04.2022.
2. Wishart D. et al. *Nucleic Acids Res.* 2018;46(1):D1074-D1082.
3. Offsides and Twosides (NSIDES v0.1) – TLab URL [<http://tatonettilab.org/offsides>]. Accessed: 24.04.2022.
4. Filimonov D.A. et al. *Chem Heterocycl Compd.* 2014;50(3):444-457.
5. Filimonov D.A. et al. *J Chem Inf Comput Sci.* 1999;39(4):666-670.
6. Filimonov D.A. et al. *Biomed Chem Res Methods.* 2018;1(1):e00004.

## Network pharmacology and molecular docking techniques to elucidate the mechanisms of *Ocimum sanctum* against tuberculosis

Tabassum S.<sup>1\*</sup>, Khalid H.R.<sup>1</sup>, ul Haq W.<sup>2</sup>, Aslam S.<sup>1</sup>, Alshammari A.<sup>3</sup>, Alharbi M.<sup>3</sup>, Rajoka M.S.R.<sup>4</sup>, Khurshid M.<sup>5</sup>, Ali Ashfaq U.<sup>1</sup>

<sup>1</sup> Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan

<sup>2</sup> Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan

<sup>3</sup> Department of Pharmacology and Toxicology, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

<sup>4</sup> Laboratory of Animal Food Function, Graduate School of Agricultural Science, Tohoku University, Sendai, Japan

<sup>5</sup> Department of Microbiology, Government College University, Faisalabad, Pakistan

\* tabasumughal@gmail.com

**Key words:** Network pharmacology; molecular docking; *Ocimum sanctum*; tuberculosis

*Aim of Study:* Mycobacterium tuberculosis (TB) is a notable health care load that imposes a serious impact on the quality of life of patients. The least reported data and multiple spectra of pathophysiological mechanisms of TB make it a challenging task and a serious economic burden in health care management. The presented study aims to discover the potential phenomenon of *Ocimum sanctum* in the medication of tuberculosis. This study uncovers the active ingredients [1], their potential targets, and signaling pathways in *O. sanctum* for the treatment of TB.

*Methodology:* In the framework of this study, we used network pharmacology approach to explore the "compound-target-pathway" network [2]. The targets of the active ingredients of *Ocimum sanctum* and tuberculosis were collected by SwissTargetPrediction, DisGeNet and GeneCard. After the screening of mutual targets of TB and active ingredients, enrichment analysis through DAVID was performed. Then the compound-target-pathway network was constructed with Cytoscape. Later, molecular docking was employed to validate the successful activity of the active compounds against potential targets of TB. *Results:* we get 8 active ingredients out of 48 available compounds of *O. sanctum*, after that we get 72 functionally enriched potential target of tuberculosis found in active ingredients. By network analysis five hub genes (AKT1, MAPK14, TNF, CASP3 and MAPK1) were screened to check the interaction between active compounds and potential targets. The docked complexes reveal that compounds of *O. sanctum* like flavones, terpenoids and sterol play a crucial role in the medication of tuberculosis.

*Conclusion:* Lastly, we conclude that eight highly active constituents help in the medication of disease by regulating the expression of AKT1, MAPK14, TNF, CASP3, and MAPK1, which may act as potential therapeutic targets for TB. These results revealed that bioactive compounds of *O. sanctum* act on proteins along with their pathways to frame a network analysis in systems pharmacology, having great worth in

drug development and utilization. These findings serve as a starting point for further research to explore the potential mechanism of *O. sanctum* for TB treatment.

### *References*

1. Borah R., Biswas S.P. Tulsi (*Ocimum sanctum*), excellent source of phytochemicals. *Int J Environ Agric Biotech.* 2018;3(5):265258.
2. Zhang X. et al. Network pharmacology based virtual screening of active constituents of *Prunella vulgaris* L. and the molecular mechanism against breast cancer. *Sci Rep.* 2020;10(1):15730.

## Computational analysis of viral drug resistance and treatment efficacy: case study for HIV-infection

Tarasova O.\*, Ivanov S.M., Poroikov V.V.

*Institute of Biomedical Chemistry, Moscow, Russia*

\* [olga.a.tarasova@gmail.com](mailto:olga.a.tarasova@gmail.com)

**Key words:** viral infections, drug resistance, treatment efficacy, HIV

*Motivation and Aim:* Treatment of most viral diseases may include prescription of several different therapeutic agents for inhibition of a virus replication, immunomodulation, and comorbidities therapy [1–4]. Viral drug resistance and individual patients' response to therapy can significantly decrease the efficacy of antiviral treatment [3–5]. Therefore, to overcome these limitations development of new therapeutic strategies is constantly needed along with the search for new mechanisms of viral infections progression. In our study, we developed a multifaceted computational approach for analysis of viral drug resistance and treatment efficacy. The developed approach can be used for selection of the best therapeutic approaches and implementing new strategies to treat viral infections. We tested this approach on the HIV infection as a case study. Highly active antiretroviral therapy is the gold standard in treatment and cure of HIV-positive patients [4]. HIV drug resistance decreases the outcome of antiretroviral therapy. The methods designed for predicting treatment efficacy and novel therapeutic strategies are helpful for these purposes.

*Methods and Algorithms:* The approach is based on the computational analysis of (i) HIV drug resistance, (ii) HIV drug exposure using treatment history, and (iii) the transcriptomic data of human blood samples obtained from HIV-positive patients. Data on drug resistance including HIV proteins sequences associated with drug resistance and HIV treatment history were obtained from the Stanford HIV drug resistance (StDB) [6] and RHIVDB databases [7, 8]. The transcriptomic data were obtained from the key transcriptomic repositories (GEO, ArrayExpress).

*Results:* We implemented a web-interface of RHIVDB [8] that include the sequences of HIV-1 proteins (reverse transcriptase (RT), protease (PR), and integrase (IN)) with the data on the drug treatment schemas for the patients carrying a particular viral variant. Using the data available from RHIVDB combined with the data from StDB, we created models for predicting the drug exposure and therapy efficacy based on the Bayesian approach and the set of machine learning methods including random forest, deep learning, support vector machines (Fig. 1). The features for the models were built based on the amino acid and nucleotide sequences of HIV RT, PR and IN.

Additionally, we considered the prediction of treatment efficacy based on the transcriptomic data of patients with HIV infection. We identified the feature importance for the models using amino acid and nucleotide sequences and based on the transcriptomic data. In particular, we determined the level of transcription of particular genes that contributed to the antiretroviral therapy outcome.

**Classify nucleotide sequences of HIV protease and reverse transcriptase gene into resistant and susceptible to antiretroviral treatment using PASS approach**

The classification is based on the analysis of nucleotide sequences the HIV protease or reverse transcriptase gene.

The method of classification is based on the estimation of the occurrence  $F_r$  and  $F_s$  of short nucleotide sequence fragments in the set of genes of Resistant (R) and susceptible (S) HIV variants respectively.

The  $F_r$  and  $F_s$  values are then processed in the Bayesian-based algorithm PASS, which makes it possible to calculate  $P_r$  and  $P_s$  values of probability for the particular sequence to belong either to R or S variant.

Upload Sequence:  Не выбран ни один файл

Enzyme encoded:

Download Results

\* indicates required fields

Results are saved in the the tab-delimited file containing the rows, each of which has sequence identifier of an uploaded FASTA file, values  $P_r$  and  $P_s$ . See [results\\_example.txt](#) as an example

If you have any question, please [contact us](#)

**Fig. 1.** A part of web-platform developed for predicting HIV drug resistance and treatment efficacy in application of highly active antiretroviral therapy: web-interface of naïve-Bayes viral drug resistance predictor

The average balanced accuracy of prediction for drug resistance and drug exposure ranged from 0.85 to 0.96 for different drugs and drug combinations, the balanced accuracy of therapeutic efficacy prediction ranges from 0.83 to 0.91. All our models are freely available on the Internet [9].

**Conclusion:** We developed a computational approach for predicting drug resistance and therapy efficacy and tested it in the case study of HIV infection. The average balanced accuracy of therapy efficacy prediction reaches 0.96 for particular drugs and drug combinations. Therefore, the developed approach can be used for drug therapy regimen optimization and search for new treatment strategies.

**Acknowledgements:** This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030), project No. 122030100170-5.

## References

1. Nguyen M.H. et al. Hepatitis B virus: Advances in prevention, diagnosis, and therapy. *Clin Microbiol Rev.* 2021;33(2):e00046-19.
2. McFee R.B. COVID-19: Therapeutics and interventions currently under consideration. *Dis Mon.* 2020;66(9):101058.
3. Naggie S., Lok A.S. New therapeutics for Hepatitis B: the road to cure. *Ann Rev Med.* 2021;72:93-105.
4. Masters M.C. et al. Beyond one pill, once daily: current challenges of antiretroviral therapy management in the United States. *Expert Rev Clin Pharmacol.* 2020;12(12):1129-1143.
5. Lu D.-Y. et al. HAART in HIV/AIDS treatments: future trends. *Infect Disord Drug Targets.* 2018;18(1):15-22.
6. Shafer R.W. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis.* 2006;194(Suppl. 1):S51-S58.
7. Tarasova O. et al. RHIVDB: A freely accessible database of HIV amino acid sequences and clinical data of infected patients. *Front Genet.* 2021;12:679029.
8. RHIVDB URL: [<http://www.way2drug.com/rhivdb/>]. Accessed on 24.04.2022.
9. Naïve Bayes HIV viral drug resistance predictor URL: [<http://www.way2drug.com/resistance/>]. Accessed on 24.04.2022.

## **Catalytic antibodies after COVID-19 and vaccination hydrolyze SARS-CoV-2 S-protein and corresponding oligopeptides**

Timofeeva A.M.<sup>1\*</sup>, Shayakhmetova L.Sh.<sup>1,2</sup>, Sedykh S.E.<sup>1,2</sup>, Nevinsky G.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *anna.m.timofeeva@gmail.com*

The most important part of the humoral response to SARS-CoV-2 are antibodies against the S-protein and RBD. The presence of anti-S-IgG and anti-RBD-IgG has been correlated with SARS-CoV-2 neutralizing activity. Antibodies to the N-protein indicate the past disease. In addition, the S-protein is the primary immunogen for the development of vaccines such as Sputnik V, Moderna, Pfizer-BioNTech and AstraZeneca. Therefore, the study of antibodies to the S-protein and RBD is essential for understanding the immune processes during SARS-CoV-2 infection and vaccination. The donors' blood plasma was screened for antibodies to S- and N-proteins of the SARS-CoV-2 virus. Using affinity chromatography on protein-G-Sepharose we isolated IgG preparations from the blood plasma of donors, and show their electrophoretic homogeneity. After affinity chromatography on sorbents with immobilized recombinant SARS-CoV-2 S-protein and its RBD fragment, we obtained antibody subfractions with an affinity for these antigens (named S-IgG and RBD-IgG respectively).

The content of RBD-IgG in the blood plasma of COVID-19 convalescents and patients vaccinated against SARS-CoV-2 was about 1 %. Antibodies not bound to the sorbent after chromatography on RBD-Sepharose, were applied to Sepharose with immobilized S-protein. The amount of this subfraction of antibodies was less than 1 %.

The S-IgG and RBD-IgG subfractions were analyzed in hydrolysis of RBD and corresponding fluorescent 18-mer oligopeptides. All RBD-IgG and S-IgG subfractions we analyzed were inactive in RBD hydrolysis, but RBD-IgG were more active in the hydrolysis of oligopeptides compared to S-IgG. It has been shown that some oligopeptides were hydrolyzed in a wide pH range or within a pH optimum 6.5–7.5. The data we obtained indicate the differences in the spectrum of antibodies generated during the immune response against the S-protein in COVID-19 convalescents and patients vaccinated with Sputnik V.

*Acknowledgements:* The work was supported by the Russian Science Foundation 21-75-10105 (to Timofeeva A.M.)

## ***Nigella sativa* compounds as novel EGFR inhibitor using DFT, molecular docking and molecular dynamics**

Tufail A.<sup>1\*</sup>, Dubey A.<sup>2,3\*</sup>

<sup>1</sup> Department of Biochemistry, University of Allahabad, Allahabad, India

<sup>2</sup> Computational Chemistry and Drug Discovery Division, Quanta Calculus Pvt. Ltd., Kushinagar, India

<sup>3</sup> Department of Pharmacology, Saveetha Dental College and Hospital, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

\* aishatufailansari@gmail.com, ameebioinfo@gmail.com

**Key words:** EGFR, *Nigella sativa*, DFT, Molecular Docking, Molecular Dynamics

**Motivation and Aim:** Health problems are increasing around the globe regarding cancer modalities. With approximately 10 million deaths expected in 2020, cancer is the leading cause of death worldwide [1]. The epidermal growth factor receptor (EGFR; HER1 in humans) is a transmembrane protein that belongs to extracellular protein ligands. Mutations that impact EGFR expression or function have the potential to cause cancer in a variety of cancer types. Over expression of EGFR is linked to the development of a range of malignancies, while faulty EGFR signaling is linked to disorders including Alzheimer's. The patient condition can be improved by preventing the growth of EGFR tumors. This can be accomplished by either blocking the EGFR receptor sites or by inhibition of its intracellular tyrosine kinase activity [2]. For cancer treatment, EGFR can be used as a promising target for the discovery of novel drugs [3]. As a result, computer-assisted research was used to find effective inhibitors of the EGFR receptors. Plants have been utilized as a significant part of treatment for a variety of disorders all around the world for ages [4]. Many plants have been proven to have potential therapeutic activity without undesired effects. Hence, it has created a strong desire to identify compounds from plants that have been used to treat ailments for centuries. The application of new approaches to an established old discipline may lead to the discovery of effective medications [5]. A plant of the family Ranunculaceae known as *Nigella sativa* is a widely used medicinal plant throughout the globe [6]. *Nigella sativa* seeds have long been used to cure a variety of ailments. To evaluate the potential ability of compounds from *N. sativa* to bind EGFR receptor, we used a computational approach based on Molecular docking, Density Function Theory (DFT), and Molecular Dynamics (MD) simulations. The findings showed that *N. sativa* compounds can bind to the human EGFR receptor's active site residues. Alpha-herederin, in instance, exhibited a projected binding energy that was equivalent to or better than certain existing medicines and interacted well with critical amino acids in the EGFR active site region. Our findings also indicate that several *N. sativa* compounds may be act as effective EGFR receptor inhibitors. We feel compelled to pursue this result further in the development of a possible medication and therapeutics in cancers.

**Methods and Methodology:** We came across *N. sativa* compounds. On the investigated compounds, DFT calculations were performed, as well as geometry optimizations, and Molecular Electrostatic Potential (MESP) map. Following DFT calculations, optimized structures underwent through molecular docking calculations. In the Protein Data Bank (<https://www.rcsb.org/>), we find the crystal structure of EGFR receptor complexes with

its known inhibitor (PDB code 1I8I). Pubchem (<https://pubchem.ncbi.nlm.nih.gov/>) was used to download the 3D structures of active compounds from *N. sativa* as well as known drugs as controls (temozolomide, lapatinib, and procarbazine) that have activity against brain tumors. Auto Dock version 1.5.7 and Maestro (Schrodinger suit) were used to prepare and dock the receptor and *N. sativa* compounds. Docking simulations using Auto Dock were set up according to the instructions in the studies. The protein structures of Maestro (Schrodinger suit) were created using the protein preparation wizard designed for this program, and explicit hydrogen were added. Using its cavity identification algorithm, this application automatically detects possible binding sites (active sites). The conformations of the best poses for each molecule examined were chosen, recorded in.pdb format, and assessed for interactions with the receptor using the Discovery Studio tools. Simulations of molecular dynamics MD simulations lasting 100ns were carried out on a Linux cluster computer running the MD application.

**Results:** The bulk of the chemical compounds identified from *N. sativa* are projected to bind the EGFR receptor with binding energies equivalent to, if not better than, recognized brain tumor treatments. All of the compounds selected from *N. sativa* have a negative binding energy, indicating a potential ability to interact with the receptor (Table 1). According to the analysis, the best-predicted-energy complexes show that these compounds can interact with critical residues involved in substrate binding and catalysis.

**Table 1.** Shows the results of ligand docking simulations in the active site of the EGFR receptor. The representative pose of the cluster of solutions discovered corresponds to the best-predicted binding energy. For procarbazine, temozolomide, and lapatinib, the projected binding energy of the representative posture and the number of poses of the associated cluster are used as a comparative control

Compounds Pubchem ID	Auto dock		Maestro (Schrodinger)	
	No of poses	Binding energy kcal/mol	XP Score	Glide Emodel
Controls for analysis				
Procarbazine(4915)	47	-6.24	-3.676	-39.851
Temozolomide (5394)	18	-5.86	-5.326	-39.384
Lapatinib (208908)	91	-6.76	-2.071	-72.140
<i>Nigella sativa</i> compounds				
Alpha-herederin (73296)	28	-7.27	-6.417	-59.382
Nigellidine (136828302)	99	-7.11	-3.322	-38.204
Thymol (6989)	100	-5.18	-3.832	-24.566
Thymoquinone (10281)	77	-5.17	-4.640	-21.810
Thymohydroquinone (95779)	91	-4.88	-4.640	-28.082

**Conclusion:** The goal of this study was to find novel EGFR receptor inhibitors from *N. sativa* that will bind to the receptor's active site selectively. Alpha-herederin was discovered to be one of the chemical compound with the potential to bind to the EGFR receptor. The docking method also identified nigellidine as a possible receptor ligand. MD simulations corroborated the docking research that nigellidine may bind stably to the receptor's active site, while the result for Alpha-herederin did not match docking results completely. The ability of nigellidine and  $\alpha$ -herederin to persist in the active site of the EGFR receptor suggests that modifying this compound to improve its specificity and affinity could lead to the identification of a new compound of EGFR inhibitors. As a result, we believe that this research will serve as a first step toward the creation of new EGFR inhibitors for the treatment of brain cancers.



## References

1. Ferlay J., Ervik M., Lam F., Colombet M., Mery L., Piñeros M. et al. Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer, 2020. (<https://gco.iarc.fr/today>). Accessed February 2021.
2. Sinto S., Settleman J., Reshkin S.J., Azzariti A., Bellizzi A., Paradiso A. The complexity of targeting EGFR signalling in cancer: From expression to turnover. *Biochim Biophys Acta Bioenerg.* 2006;1766:120-39. doi: 10.1016/j.bbcan.2006.06.001.
3. He K., Xu J., Liang J., Jiang J., Tang M., Ye X., Zhang Z., Zhang L., Fu B., Li Y., Bai C., Zhang L., Tao W. Discovery of a novel egfr-targeting antibody-drug conjugate, SHR-A1307, for the treatment of solid tumors resistant or refractory to anti-egfr therapies. *Mol Cancer Ther.* 2019;18(6):1104-1114.
4. Ahmad A., Husain A., Mujeeb M., Khan S.A., Najmi A.K., Siddique N.A., Damanhoury Z.A., Anwar F. A review on therapeutic potential of *Nigella sativa*: A miracle herb. *Asian Pac J Trop Biomed.* 2013;3(5):337-52.
5. Dubey A., Dotolo S., Ramteke P.W., Facchiano A., Marabotti A. Searching for chymase inhibitors among chamomile compounds using a computational-based approach. *Biomolecules.* 2018;9(1):5. doi: 10.3390/biom9010005.
6. Ahmad A., Husain A., Mujeeb M., Khan S.A., Najmi A.K., Siddique N.A., Damanhoury Z.A., Anwar F. A review on therapeutic potential of *Nigella sativa*: A miracle herb. *Asian Pac J Trop Med.* 2013;3(5):337-352. doi: 10.1016/S2221-1691(13)60075-1.

# Exploration of natural flavonoids as novel Topoisomerase II inhibitors using DFT, molecular docking, and molecular dynamics

Tufail A.<sup>1\*</sup>, Dubey A.<sup>2,3\*</sup>

<sup>1</sup> Department of Biochemistry, University of Allahabad, Allahabad, India

<sup>2</sup> Computational Chemistry and Drug Discovery Division, Quanta Calculus, Kushinagar, India

<sup>3</sup> Department of Pharmacology, Saveetha Dental College and Hospital, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

\* aishatufailansari@gmail.com, ameebioinfo@gmail.com

**Key words:** topoisomerase II, DFT, molecular docking, molecular dynamics

*Motivation and Aim:* Topoisomerase II are seen as an important therapeutic target for cancer therapy now a days [1] as they are found in increased amount in cancerous cells. Topoisomerase II inhibitors are used as chemotherapeutic agents due to their ability to induce tumour cell death [2]. Role of p53 is important in controlling the cell cycle and apoptosis. As already known, p53 is unable to express in cancerous cells and is found in large amount as mutant p53. MDM2 protein is a product of one of the responsive genes of p53. As stated by the recent paper [7], the MDM2 protein, binds to the p53 protein directly through their amino termini and the function of p53 is inhibited through three major mechanisms: (1) On binding, MDM2 directly ubiquitinates p53 through its E3 ligase activity, which therefore promotes proteasomal degradation of p53; (2) p53 and MDM2 interaction blocks the p53 from binding to its targeted DNA, and hence blocking p53's transcriptional activity; and (3) export of p53 is promoted by MDM2 out of the cell nucleus, making p53 inaccessible to the targeted DNA and further its transcriptional ability is reduced. This provides a p53-MDM2 auto regulatory feedback loop. Non-Functional p53 leads to development as well as progression of tumor cells. A study shows the effect of Topoisomerase II inhibitors on the auto regulatory feedback loop of p53-MDM2 complex, which suggests an involvement of Topoisomerase II in p53 regulation [3]. This is done by activating the G1 check-point and p53-p21CIP1/WAF1 pathway [4]. One of the reasons for cancer progression is mutations in human cancer (more than 50 %) along with increased expression of MDM2 [5]. This has led to numerous evaluations and investigations in its role and potential as a therapeutic target in the sense of restoring wild type p53. As seen by the already known drugs of topoisomerase II inhibitors (e.g. doxorubicin), cytotoxic concentrations are needed for inhibiting the p53-MDM2 auto regulatory feedback loop [6]. In this study, we are targeting multiple pathway for killing cancerous cells using natural compounds. This is done by topoisomerase II inhibition along with activation of p53 by providing optimum drug concentration which would be less toxic.

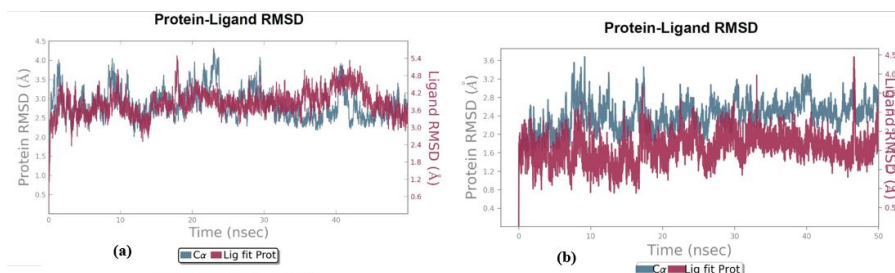
*Methods and Methodology:* A library of natural compounds was screened *in-silico* for determining the interaction of each compound with p53-MDM2 complex (1T4F) and Topoisomerase II (4FM9) through molecular docking experiments. The already known drugs, Doxorubicin and Mitoxantrone (Topoisomerase II inhibitors) were taken as control for validation. Density Function Theory (DFT) and Electrostatic Potential

(MESP) map calculations were also performed. Following the molecular docking experiments, best protein-ligand complexes were also assessed for 100ns long molecular dynamics Simulations to record the Root means Square deviation (RMSD) between protein and ligand. RMSF were also calculated to see the variations in protein fluctuations.

**Results:** Each natural compound thus shortlisted, showed almost equal binding energy with both p53-MDM2 complex and Topoisomerase II (Table 1). Molecular dynamics results also shows the stability of None2 compound as compared to control Compound K (Fig. 1). Hence, indicating a possibility that same drug concentration can be given to activate p53 as needed for the inhibition of Topoisomerase II.

**Table 1.** Results of Molecular Docking Simulations of ligands into the active site of Topoisomerase II and p53-MDM2 complex

Compounds PubChem ID	Binding energy with topoisomerase II (4FM9)				Binding energy with p53+MIDM2 complex (1T4F)			
	Autodock		Glide		Autodock		Glide	
	best predicted binding energy, kcal/mol	number of poses in cluster with best predicted binding energy	score	Emodel	best predicted binding energy, kcal/mol	number of poses in cluster with best predicted binding energy	score	Emodel
Doxorubicin* (31703)	-6.55	2	-6.942	-77.790	-5.61	1	-3.417	-36.789
Mitoxantrone* (4214)	-5.58	1	-6.028	-62.840	-3.89	1	-4.259	-51.713
None 2	-6.02	4	-5.094	-40.652	-5.02	26	-5.867	-67.716
Acetovanillone (2214)	-4.69	2	-5.100	-32.137	-4.04	13	-4.56	-28.363
Vanillic Acid (8468)	-4.31	8	-3.995	-24.447	-4.49	21	-4.688	-23.762
Compound K (5481990)	-8.34	1			-6.18	1	-5.669	-60.991



**Fig. 1.** Evolution of the Root mean square deviation (RMSD) during MD simulations of Topoisomerase II in complex with the 2 different ligands: (a) Compound K (b) None 2. On the left axis, the protein RMSD value (represented as the blue trace in the graph) is shown, whereas on the right axis, the ligand RMSD value (represented as the red trace in the graph) is shown. Values are reported in Å. The graphs were produced by Desmond analysis tools

**Conclusion:** This can probably be used as an effective therapeutic approach for curing cancers as per the strategy of targeting two pathways with the same drug concentration, unlike before where more cytotoxic concentration was needed for activating p53 apoptosis pathway than for the inhibition of Topoisomerase II. As a result, lower side effects would be seen because of drugs made of natural compounds.

## References

1. Malina J., Vrana O., Brabec V. Mechanistic studies of the modulation of cleavage activity of topoisomerase I by DNA adducts of mono- and bi-functional PtII complexes. *Nucleic Acids Res.* 2000;37(16):5432-5442.
2. Thomas A., Perry T., Berhane S., Oldreive C., Zlatanou A., Williams L.R., Weston V.J., Stankovic T., Kearns P., Pors K., Grand R.J., Stewart G.S. The dual-acting chemotherapeutic agent Alchemix induces cell death independently of ATM and p53. *Oncogene.* 2015;34:3336-3348.
3. Valkov N.I., Sullivan D.M. Tumor p53 status and response to topoisomerase II inhibitors. *Drug Resist Updat.* 2003;6(1):27-39. doi: 10.1016/s1368-7646(02)00143-7.
4. Siu W.Y., Lau A., Arooz T., Chow J.P., Ho H.T., Poon R.Y. Topoisomerase poisons differentially activate DNA damage checkpoints through ataxia-telangiectasia mutated-dependent and -independent mechanisms. *Mol Cancer Ther.* 2003;3(5):621-632.
5. Gupta A., Shah K., Oza M.J., Behl T. Reactivation of p53 gene by MDM2 inhibitors: A novel therapy for cancer treatment. *Biomed Pharmacother.* 2019;109:484-492.
6. Arriola E.L., Lopez A.R., Chresta C.M. Differential regulation of p21waf-1/cip-1 and Mdm2 by etoposide: etoposide inhibits the p53-Mdm2 autoregulatory feedback loop. *Oncogene.* 1999;18(4):1081-1091.
7. Wang S., Zhao Y., Aguilar A., Bernard D., Yang C.Y. Targeting the MDM2-p53 protein-protein interaction for new cancer therapy: Progress and challenges. *Cold Spring Harb Perspect Med.* 2017;7(5):a026245.

# Nanoteranostics based on multimodal nanoparticles and nanosystems

Zavestovskaya I.

*P.N. Lebedev Physical Institute of RAS, Moscow, Russia*

*National Research Nuclear University MEPhI, Moscow, Russia*

*zavestovskayain@lebedev.ru*

**Key words:** nanoparticles, nanotheranostics, nuclear nanomedicine, radiation therapy, radiosensitizers

*Motivation and Aim:* Cancer is one of major problems of modern society counting about 15–16 % of all death cases per year and the death toll from cancer continues to rise. Nanotheranostics based on multimodal nanoparticles (NPs) have emerged as one of promising techniques for cancer treatment, which can offer both imaging and therapy modalities [1]. We will highlight our recent achievements in the development of nanotechnologies for nuclear biomedicine, including nuclear medicine and radiology and some important trends in the development of this technology.

Laser ablation has appeared as a new non-chemical pathway for the synthesis of nanomaterials, which is free of limitations of conventional chemical approaches and makes possible the synthesis of ultrapure nanostructures. In this approach, small nanoscale clusters are naturally formed during laser ablation from a solid target, and then released into a liquid ambient to form a colloidal nanoparticle solution. Laser ablation can provide nanomaterials exhibiting unique properties and functionalities.

The properties of NPs as an increased ratio of surface area to volume, the ability of passive/active guidance and high load capacity, a large cross-section of interaction with biological tissues, unique properties of the surface of nanomaterials, easy giving of many functions to nanomaterials, etc. are used. All these properties of laser-synthesized nanomaterials promise their successful employment in theranostics.

We propose silicon (Si) NPs (Si\*NPs) synthesized by promising laser-based approaches as a nearly ideal NPs for nanotheranostics [2]. One can use these methods to make stable colloidal dispersions of silicon nanoparticles in both organic and aqueous media, which are suitable for a multitude of applications across the important fields of health care. Size tailoring allows production of Si\*NPs with efficient photoluminescence that can be tuned across a broad spectral range from the visible to near-IR by varying particle size and surface functionalization. These applications encompass several types of bioimaging and various therapies, including phototherapy, RF thermal therapy, and radiotherapy [3]. The uniqueness of such Si\*NPs is based on their biodegradability, which makes possible rapid elimination of these structures from the organism within several days even if their initial size is large (30–80 nm) under absence of any toxic effects, which was earlier confirmed in a mice model. In addition, in contrast to Si nanostructures prepared by conventional chemical or electrochemical routes, laser-synthesized Si\*NPs have ideal round shape, controllable size with low size dispersion, and are free of any toxic impurities, which promises a better transport *in vivo* and the absence of side effects.

Synthesized nanoparticles were tested as carriers for promising radionuclides (Re-188, Ga-68) in nuclear medicine, as well as sensitizers in radiation therapy. We demonstrate the possibility for fast PEGylation and conjugation of laser-synthesized Si\*NPs with

Rhenium-188 (<sup>188</sup>Re) radionuclide, which is one of most promising generator-type therapeutic beta-emitters with the energy of positron emission of 1.96 MeV (16.7 %) and 2.18 MeV (80 %) and half-decay time of 17 hours [4]. We show that such conjugates can efficiently deliver the radionuclide through the blood stream and retain it in the tumor region. We also show that Si NPs ensure excellent retention of <sup>188</sup>Re in tumor, not possible with the salt, which enables one to maximize therapeutic effect, as well as a complete time-delayed conjugate bioelimination. Finally, our tests on rat survival demonstrate excellent therapeutic effect (72 % survival compared to 0 % of the control group). Combined with a series of imaging and therapeutic functionalities based on unique intrinsic properties of Si\*NPs, the proposed biodegradable complex promises a major advancement of nuclear nanomedicine.

Technologies of combined action of various types of radiation (protons-neutrons, protons-carbon ions, multi-ion therapy); targeted proton therapy technologies using promising nanoparticles and systems based on them as therapy sensitizers and active agents for diagnostics. The latter direction involves a significant expansion of the field of modern nuclear medicine through integration with nanotheranostics, which uses nanoparticles for the diagnosis and therapy of cancer, using their unique properties. The introduction of non-radioactive materials that can be activated from the outside using various external sources of nuclear particles to produce radioactivity in situ is one of the new directions of activation of nano-drugs at the site of a cancerous tumor, which can be considered as in situ production of radiopharmaceuticals. Such binary radiotherapy technologies become especially efficient when one can achieve a high tumor/non-tumor action contrast, which enables to minimize side effects related to the irradiation of healthy issues.

*Conclusion:* Our recent data show that the field of nuclear medicine can be significantly expanded by integrating with nanomedicine, which utilizes nanoparticles (NPs) as carriers of radionuclides or as radiosensitizers for radiation therapy and/or active agents for imaging (radiopharmaceutical medicines in situ). The synergy of nanotechnology with nuclear medicine opens up a new direction of cancer imaging and therapy – nuclear nanotheranostics.

*Acknowledgements:* The study is supported by the Ministry of Science and Higher Education of RF (project No. 075-15-2021-1347), RFBR (project No. 20-20-00861).

### References

1. Roy I., Krishnan S., Kabashin A., Zvestovskaya I., Prasad P. Transforming Nuclear medicine with nanoradiopharmaceuticals. Review. *ACS Nano*. 2021. doi: 10.1021/acsnano.1c10550. Online ahead of print.
2. Kabashin A.V., Singh A., Swihart M.T., Zvestovskaya I.N., Prasad P.N. Laser-processed nanosilicon: a multifunctional nanomaterial for energy and healthcare. Review. *ACS Nano*. 2019;13(9):13(9):9841-9867. doi: 10.1021/acsnano.9b04610.
3. Oleshchenko V.A., Kharin A.Yu., Alykova A.F., ..., Timoshenko V., Zvestovskaya I.N. Localized infrared radiation-induced hyperthermia sensitized by laser-ablated silicon nanoparticles for phototherapy applications. *Appl Surf Sci*. 2020;516:145661.
4. Petriev V.M. et al. Nuclear nanomedicine using si nanoparticles as safe and effective carriers of <sup>188</sup>Re radionuclide for cancer therapy. *Sci Rep*. 2019;9:2117.

## Modeling of mammalian ribonuclease inhibitor complex with bacterial ribonuclease binase

Zheleznyak A.R.<sup>1,2</sup>, Kechko O.I.<sup>1</sup>, Anashkina A.A.<sup>1\*</sup>, Petrushanko I.Yu.<sup>1</sup>, Mitkevich V.A.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> I.M.Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

\* anastasia.a.anashkina@gmail.com

**Key words:** binase, ribonuclease inhibitor, RNase A

*Motivation and Aim:* Ribonucleases appear to be a promising replacement for conventional chemotherapy [1] because the side effects of systemic chemotherapy used to treat cancer and directly target DNA are often severe [2]. RNases play a key role in RNA metabolism [3]. One of the promising representatives of therapeutic RNases is bacterial ribonuclease binase, which is not inhibited by the mammalian ribonuclease inhibitor (RI) and selectively eliminates a number of malignant cells. At the same time, pancreatic RNase A is inhibited with high efficiency by RI and does not have a cytotoxic effect on malignant cells [4]. It was postulated that the cytotoxic effect of bacterial RNases is due to the fact that they do not interact with RI. However, this was not verified. The purpose of this work was to test the possibility of interaction between binase and RI.

*Methods and Algorithms:* molecular modeling of binase complexes with RI was performed using the MOE program, protein-protein docking was performed by the ZDock server [5], and model analysis was performed by the Qasdom server [6]. The interaction of binase with RI in solution was evaluated by microscale thermophoresis (MST). Analysis was performed using the Monolith NT.115 (NanoTemper technology) and data analysis was performed using MO.Affinity Analysis software v.2.3.

*Results:* An experimental verification of the interaction of binase with RI, carried out in solution by the MST method, showed that they form a stable complex with a dissociation constant in the nanomolar range. A model structure of the complex of RI with binase was constructed for both reduced form of RI and oxidized one. It was found that in both cases the binase active site in such a complex is oriented away from RI and remains available for ribonuclease activity, while the RNase A catalytic site is blocked. Significant amino acids of the binding site of binase with barstar (a natural inhibitor of binase) – Gln28, Arg58, Arg82, Arg86 [7] – as well as the active site of binase – Lys26, Asp53, Val54, Phe55, Ser56, Asn57, Arg58, Glu59, Glu72, Arg82, Arg86, His101, His102 [8] do not interact with RI in our model. Energies of binase complex with the oxidized and reduced forms of the ribonuclease inhibitor were –575 and –617 kcal/mol respectively.

*Conclusion:* For the first time, it was found that bacterial RNase binase forms a complex with RI. In contrast to RNase A, the active site of binase in complex with RI is not blocked, and binase can degrade the substrate, realizing its cytotoxic effect.

*Acknowledgements:* This research was supported by the grant ICGEB No. CRP/RUS20-01.

## References

1. Makarov A.A., Ilinskaya O.N. Cytotoxic ribonucleases: molecular weapons and their targets. *FEBS Lett.* 2003;540:15-20. doi: 10.1016/S0014-5793(03)00225-4.
2. Schirmmacher V. From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). *Int J Oncol.* 2019;54:407-419. doi: 10.3892/ijo.2018.4661.
3. Deutscher M.P., Li Z. Exoribonucleases and their multiple roles in RNA metabolism. *Prog Nucleic Acid Res Mol Biol.* 2001;66:67-105. doi: 10.1016/S0079-6603(00)66027-0.
4. Blackburn P., Jailkhani B.L. Ribonuclease inhibitor from human placenta: interaction with derivatives of ribonuclease A. *J Biol Chem.* 1979;254:12488-12493.
5. 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;30:1771-1773. doi: 10.1093/bioinformatics/btu097.
6. Anashkina A.A., Kravatsky Y., Kuznetsov E., Makarov A.A., Adzhubei A.A. Meta-server for automatic analysis, scoring and ranking of docking models. *Bioinformatics.* 2018;34:297-299. doi: 10.1093/bioinformatics/btx591.
7. Wang L., Pang Y., Holder T., Brender J.R., Kurochkin A.V., Zuiderweg E.R.P. Functional dynamics in the active site of the ribonuclease binase. *Proc Natl Acad Sci.* 2001;98:7684-7689. doi: 10.1073/pnas.121069998.
8. Ermakova E.A. A comparative study of the interaction of *Bacillus amyloliquefaciens* ribonuclease (barnase) and *Bacillus intermedius* ribonuclease (binase) with barstar by brownian dynamics simulation. *Biophysics.* 2006;51:202-208. doi: 10.1134/S0006350906020072.



# 4 Symposium “Evolutionary, population and medical genomics/ genetics of human”



- 4.1 Section “Human medical/  
population genomics and genetics” [377](#)
- 4.2 Section “Genome-wide association  
studies” [463](#)
- 4.3 Section “Human origin  
and evolution” [479](#)

## Searching genomic signals of natural selection and local adaptation in populations of Russia

Aliev A.M., Khvorykh G.V., Khrunin A.V.\*

*Institute of Molecular Genetics of National Research Center "Kurchatov Institute", Moscow, Russia*

\* [khrunin@img.msk.ru](mailto:khrunin@img.msk.ru)

**Key words:** DNA polymorphisms, signals of natural selection, human populations

The identification of genomic loci that are subjected to natural selection provides both deeper insights into the existence of population genetic structures and understanding the genetic bases of population-level differences in the distribution of common diseases and traits. In our study we searched the signatures of positive selection in the genome-wide data of 10 populations from Russia that, until recently, remained a white spot on the genomic map of human local adaptations. To detect signals of selection, we used the population branch statistic (PBS), which was calculated for sliding windows of 20 SNPs with a step size of 5 SNPs. The chromosome regions containing windows with PBS values above the 99.9th percentile of the empirical distribution were candidate regions and were further annotated. Totally, 936 genes were annotated (72 to 169 genes per population). The functional evaluation of the genes showed that many of them were significantly overrepresented in particular biological processes (metabolic pathways). Some of these pathways were identified across multiple populations – groups of populations, united by their history of origin or geography of residence: 1) the processes of alcohol metabolism, retinol and retinoic acid associated with ADH1A, ADH1B and ADH1C in Veps, Komi Priluzski, Khanty, Mansi and Nenets, 2) the metabolism of unsaturated fatty acids associated with FADS1, FADS2 and FADS3 in Khanty, Mansi, Nenets and Yakuts, and 3) the metabolism of thyroid hormones associated with DUOX1, DUOX2, DUOXA1 and DUOXA2 in Nenets and Yakuts. Others were population-specific: glucuronization processes (Veps), retinoic acid metabolism by cytochromes (northern Russians), cornification (Komi Izhemski), etc. At the same time, the vast majority of the genes were not associated with any pathways and processes. This suggests the complexity of adaptation processes, which in each specific case could be going or supplemented by adjusting individual elements of a variety of metabolic pathways to the characteristics of the habitat.

*Acknowledgements:* The study was funded by Russian Foundation for Basic Research (grant No. 20-04-00824).

## Alleles and haplotypes diversity of HLA-A, -B, -C, -DRB1, and -DQB1 genes in Russia

Ananeva A.<sup>1\*</sup>, Ryzhkova T.V.<sup>2</sup>, Gaifullina R.<sup>3</sup>, Gusev O.<sup>1</sup>, Shagimardanova E.<sup>1</sup>

<sup>1</sup> *Extreme Biology Laboratory, Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>2</sup> *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>3</sup> *National Bone Marrow Donor Registry, Moscow, Russia*

<sup>4</sup> *Volga-Region Bone Marrow Donor Registry, Kazan, Russia*

\* *nasty.aananeva@gmail.com*

**Key words:** HLA, novel allele, diversity, haplotype

*Motivation and Aim:* HLA is well known as a highly polymorphic genetic system. HLA genes are in the MHC (major histocompatibility complex) and the frequency of alleles and haplotypes varies widely among human populations. The aim of our study was to characterize allele frequency distribution of the HLA-A, HLA-B, HLA-C, HLA-DQB1, HLA-DRB1 loci of the Russian population from different regions of Russia.

*Methods and Algorithms:* Blood samples from 5492 adult potential donors of hematopoietic stem cells recruited by National bone marrow donor registry named after Vasya Perevoshchikov were analyzed. We implemented an NGS-based workflow for the full length typing of HLA-A, -B, -C, -DRB1 and -DQB1 using Holotype HLA™ (Omixon, Hungary). Sequencing was performed using the MiSeq reagent kit v.2 through MiSeq instrument (Illumina, USA). Analysis of resulting FastQ data files was carried by HLA Twin™ software (Omixon, Hungary). HLA allele and haplotype frequencies were estimated via maximum-likelihood analysis from genotypic data through an expectation-maximization (EM) algorithm for unknown gametic phase implemented in the Arlequin program version 3.5.2.2.

*Results:* A total of 72 HLA-A, 122 HLA-B, 67 HLA-C, 34 HLA-DQB1, 89 HLA-DRB1 alleles were identified. The most frequent among the HLA-A were A\*02:01:01 (26,7 %), A\*03:01:01 (13,5 %), A\*01:01:01 (11,2 %) and A\*24:02:01 (11,2 %). Among the HLA-B locus, the most frequent variants were B\*07:02:01 (11,3 %), B\*13:02:01 (7,1 %), B\*35:01:01 (6,4 %), B\*18:01:01 (6,4 %), B\*08:01:01 (6,3 %). Among the HLA-C locus C\*07:02:01 (13,1 %), C\*04:01:01 (13 %), C\*06:02:01 (12,8 %), C\*07:01:01 (11 %), C\*12:03:01 (9,3 %). The most frequent allele of HLA-DQB1 was defined as DQB1\*03:01:01 (19.1 %), other variants were DQB1\*02:02:01 (14,7 %), DQB1\*05:01:01 (14,1 %), DQB1\*06:02:01 (10,9 %) and finally among the DRB1 locus - DRB1\*07:01:01 (14,8 %), DRB1\*01:01:01 (12 %) and DRB1\*15:01:01 (11,7 %). There were a trend for higher frequency of HLA haplotype A\*01:01:01~B\*08:01:01~C\*07:01:01~DQB1\*02:01:01~DRB1\*03:01:01, determined for 2,38 % and haplotype A\*02:01:01~B\*13:02:01~C\*06:02:01~DQB1\*02:01:01~DRB1\*07:01:01 determined for 1.05 % of donors.

Among 5,492 potential donors we identified 18 rare alleles and 17 very rare alleles. Also, we identified 16 novel alleles: 6 HLA-A, 7 HLA-B, 8 HLA-C, 4 HLA-DQB1, 3 HLA-DRB1. The names A\*02:974, C\*06:327, A\*68:01:58, C\*04:441:01:02, A\*02:01:204,

C\*12:349, A\*03:444, B\*39:189, B\*07:458, B\*44:348, A\*03:390:02, B\*27:05:57, B\*08:295, C\*05:269, A\*02:1042, C\*01:230 has been officially assigned by the WHO Nomenclature Committee. We also identified 12 potentially novel alleles among this group, which were submitted to the WHO Nomenclature Committee, but official names have not been assigned yet. This follows the agreed policy that, subject to the conditions stated in the most recent Nomenclature Report (Marsh et al. 2010), names will be assigned to new sequences as they are identified. Lists of such new names will be published in the following WHO Nomenclature Report.

*Conclusion:* The most frequent alleles and haplotype of Russian population are typical for European populations. In the cohort analyzed 0.64 % of potential donors can be assumed as bearing rare or very rare HLA alleles. 28 novel and potentially novel alleles have been described among 5492 potential donors. Analysis of the HLA haplotypes frequencies in the population is necessary to assess the probability of detecting a matching donor. Also, such studies can be used for population genetic and HLA-disease association fields.

*Acknowledgements:* We would like to thank the Charity fund “National Bone Marrow Donors Registry named after Vasya Perevoshchikov” for support of the HLA typing and for the donor recruiting. This study was the part of the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

### *References*

1. Marsh S.G.E., Albert E.D., Bodmer W.F., Bontrop R.E., Dupont B., Erlich H.A., Fernández-Viña M., Geraghty D.E., Holdsworth R., Hurley C.K., Lau M., Lee K.W., Mach B., Maiers M., Mayr W.R., Müller C.R., Parham P., Petersdorf E.W., Sasazuki T., Strominger J.L., Svejgaard A., Terasaki P.I., Tiercy J.M., Trowsdale J. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010;75(4):291-455.

## Re-analysis of DNA methylation data from tuberculosis-infected individuals with a focus on target genes

Babushkina N.P. \*, Bragina E.Yu., Goncharova I.A., Markov A.V., Kucher A.N., Nazarenko M.S.

Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia

\* [nad.babushkina@medgenetics.ru](mailto:nad.babushkina@medgenetics.ru)

**Key words:** tuberculosis, methylation

*Motivation and Aim:* Along with the structural variability of the genome, epigenetic modifications can be significant for susceptibility to various, including infectious diseases. The aim of this study was to re-analyze the differential methylation of genes (DMG), for which associations with tuberculosis and its pathogenetically significant features were previously established, as well as for those of interest from the point of view of studying the phenomenon of reverse comorbidity of asthma and tuberculosis.

*Methods and Algorithms:* Re-analysis of two DMG datasets presented in the public repository of functional genomic data Gene Expression Omnibus [1] was carried out. The GSE72338 dataset [2] was obtained from the study of DNA methylome, transcriptome (mRNA and microRNA), and proteome. The study was performed on peripheral blood monocytes (PBMC) and neutrophils of patients with active tuberculosis (TB,  $n = 8$ ) and clinically healthy individuals, with household contact with active TB patients (LTBI,  $n = 8$ ). The patients lived in an area with a high incidence TB in Cape Town, South Africa [2]. The second dataset, GSE118469 [3], was obtained from a study of TB in Taiwanese residents. The study included patients with TB ( $n = 12$ ) without concomitant infections (including HIV) and malignant neoplasms. The control group comprised of healthy individuals ( $n = 6$ ). Methylation analysis was performed on an Infinium HumanMethylation450K BeadChip v1.2 chip [3].

In the GSE72338 and GSE118469 datasets we were primarily interested in genes that, according to our previously data were associated with infectious diseases, in particular thirteen genes *IFNG*, *SOCS5*, *TNFB*, *TNFRSF1B*, *PIAS3*, *PIASY*, *CXCL10*, *ATM*, *NBN*, *MRE11*, *MLH1*, *PMS2*, and *TP53BP1*.

*Results:* Only the *SOCS5* gene (out of 13 genes) included in the Top-50 DMG list, topped the DMG list when comparing neutrophils from TB and LTBI patients ( $p = 0.00000226$ ), but adjusting for multiple comparisons makes these differences not statistically significant ( $\text{adj.}p = 0.595$ ). It should be noted that, taking into account the FDR correction no significant differences were obtained in any of the comparisons. Further, we carried out the analysis, focusing on the initial level of significance achieved, without introducing any corrections.

In monocytes, 13 sites in seven target genes (*NBN*, *MRE11A*, *MLH1*, *IFNG*, *LTA*, *TNFRSF1B*, *CXCL10*) were differentially methylated in patients with active TB compared with LTBI. In neutrophils, 7 sites in regions of genes *ATM*, *MLH1*, *PMS2*, *SOCS5*, *TNFRSF1B*, *CXCL10* were differentially methylated in patients with active TB compared with LTBI. 23 sites in *ATM*, *PMS2*, *TP53BP1*, *SOCS5*, *TNFRSF1B*, *CXCL10* genes were differentially methylated in PBMC when comparing between TB patients and healthy individuals.

Then, we annotated these loci with the UCSC resource. In two datasets, 39 CpG-sites were identified in target genes that were differentially methylated between TB patients and healthy controls. The position of three CpG-sites in the genome allows them to be attributed to the sphere of influence of two genes: cg11588932 (*ATM*, *NPAT*), cg13846866 (*MLH1*, *EPM2AIP1*), cg23492613 (*MRE11A*, *ANKRD49*). There are 32 sites in the gene body (targeted or co-localized) (4 in exons and 28 in introns); 7 sites are located in the intergenic space near the promoter region of the gene.

The methylation level of most of the studied sites (> 200 or even > 600) indicates their low transcriptional activity. This is supported by a small number (8 CpG) sites localized in the RNA polymerase II subunit A binding region and in open heterochromatin (12 CpG sites). The CpG-islands include 5 sites, and 21 sites are located near the islands, 12 sites are single. At that, 26 sites are localized near the promoters of target (and co-localized) genes; only cg21288207 is included directly in the promoter of the target gene (*PMS2*). In addition, 2 CpG-sites are included in the co-localized genes promoter sequence (cg23492613: *ANKRD49*; cg16280132: *ATP6VIG2*). Only 9 CpG-sites are located in transcription factor binding regions. However, cg24682307 (*TNFRSF1B*) is located in the IRF1 binding region, which activates the transcription of genes involved in the response to viruses and bacteria [4]; cg21288207 and cg19570749 (*PMS2*) in the ELF1 and GATA3 binding regions respectively, that regulate gene expression during the development of T-lymphocytes [5, 6].

**Conclusion:** Thus, these findings indicated that DNA methylation is important in TB development. The differential value of DNA methylation of 13 target genes associated with TB, as well as the localization of these sites in the binding regions of transcription factors involved in the development of the immune response to infection, confirm the involvement of protein products of target genes in the pathogenesis of TB.

**Acknowledgements:** The work was carried out within the framework of the State Task of the Ministry of Science and Higher Education No. 122020300041-7.

### References

1. GEO, <https://www.ncbi.nlm.nih.gov/geo>.
2. Esterhuyse M.M. et al. Epigenetics and proteomics join transcriptomics in the quest for tuberculosis biomarkers. *mBio*. 2015;6(5):e01187-e01115. doi: 10.1128/mBio.01187-15.
3. 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.
4. Taki S. et al. Multistage regulation of Th1-type immune responses by the transcription factor IRF-1. *Immunity*. 1997;6(6):673-679. doi: 10.1016/s1074-7613(00)80443-4.
5. Leiden J.M. et al. A novel Ets-related transcription factor, Elf-1, binds to human immunodeficiency virus type 2 regulatory elements that are required for inducible trans activation in T cells. *J Virol*. 1992;66(10):5890-5897. doi: 10.1128/JVI.66.10.5890-5897.1992.
6. Wan Y.Y. GATA3: a master of many trades in immune regulation. *Trends Immunol*. 2014;35(6):233-242. doi: 10.1016/j.it.2014.04.002.

## Re-analysis of DNA methylation in genes associated with bronchial asthma

Babushkina N.P. \*, Bragina E.Yu., Goncharova I.A., Markov A.V., Kucher A.N., Nazarenko M.S.

Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia

\* [nad.babushkina@medgenetics.ru](mailto:nad.babushkina@medgenetics.ru)

**Key words:** bronchial asthma, methylation, re-analysis

**Motivation and Aim:** To study the variability of gene methylation patterns, which in our previous studies showed associations with bronchial asthma and tuberculosis (which are dystrophic diseases), we re-analyzed two datasets GSE104471 and GSE59339 from the public repository of functional genomic data Gene Expression Omnibus (GEO) [1]. The goal of the study was to compare DNA methylation level associated with bronchial asthma genes in different groups.

**Methods and Algorithms:** The GSE104471 dataset was received by I.V. Yang et al. (2017) comparing methylation profiles (Illumina 450k) in peripheral blood mononuclear cells (PBMC), nasal and bronchial epithelium in 12 patients with allergic bronchial asthma and 12 healthy individuals without asthma: all white non-smoking adults are non-Hispanic [2]. The GSE59339 dataset was received by L.P. Gunawardhana et al. (2014) at studying the profile of DNA methylation (Illumina Infinium Methylation27) in blood monocytes in adult patients with eosinophilic asthma (EA;  $n = 21$ ), paucigranulocytic asthma (PGA;  $n = 22$ ), neutrophilic asthma (NA;  $n = 9$ ) and healthy individuals ( $n = 10$ ). Over half (67 %) of NA patients were ex-smokers [3].

The analyzed target genes are divided into 3 subgroups: 1) associated with asthma/tuberculosis (13 genes (*IFNG*, *SOCS5*, *TNFB*, *TNFRSF1B*, *PIAS3*, *PIASY*, *CXCL10*, *ATM*, *NBN*, *MRE11*, *MLH1*, *PMS2*, *TP53BP1*), were considered in all groups comparison of GSE104471 and GSE59339 datasets); 2) interesting from the point of view of atopy (14 genes (*TNF*, *IL13*, *IL4*, *IL4R*, *TGFB1*, *MS4A2*, *HLA-DRB1*, *HLA-DQB1*, *CD14*, *LTC4S*, *IL10*, *TLR2*, *CTLA4*, *HLA-DQA1*), were considered in all groups comparisons of the GSE104471 data set and in the comparison groups of the GSE59339 set, including eosinophilic asthma); 3) interesting from the point of view of syntropy of asthma and diseases of the cardiovascular system (7 genes (*ANG*, *RNASE4*, *LOC105376244*, *TLR4*, *AL160272.2*, *ABTB2*, *CAT*) were considered in the comparison groups of the GSE59339 set, including neutrophilic and paucigranulocytic asthma).

**Results:** Only *MRE11* out of the target genes, was included in the list of Top 50 differentially methylated genes (DMG) when comparing eosinophilic and paucigranulocytic asthma ( $p = 0.00028$ ), however, the introduction of a correction for multiple comparisons makes these differences statistically insignificant (adj.  $p = 0,49$ ). Significant differences in the level of DNA methylation of target genes, taking into account the FDR correction, were obtained only when comparing methylation between asthma patients and healthy individuals – in both nasal and bronchial epithelial cells. In bronchial epithelial cells, for 8 genes (*MRE11*, *MLH1*, *TP53BP1*, *TNF*, *TGFB1*, *HLA-DRB1*, *HLA-DQB1*, *CD14*), which were analyzed for this pair of comparisons using the GSE104471 dataset, there is information on the level of methylation of 175 CpG-sites,

differential methylation was shown with  $p < 0.05$  – for 41, with  $p\text{FDR} < 0.05$  – for 6 CpG-sites: 3 in the *MLH1* gene, one each in the *TNF*, *TGFB1*, and *HLA-DRB1* genes. In nasal epithelial cells, information on differences in methylation levels is available for 103 CpG-sites in 11 genes (*ATM*, *TP53BP1*, *IFNG*, *LTA*, *TNFRSF1B*, *PIAS3*, *IL4R*, *TGFB1*, *CD14*, *LTC4S*, *TLR2*), differential methylation is shown with  $p < 0,05$  – for 39 sites in these genes, with  $p\text{FDR} < 0.05$  – for 18.

In PBMC, it was possible to analyze the information on 228 CpG-sites in 13 genes (*ATM*, *MLH1*, *MRE11*, *PMS2*, *TP53BP1*, *IL13*, *IL4*, *TGFB1*, *HLA-DRB1*, *HLA-DQB1*, *TLR2*, *CTLA4*, *HLA-DQA1*). Differential methylation is recorded at  $p < 0.05$  – for 18 sites in the 11 genes studied (except for *IL13* and *IL4*), with  $p\text{FDR} < 0.05$  all differences do not reach a significant level.

When analyzing the GSE59339 dataset after the FDR correction, no statistically significant differences were found for any of the CpG-sites of the target genes. However, without an FDR test, differences in methylation are detected in all types of comparisons. Thus, compared with the control, differences in the level of methylation were revealed for eosinophilic AD – at two sites (in the *MRE11* and *LTA* genes), for neutrophilic asthma – at two sites (*MLH1*, *TLR4*), for paucigranulocytic – at two sites (*LTA*, *ANG*); when comparing the forms of asthma, depending on the types of inflammation, differences were found between eosinophilic asthma and neutrophilic asthma – by loci in the *MLH1* and *SOCS5* genes, between eosinophilic asthma and paucigranulocytic asthma – by the CpG-site in the *MRE11* gene, between neutrophilic asthma and paucigranulocytic asthma – by sites in the *MRE11* and *SOCS5* genes.

*Conclusion:* The data obtained as a result of reanalysis on differences in the level of DNA methylation of a number of genes associated with bronchial asthma indicate the potential significance of epigenetic modifications in predisposition to this pathology.

*Acknowledgements:* The work was carried out within the framework of the State Task of the Ministry of Science and Higher Education No. 122020300041-7.

## References

1. GEO, <https://www.ncbi.nlm.nih.gov/geo/>.
2. 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
3. Gunawardhana L.P. et al. Characteristic DNA methylation profiles in peripheral blood monocytes are associated with inflammatory phenotypes of asthma. *Epigenetics.* 2014;9(9):1302-1316. doi: 10.4161/epi.33066.



## Seeking an optimal approach to variant calling in medical genetics

Barbitoff Y.A.<sup>1,2\*</sup>, Abasov R.<sup>2</sup>, Predeus A.V.<sup>2</sup>, Glotov A.S.<sup>1</sup>

<sup>1</sup> *Dpt. of Genomic Medicine, D.O. Ott Research Institute of Obstetrics, Gynaecology, and Reproductology, St. Petersburg, Russia*

<sup>2</sup> *Bioinformatics Institute, St. Petersburg, Russia*

\* [barbitoff@bioinf.me](mailto:barbitoff@bioinf.me)

**Key words:** exome sequencing, coverage bias, variant calling, benchmark

*Motivation and Aim:* Accurate variant detection in the coding regions of the human genome is a key requirement for molecular diagnostics of Mendelian disorders. Efficiency of variant discovery from next-generation sequencing (NGS) data depends on multiple factors, including reproducible coverage biases of NGS methods and the performance of read alignment and variant calling software.

*Methods and Algorithms:* In this work, we systematically evaluated the performance of 5 kits for whole-exome (WES) and whole-genome (WGS) sequencing, as well as 4 popular short read aligners and 9 novel and well-established variant calling and filtering methods. We used a large dataset of both “gold standard” (provided by the Genome In A Bottle (GIAB) consortium) and in-house WES and WGS samples.

*Results:* We showed that, contrary to a common view, most of the observed bias in modern WES stems from mappability limitations of short reads and exome probe design rather than sequence composition. When non-repetitive targeted exome regions are considered, WES performs similarly to WGS in terms of coverage and variant calling accuracy. Among software tools tested, Bowtie2 performed significantly worse than other aligners, suggesting it should not be used for medical variant calling. When other aligners were considered, the accuracy of variant discovery mostly depended on the variant caller and not the read aligner. Among the tested variant callers, DeepVariant consistently showed the best performance and the highest robustness. We have also compared the consistency of variant calls in GIAB and non-GIAB samples. With few important caveats, best-performing tools have shown little evidence of overfitting.

*Conclusions:* Our analysis showed that modern strategies of exome sequencing and NGS data analysis allow for high accuracy of genetic variant discovery within coding regions of the human genome. However, there is still a need for development of new library preparation and variant calling methods to enhance variant discovery in the repetitive regions of the coding genome.

*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of Russian Federation (project “Multicenter research bioresource collection – Human Reproductive Health” contract No. 075–15–2021-1058 from September 28, 2021). We thank the Systems Biology Fellowship for providing financial assistance to Y.A.B.

## On the role of genetic factors in the development of preeclampsia

Batlutskaya I.V., Abramova M.Yu.\*

BELGOROD State National Research University, Belgorod, Russia

\* [abramova\\_myu@bsu.edu.ru](mailto:abramova_myu@bsu.edu.ru)

**Key words:** preeclampsia, polymorphism, rs2681472

*Introduction:* Preeclampsia (PE) is one of the most severe complications of pregnancy and, traditionally, occupies one of the leading places among the causes of maternal mortality [1]. There are many theories about the causes of this complication of pregnancy, but a consensus on this issue has not yet been formed. The only indisputable fact is that PE, like most other pathologies, refers to multifactorial diseases, the development of which is due to the interaction of environmental and genetic factors. According to the data of foreign and domestic literature, the genetic component in the formation of this complication of gestation is more than 50 % [2]. However, despite the huge array of studies conducted, the results obtained are often ambiguous and vary depending on the ethno-territorial characteristics of the groups studied.

*Goal:* To study the associations of the rs2681472 polymorphism of the *ATP2B1* gene with the development of preeclampsia in the population of the Central Black Earth Region of the Russian Federation.

*Materials and Methods:* The samples for this study included 236 women with PE (the mean  $\pm$  SD age of  $26.97 \pm 5.38$  years) and 97 individuals with a physiological course of gestation (the mean  $\pm$  SD age of  $26.49 \pm 5.48$  years) of Russian nationality, residents of the Central Black Earth Region of the Russian Federation without any kinship, and who gave voluntary consent to participate in the study. The samples were formed on the basis of the perinatal center of the Saint Joasaph Belgorod Regional Clinical Hospital. The venous blood of all women served as the material for the study. DNA was extracted from whole blood by phenol-chloroform method. Genotyping of the polymorphic variant rs2681472 of the *ATP2B1* gene performed by PCR DNA synthesis on a CFX-96 Real-Time System (Bio-Rad, USA) using standard oligonucleotide primers and probes polymorphism analysis by allele discrimination. The statistical data were analyzed using STATISTICA (V. 6.0.) software for Windows.

*Results and Discussion:* When conducting a population-genetic analysis of the observed frequency distribution of alleles and genotypes according to the polymorphic marker under consideration, it showed its compliance with the theoretically expected distribution according to the Hardy-Weinberg equilibrium. The following frequency of genotypes and alleles of the polymorphic locus rs2681472 of the *ATP2B1* gene established in a sample of women with PE: GG — 2.30 %, AG — 25.74 %, AA — 71.96 %, the frequency A and G alleles was reported to be 84.83 and 15.17 % respectively. In the control group, the frequency was also calculated: homozygote GG — 2.89 %, heterozygote AG — 28.23 % and homozygote AA — 68.87 %. The frequency of A and G alleles in the control group respectively, was reported to be 82.99 % and 17.01 %.

When conducting a comparative analysis of the frequencies of alleles and genotypes rs2681472 of the *ATP2B1* gene, there were no statistically significant differences between the studied samples ( $p > 0.05$ ).

*Conclusions:* The polymorphic locus rs2681472 of the *ATP2B1* gene is not associated with the development of preeclampsia in the population of the Central Black Earth Region of the Russian Federation.

### *References*

1. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin Summary, Number 222. *Obstet. Gynecol.* 2020;135(6):1492-1495. doi: 10.1097/AOG.0000000000003892.
2. Galaviz-Hernandez C., Sosa-Macias M., Teran E. et al. Paternal determinants in preeclampsia. *Front Physiol.* 2019;9:1870. doi: 10.3389/fphys.2018.01870.

## Search for novel genetic variants associated with the variability of cognitive functions in the elderly according to whole exome sequencing data

Bocharova A.<sup>1\*</sup>, Vagaitseva K.<sup>1</sup>, Zarubin A.<sup>1</sup>, Makeeva O.<sup>2</sup>, Zhukova N.<sup>2,3</sup>, Zhukova I.<sup>2,3</sup>, Stepanov V.<sup>1</sup>

<sup>1</sup> Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia

<sup>2</sup> Nebbiolo Center for Clinical Trials, Tomsk, Russia

<sup>3</sup> Siberian State Medical University, Tomsk, Russia

\* [anna.bocharova@medgenetics.ru](mailto:anna.bocharova@medgenetics.ru)

**Key words:** cognitive function, whole exome sequencing, genome, cognitive decline

*Motivation and Aim:* Cognitive decline with the age, both in the normal aging and in the pathological manifestations in form of dementia, is an important public health and social challenge. Next generation sequencing (NGS) is increasingly being used to identify known and novel gene mutations. In this study we used deep (> 50x) whole-genome sequencing (WGS) to search for novel potential markers of cognitive functions in the coding and regulatory parts of the genome in 86 elderly individuals without established diagnoses of neurodegenerative and mental diseases.

*Methods and Algorithms:* The process for enrichment of genomic DNA sample libraries was carried out using SeqCap EZ Library Kit (Roche, USA) according to the SeqCap EZ Library SR User's Guide Version 5.1 following the recommended protocols. Whole exome sequencing (WES) was performed using Illumina HiSeq 2500 platform (San Diego, CA, USA) on 86 samples from a group of elderly individuals without established diagnoses of neurodegenerative and mental diseases. A SeqCap EZ MedExome Enrichment Kit (Roche, USA) was used for exome sequencing. The total volume of nucleotide sequenced sequences is 47 million bp and includes exons, introns, 5'- and 3'-untranslated regulatory regions. The whole sequenced data aligned to the human reference genome (GRCh38) using the BWA-mem algorithm, version 0.7.10. Further, Genome Analysis Toolkit v.3.3 (GATK; Broad Institute) was used for data quality assurance as well as variant discovery. Variant characterization, including filtering, annotation, classification, prioritization and inheritance pattern analysis, was performed using the SnpEff toolbox and proprietary algorithms in the R environment. To elucidate the novel genetic variants associated with the variability of cognitive functions in the elderly we performed an exome-wide association study (EWAS). The compared groups comprised samples of individuals with the highest MoCA test values versus a sample of individuals with low test values. Fisher's exact test was chosen to compare the samples. The analysis was performed in the R environment using own script. The String and Gene Ontology databases were used for pathway and disease association enrichment analysis.

*Results:* In total, 310138 variants were found as a result of exome sequencing of 86 samples: 257711 were single nucleotide variants (SNV), 52427 are insertion-deletion polymorphisms (INDEL). An analysis of the distribution of variants by localization in the gene showed that 24.0 % of the detected variants are localized in introns, 12.7 % – in exons, 10.7 % – in upstream regions, 8.1 % – in downstream regions, 1.5 % – in 3'-

untranslated regions, 0.9 % – in 5'-untranslated regions, 1.3 % – in intron splicing sites. The average density of variants by exome was 1 variant per 9.98 bp. The array of variants for further analysis amounted to 28449 markers after the removal of singletons, selection of variants by frequency in the sample (at least 5 %), and additional filtering by the quality of genotypes (DP > 10 in each sample). Of these, 14528 were localized in exons, including 6473 non-synonymous mutations. After correction for multiple testing, only 4 markers were reached the exome-wide significance threshold: rs71148121 in *PKD1*, rs9898911 in *ATAD5*, rs8069561 and rs692529 in *DNAH17*.

*Conclusion:* Previously, associations with cognitive abilities, memory, Alzheimer's disease and other neurodegenerative diseases were not found for these markers. Analysis of these genes using the String and Gene Ontology databases showed that these genes are involved in processes associated with a change in the state or activity of the cell as a result of a stimulus indicating DNA damage due to environmental influences or metabolic errors (*ATAD5*), processes associated with cell localization (*PKD1*), and with microtubule processes, including microtubule-based movement (*DNAH17*).

*Acknowledgements:* The study is supported by The Research Institute of Medical Genetics, Tomsk National Research Medical Center (122020200083-8).

## Detection of the DPY19L2 gene deletion using nanopore sequencing

Bolshakova E.<sup>1\*</sup>, Saifitdinova A.<sup>1,2</sup>, Leonteva O.<sup>2</sup>, Pavlova O.A.<sup>2</sup>, Vasiljeva O.<sup>2</sup>, Bichevaya N.<sup>2</sup>

<sup>1</sup> Herzen State Pedagogical University of Russia, St. Petersburg, Russia

<sup>2</sup> International Center for Reproductive Medicine, St. Petersburg, Russia

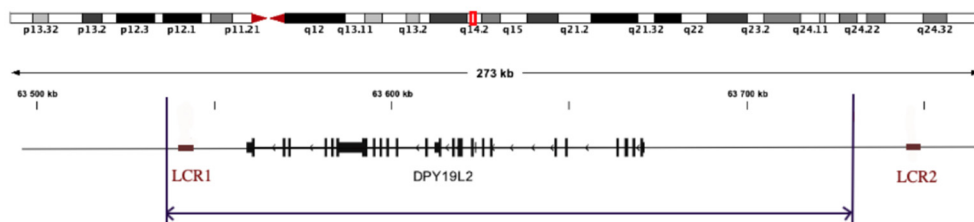
\* bolshakovaaev@gmail.com

**Key words:** globozoospermia, infertility, assisted reproductive technology, nanopore sequencing

**Introduction:** Globozoospermia is a rare (less than 0.1 %) severe disorder in male infertility [2]. Two types of globozoospermia have been described in the literature. Type 1 or total globozoospermia is diagnosed by the presence of 100 % of the spermatozoa with a small, round, and acrosome-free head. Type 2 or partial globozoospermia is classified if the quantity of round-headed spermatozoa in a semen sample is less than 100 % [1]. Since round-headed spermatozoa cannot penetrate the zona pellucida due to the lack of an acrosome, this type of disorder causes primary male infertility.

**Methods and Algorithms:** A married couple came to the International Center for Reproductive Medicine (St. Petersburg, Russia) with complaints of infertility. The man was diagnosed with total globozoospermia, the semen analysis revealed the spherical heads without an acrosome in 100 % of the sperm cells. The sperm was taken for DNA extraction and genome analysis using Oxford Nanopore long read sequencing technology. For DNA extraction was used The Monarch DNA Extraction Kit by New England Biolabs. This was followed by precipitation with ethanol in the presence of 0.3 M sodium acetate. High quality DNA samples were processed using Nanopore Ligation Sequencing Kit SQK-LSK110 to prepare the libraries, which were sequenced using the MinION R9.4.1 flowcells. Obtained FAST5 files were basecalled using Guppy (Oxford Nanopore Technologies) and mapped with Minimap2 to the GRCh38/hg38 genome assembly (Human Genome Reference Consortium). Aligned file was sorted by Samtools. Finally, we used the Integrative Genomics Viewer to visualize your data with masked secondary reads.

**Results:** The analysis revealed a homozygous deletion of a region of about 200 kb at 12q14.2 carrying the entire gene *DPY19L2* (Fig. 1). We confirmed the deletion by amplification with the following primers: FORWARD CTCCAAGATGTCTGCTGGC and REVERSE CACACTTTCCTAAAACATATGCTG, followed by Sanger sequencing.



**Fig. 1.** Schematic representation of the analyzed region. Blue arrow represents the deletion region

*Discussion:* Globozoospermia has been associated with alterations in specific genes. To type I globozoospermia lead mutations in the *PICK1*, *DPY19L2*, and *SPATA16* genes. It has been shown that *DPY19L2* mutations can be detected in a vast majority of globozoospermic patients (66.7 %) [4]. The most common *DPY19L2* mutation is a 200 kb deletion between two very similar 28 kb low copy copy (LCR) flanking the gene. The mechanism underlying the deletion is most likely non-allelic homologous recombination (NAHR) between flanking LCRs [3]. In our case, the extreme points of the about 200-kbp deletion site are located close to the LCRs (Fig. 1).

Detecting a deletion of this size, given the presence of highly homologous sequences to the lost gene in the human genome, is difficult. The use of sequencing, which allows long reads, made it possible to do this reliably and quickly.

In this case, the homozygous form of the mutation arose as a result of closely related marriages and is explained by the region of residence of the patient. Consanguineous marriages increase homozygosity in offsprings leading to an increase in the risk of inherited disorders. This mutation was not observed in other relatives and was found *de novo*.

The identified gene *DPY19L2* deletion in this clinical case led to the proposal of an oocyte fertilisation method using calcium ionophore. Intracytoplasmic Sperm Injection (ICSI) has produced 4 developing embryos that have formed a blastocyst on day 5 of development and can develop into a healthy baby for this family. Genetic counselling based on the information we have identified will enable the family to avoid developing abnormalities in their offspring in the future.

*Conclusion:* Third-generation sequencing technologies give us the ability to quickly identify genomic variations that cause infertility in particular and develop a case management strategy. Since globozoospermia is considered to be a severe genetic disease, descendants of that family should be tested for deletion of this gene, which could prevent an increased risk of transmission of the disease to offspring, especially in case of consanguineous relationships. Accurate deletion mapping will allow us to develop a fast and reliable way to detect pathology.

*Acknowledgements:* The research is conducted at the Department of Human and Animal Anatomy and Physiology of the Herzen State Pedagogical University of Russia within the framework of the research area No. 21 "Adaptive responses of biological systems to specific and nonspecific effects of environmental factors" on "From gene to adaptation: genome lability in implementation of adaptive response and organism development programme in norm and pathology".

## References

1. Anton-Lamprecht I., Kotzur B., Schopf E. Round-headed human spermatozoa. *Fertil Steril.* 1976;27(6):685-693.
2. Dam A.H., Feenstra I., Westphal J.R., Ramos L., van Golde R.J., Kremer J.A. Globozoospermia revisited. *Hum Reprod Update.* 2007;13(1):63-75.
3. Elinati E., Kuentz P., Redin C. et al. Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Hum Mol Genet.* 2012;21(16):3695-3702.
4. Faja F., Pallotti F., Cargnelutti F. et al. Molecular Analysis of DPY19L2, PICK1 and SPATA16 in Italian Unrelated Globozoospermic Men. *Life (Basel).* 2021;11(7):641.

## Loss-of-function mutations in athletes

Boulygina E.A.<sup>1\*</sup>, Valeeva E.V.<sup>1,2</sup>, Semenova E.A.<sup>3,4</sup>, Larin A.K.<sup>3</sup>, Nabiullina R.M.<sup>2</sup>, Mavliev F.A.<sup>4</sup>, Akhatov A.M.<sup>4</sup>, Generozov E.V.<sup>3</sup>, Ahmetov I.I.<sup>2,3,5</sup>

<sup>1</sup> Kazan (Volga region) Federal University, Kazan, Russia

<sup>2</sup> Kazan State Medical University, Kazan, Russia

<sup>3</sup> Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia

<sup>4</sup> Volga Region State Academy of Physical Culture, Sport and Tourism, Kazan, Russia

<sup>5</sup> Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

\* boulygina@gmail.com

**Key words:** LoF mutation, athletes, muscle mass, peak power, reaction time, speed performance

*Motivation and Aim:* Whole-genome and whole-exome sequencing data reveal that individuals carry many variants predicted to inactivate genes (knockouts). Some of naturally-occurring knockout or loss-of-function (LoF) variants may potentially influence athletic performance. The aim of the study was to identify loss-of-function variants in the whole genome of athletes and establish an association between such variants with exercise-related traits.

*Methods and Algorithms:* Twenty Tatar sub-elite male wrestlers participated in the study. DNA was extracted from fasting blood, and whole-genome libraries were sequenced on the Illumina HiSeq 2500 platform. Raw reads were mapped to the human genome hg19 reference using BWA [1], and the variant calling was performed using Strelka2 [2] followed with variants filtering and annotation [3]. Whole-genome variants were validated using the microarray data (HumanOmniExpress BeadChip, 900K SNVs). Athletes have been subjected to several functional tests. The anaerobic lower-limb peak power was assessed during the Wingate 30-sec test. The reaction time was evaluated using the visual computer “Traffic light” test. Specifically, the participants were asked to press the mouse button when the green signal appeared on the screen, and the reaction time was calculated as an average of 3 best attempts out of 5. The muscle mass of wrestlers was measured by the bioimpedance method using a Tanita MC 980 MA electronic scales after fasting for at least 8 hours. In addition, athletes were asked to evaluate their ability to run short distances as “good”, “average” or “poor”. The list of known loss-of-function variants was extracted from the gnomAD v.2.1.1 database [4] (variants annotated as “stop gained”, “frameshift”, “splice donor”, “splice acceptor”) and filtered to exclude low confidence LoF and multi-allelic variants. To test the association between high-quality LoF variants in athletes and their functional characteristics, Spearman rank correlation coefficient was calculated using the additive, dominant, and recessive genetic model. P value less than 0.05 was considered as statistically significant.

*Results:* Over 580 loss-of-function variants (SNPs and indels) were detected in athletes’ genomes. Among them, 263 polymorphisms were frameshift mutations, 170 variants caused a premature stop codon, other mutations occurred in acceptor ( $n = 105$ ) and donor ( $n = 50$ ) splicing site. Fifteen high-confidence polymorphisms were found to be



associated with at least one evaluated exercise-related parameter. Variants in genes *BCKDHA* (rs3217385), *RHBG* (rs2245623), *ATP13A5* (rs74437357), *SLC6A18* (rs7447815), *ANKDD1B* (rs34358), *OR2J1* (rs2394517), and *KIAA1755* (rs41282820) showed an association with lower-limb peak power. Muscle mass values were shown to have correlation with polymorphisms in genes *PRSS48* (rs77216366), *ZACN* (rs1043149), *FUT2* (rs601338), and *H2BW2* (rs2301384). Three SNPs were demonstrated to reduce reaction time, namely, *VWDE* (rs17165936), *APOBEC1* (rs34275479), and *STK31* (rs6945306). Finally, self-reported speed running performance was associated with variants in genes *ANKDD1B* (rs34358), *IDO2* (rs4503083), and *BCKDHA* (rs3217385).

*Conclusion:* In conclusion, we identified 15 LoF mutations associated with exercise-related traits in athletes. Further studies involving larger cohorts of athletes are warranted to replicate and extend these findings.

*Acknowledgements:* This work was part of Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project No. 0671-2020-0058).

### References

1. Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-595.
2. Kim S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594.
3. Boulygina E.A. et al. Whole genome sequencing of elite athletes. *Biol Sport*. 2020;37(3):295-304.
4. Karczewski K.J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.

## Understanding the role of genetics in comorbidity

Bragina E.Yu.<sup>1\*</sup>, Goncharova I.A.<sup>1</sup>, Zhalsanova I.Zh.<sup>1</sup>, Nemerov E.V.<sup>2</sup>,  
Nazarenko M.S.<sup>2</sup>, Freidin M.B.<sup>1</sup>, Puzyrev V.P.<sup>1,2</sup>

<sup>1</sup> *Research Institute of Medical Genetics, TNMRC, Tomsk, Russia*

<sup>2</sup> *Siberian State Medical University, Tomsk, Russia*

\* *elena.bragina72@gmail.com*

**Key words:** bronchial asthma, cardiovascular/metabolic diseases, comorbidity, syntropy, dystropy, gene

*Motivation and Aim:* The pattern of diseases in humans is heterogeneous, which is manifested by various combinations of diseases, including comorbidities associated with a common pathogenetic mechanism, as well as diseases that rarely manifest together at the phenotypic level. According to the concept of syntropy/dystropy or often/rare co-existence diseases, the accumulation of cases of comorbidity in families indicates the non-random connection between diseases in individuals and their relatives, and the importance of genes called syntropic/dystropic respectively, which determine their mutual development [1]. We know little about genes involved in the comorbidity of bronchial asthma with other diseases for example of cardiovascular continuum, with special attention to the mechanisms that might underlie a relationship between them.

*Methods and Algorithms:* We studied in the capacity of the syntropy model of arterial hypertension, coronary heart disease, myocardial infarction, obesity, type 2 diabetes mellitus and bronchial asthma as a significantly more common comorbidity among patients. We focus on cis/trans-eQTL of the genes involved in complex interconnection as a reason for syntropy of bronchial asthma with cardiovascular diseases identified previously using bioinformatics approaches [2, 3]. Genotyping of cis/trans-eQTL was performed using MALDI-TOF mass spectrometry in patients with bronchial asthma combined with cardiovascular/metabolic disorders as well as practically healthy individuals.

*Results:* Bronchial asthma in combination with cardiovascular/metabolic disorders was linked with particular genetic variants affecting gene expression mainly *CAT*, *TLR4*, *ELF5*, *ABTB2*, *UTP25*, *TRAF3IP3*, *NFKB1*, *LOC105377347*, *C1orf74* and *IRF6*. Only rs11590807 which regulate expression of the *UTP25*, *IRF6*, *TRAF3IP3*, *RP1-28O10.1* genes was associated with all studied comorbid phenotype of bronchial asthma and diseases of the cardiovascular continuum. Additionally we established the specificity of the associations of the cis/trans-eQTL in the development of time-ordered comorbidities. The rs1010461 variant, which regulates the expression of *RNASE4* and *ANG* genes, is linked with cardiovascular diseases as the first phenotype of this comorbidity with bronchial asthma. The rs769214, rs11032700, rs11032699, rs484214, and rs480575 variants, which regulate the expression of *CAT* gene, are associated with bronchial asthma as the first phenotype of comorbidity.

*Conclusion:* We identified the cis/trans-eQTL that are involved in the pathogenesis of bronchial asthma and cardiovascular/metabolic diseases, and may also contribute to the comorbidity of these diseases. The cis/trans-eQTL affecting the expression of many genes can serve as potential biological markers of complex causal relationships between bronchial asthma and cardiovascular/metabolic disorders.

*Acknowledgements:* The work was carried out within the framework of the State Task of the Ministry of Science and Higher Education (075-00603-19-00).

### *References*

1. Puzyrev V.P. Genetic bases of human comorbidity. *Russ J Genet.* 2015;51(4):408-417.
2. Saik O.V. et al. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med Genomics.* 2018;11(Suppl 1):15.
3. Zolotareva O. et al. Comorbidity of asthma and hypertension may be mediated by shared genetic dysregulation and drug side effects. *Sci Rep.* 2019;9(1):16302.

## Analysis of hypoxia signature in TCGA-BRCA dataset and in mcf-7 cell line

Brovkina O.\*, Yusubalieva G., Nikitin A.

*Federal Research and Clinical Center, FMBA of Russia, Moscow, Russia*

\* *brov.olia@gmail.com*

**Key words:** hypoxia, breast cancer, tumor microenvironment, transcriptome analysis, epithelial-mesenchymal transition, database screening

*Motivation and Aim:* Hypoxia is a common feature for most solid tumors and is associated with a poor prognosis, while the lack of oxygen enrichment of tumor tissues and their microenvironment reduces the effectiveness of radiation therapy and some chemotherapy drugs [1]. Therefore, the development of clinical drugs that reduce hypoxia, as well as effective and accurate measuring, are being actively pursued [2]. There are several possible methods for detection of hypoxia. One of the most promising methods is gene expression profiling associated with the transcriptional response of a tumor to its hypoxic microenvironment.

*Methods:* The analysis of large dataset was performed using open database TCGA-BRCA. The analysis included 531 tumor and normal samples. 444 samples were breast carcinoma samples from stage III and IV patients of the disease, 87 samples were normal controls. Differentially expressed (DE) genes were analyzed using the DeSeq2 package. To normalize the obtained data, the TMM (Trimmed Mean of M-values) method was applied with the calculation of CPM (counts per million) taking into account the normalization coefficients. Quasi-likelihood F-test (QLF) and Mann-Whitney U-test (MW) were used to assess the significance of changes in gene expression. High and low expressed genes were selected using  $p_{adj} < 0.05$  and  $\log_{2}FC > 1.5$ . The fgsea and gprofiler packages were used to enrich the data on DE genes. Among the list of signaling pathways, pathways associated with hypoxia were selected. Then, using the Elastic Net model, the degree of significance of the difference between tumor samples from the norm in terms of DE genes that are part of the hypoxia signaling pathways was determined.

The experimental validation of selected panel of genes was performed on 4T1 cell line. This tumor-derived cell line is an animal model for stage IV human breast cancer. The cells were cultured in RPMI 1640 medium (Life Technologies, USA) supplemented with 5 % fetal bovine serum (FBS) (Life Technologies) and 1 % PenStrep (Life Technologies), in a humidified atmosphere containing 5 % CO<sub>2</sub> and 5 % O<sub>2</sub> in an incubator (Sanyo, Japan) at 37 °C. The cells were subcultured at confluence by treatment with 0.05 % trypsin and 0.02 % EDTA in phosphate buffered saline (PBS). Total RNA isolation from adipose tissue was carried out with a Rneasy Lipid Tissue Mini Kit (Qiagen, Germany) on the automatic QIAcube station according to the manufacturer's protocol. Reverse transcription of RNA was performed using MMLV reverse transcriptase (Evrogen, Russia) and random hexanucleotide primers. Each setting of the reverse transcription reaction included negative controls containing no RNA.

We estimated expression levels of 18 mRNAs by qPCR on a CFX96 thermal cycler (Bio-Rad, USA). A SYBR Green intercalating dye, qPCRmix-HS SYBR (Evrogen), was

added to the PCR mixture. The total volume of the PCR reaction was 20 µl. Conditions for amplification of DNA fragments were: 95 °C for 20 sec – 1 cycle; 95 °C for 1 sec, 60–64 °C for 20 sec – 40 cycles. The B2M gene was a reference gene.

*Results:* A total of 738 DE genes were identified, of which 499 were up-regulated and 239 were down-regulated. Expression of 36 genes was associated with hypoxia, changes in 50 genes were caused by the process of epithelial-mesenchymal transition (EMT). In addition to previously known genes, the alpha-lactalbumin LALBA gene and the exonuclease gene stimulated by interferon 20 ISG20 were overexpressed. Analysis of signaling pathways revealed the following significant mechanisms: hypoxia-inducible factor-1 (HIF) regulatory pathways, glucose metabolism (SLC2A1, PFKFB4, LDHA, ALDOA, ALDOC), and angiogenesis (ANGPTL4, VEGFA). EMT included the following pathways: fatty acid metabolism (LEP, TGFB, FASN), apoptosis (BNIP3L, BNIP3, NDRG1), Krebs cycle (IDH, SDH), and cell adhesion (SNAIL, TWIST, ZEB1, ZEB2): GO:0001666 GO: 0005975 GO: 0005996 GO: 0006090 GO: 0006091 GO: 0006096 GO: 0033554 GO: 0035694 GO: 0036293 GO: 0036294 GO: 0046031 GO: 0046034 GO: 0046939 GO: 0055086 GO: 0070482 GO: 0071453 GO: 0071456 GO: 0072521 GO: 0072521 GO: 0072521 GO :1901135 GO:0005737

Based on the obtained signatures, we divided the patients into 2 clusters. The first cluster included patients (317/60 %) with high rates of hypoxia (according to the ConsensusClusterPlus program), the second included patients (214/40 %) whose rates were lower. The analysis of survival curves showed that the survival rate of patients with high rates of hypoxia was significantly lower compared to patients from the low hypoxia group. It is noteworthy that in the first cluster patients with metastases predominated, as well as those with impaired EMF mechanisms.

The qPCR analysis confirmed high expression for *LEP*, *ADM*, *ALDOA*, *BNIP3*, *LOX*, *NDRG1*, *P4HA1*, *PDK1*, *SLC2A1*, *LALBA*, *ISG20* genes (lgFC > 1,5, p.adj < 0,05).

*ANRD37*, *BNIP3L*, *EGLN3*, *P4HA2*, *PFKFB3* genes had no significant difference compared with cell in normoxic condition, while *FAM162A*, *KCTD11* genes had significantly low expression (lgFC < -1,5, p.adj < 0,05).

*Conclusion:* The stratification of patients based on the hypoxia signatures makes it possible to identify groups with a high risk of disease progression. Evaluation of gene expression profiles associated with hypoxia allows to detect metastatic potential and select the optimal treatment strategy. The evaluation of individual genes and common pathways in hypoxia is consistent with the global trend towards a personalized approach to the diseases treatment and contributes to the search for targeted therapeutic drugs.

*Acknowledgements:* The research was funded by the Russian Science Foundation, grant No. 21-75-00041.

## References

1. Burroughs S.K., Kaluz S., Wang D., Wang K., Van Meir E.G., Wang B. Hypoxia inducible factor pathway inhibitors as anticancer therapeutics. *Future Med Chem.* 2013;5(5):553-72. doi: 10.4155/fmc.13.17.
2. Bibby B.A.S., Thiruthaneeswaran N., Yang L., Pereira R.R., More E., McArt D.G., O'Reilly P., Bristow R.G., Williams K.J., Choudhury A., West C.M.L. Repurposing FDA approved drugs as radiosensitizers for treating hypoxic prostate cancer. *BMC Urol.* 2021;21(1):96.

## Genes, encoding heat-resistant obscure (Hero) proteins: new players in ischemic stroke pathogenesis

Bushueva O.<sup>1\*</sup>, Soldatov V.<sup>2</sup>, Belykh A.<sup>1</sup>, Kobzeva K.<sup>1</sup>, Soldatova M.<sup>1</sup>, Gurtovoy D.<sup>1</sup>, Shilenok I.<sup>3</sup>, Deykin A.<sup>2</sup>

<sup>1</sup> Kursk State Medical University, Kursk, Russia

<sup>2</sup> Belgorod State National Research University, Belgorod, Russia

<sup>3</sup> Kursk Emergency Hospital, Kursk, Russia

\* [olga.bushueva@inbox.ru](mailto:olga.bushueva@inbox.ru)

**Key words:** ischemic stroke, Hero proteins, heat shock proteins, chaperones, genetic variants

*Motivation and Aim:* Despite significant efforts made by the global medical community in the field of prevention and treatment of stroke, this disease remains the third leading cause of death worldwide and is the leading cause of disability, cognitive decline and dementia. At the same time, up to 80 % of all stroke cases are ischemic stroke (IS). A critical decrease in perfusion in IS leads to hypoxic damage of brain tissues. Cascades of ischemic and reperfusion damage to neurons and glia lead to cell death. Due to the accumulation of reactive oxygen species, calcium overload and the accumulation of hydrogen ions, hypoxic stress turns into chemical one. Violation of homeostasis leads to the proteotoxic effect of accumulated metabolites and denaturation of intracellular proteins. It is the critical accumulation of damaged enzymes and structural proteins that causes irreversible changes leading to necrosis or programmed cell death. To reduce the structural and functional consequences of hypoxic and other negative effects, a special class of proteins, chaperones, functions in cells. The main function of chaperones is to restore the native tertiary and quaternary structures of damaged proteins. In 2020, scientists from Japan discovered a new class of chaperones, "Hero proteins". Hero proteins are hydrophilic and highly charged, and function to stabilize various "client" proteins, protecting them from denaturation even under stress conditions such as heat shock, desiccation, and exposure to organic solvents [1]. The genetic aspects of the involvement of these proteins in the predisposition to ischemic stroke have not been studied.

*Methods and Algorithms:* A total of 1799 unrelated Russian individuals from Central Russia was examined: 844 patients with IS and 935 healthy individuals, without a history of cardio- and cerebrovascular diseases. The study was approved by the Regional Ethics Committee of the Kursk State Medical University. Genotyping of 12 polymorphic variants in genes encoding Hero proteins (rs11666524, rs2277947, rs346157, rs346158, rs10104 *C19ORF53*, rs2900262 *C9ORF16*, rs4644832 *SERF2*, rs4655707, rs12566098, rs6702742, rs1058074 *SERBP1*, rs6677 *C11ORF58*) was performed by TaqMan-based PCR on a CFX96 amplifier (Bio-Rad, United States). Protocols for genotyping were designed in the Laboratory of genomic research (Research Institute for Genetic and Molecular Epidemiology of Kursk State Medical University). To analyze associations of genotypes with the risk of IS, a regression model was used. All calculation were performed in the SNPStats program, available online (<https://www.snpstats.net/start.htm>) with adjustment for sex and age.

**Results:** Seven genetic variants (rs11666524, rs2277947, rs346157, rs346158, rs10104 *C19ORF53*, rs4644832 *SERF2*, rs12566098 *SERBP1*) showed associations with the increased risk of IS (Table 1).

**Table 1.** Statistically significant associations of SNPs in the genes encoding Hero chaperones with the development of ischemic stroke

SNP	Minor allele	corOR (95 % CI) <sup>1</sup>	P <sup>2</sup>
rs11666524 (G/A) <i>C19ORF53</i>	A	1.61 (1.10-2.36)	0.015 <sup>R</sup>
rs2277947 (G/A) <i>C19ORF53</i>	A	1.86 (1.24-2.78)	0.0023 <sup>R</sup>
rs346157 (A/G) <i>C19ORF53</i>	G	1.28 (1.02-1.62)	0.036 <sup>R</sup>
rs346158 (T/C) <i>C19ORF53</i>	C	1.65 (1.13-2.40)	0.0091 <sup>R</sup>
rs10104 (A/G) <i>C19ORF53</i>	G	1.75 (1.15-2.67)	0.01 <sup>R</sup>
rs4644832 (A/G) <i>SERF2</i>	G	1.33 (1.08-1.65)	0.0065 <sup>C</sup>
rs12566098 (C/G) <i>SERBP1</i>	G	1.58 (1.31-1.91)	< 0.0001 <sup>C</sup>
<sup>1</sup> – odds ratio and 95 % confidence interval;			
<sup>2</sup> – P value with adjustment for gender, age			
<sup>R</sup> – recessive regression model; <sup>C</sup> – codominant regression model			

The most significant associations with the development of ischemic stroke were found in the analysis of rs12566098 *SERBP1*. *SERBP1* (SERPINE1 MRNA Binding Protein 1) is a protein coding gene. Among its related pathways are apoptosis-related network (<https://www.genecards.org>). This genetic variant has a high regulatory potential. In particular, it affects the binding of 50 transcription factors, depending on the carriage of the reference or alternative allele. Allele G rs12566098 *SERBP1* is associated with reduced gene expression and affects histone modifications (H3K4me1 and H3K27ac) in peripheral blood cells.

**Conclusion:** The present study was the first to show that genetic variations at the Hero proteins may contribute to the risk of ischemic stroke.

**Acknowledgements:** The study is supported by Russian Science Foundation (No. 22-15-00288).

### References

1. Tsuboyama K., Osaki T., Matsuura-Suzuki E., Kozuka-Hata H., Okada Y., Oyama M., ..., Tomari Y. A widespread family of heat-resistant obscure (Hero) proteins protect against protein instability and aggregation. *PLoS Biology*. 2020;18(3):e3000632.

# Molecular mechanisms of cardio- and cerebrovascular comorbidity: from experimental analysis of structural and epigenetic variations in the human genome to post-GWAS analysis of genetic correlations between diseases

Bushueva O.<sup>1\*</sup>, Nazarenko M.<sup>2</sup>, Polonikov A.<sup>1</sup>, Ivanov V.P.<sup>1</sup>

<sup>1</sup> Kursk State Medical University, Kursk, Russia

<sup>2</sup> Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

\* [olga.bushueva@inbox.ru](mailto:olga.bushueva@inbox.ru)

**Key words:** comorbidity, hypertension, coronary heart disease, stroke, genetic variants, DNA methylation, genetic correlations, LCV analysis

*Motivation and Aim:* Common and often combined cardio- and cerebrovascular diseases (CCVDs), including arterial hypertension (AH), coronary heart disease (CHD), and cerebral stroke (CS), are the main cause of mortality and disability worldwide [1]. Despite the obvious progress in the genetics of human CCVDs, investigation aimed at studying the molecular genetic mechanisms of the formation of cardio- and cerebrovascular comorbidity are rare. The aim of our study was the molecular genetic, epigenetic and bioinformatic analysis of the formation of comorbid cardio- and cerebrovascular diseases.

*Methods and Algorithms:* A total of 2685 unrelated Russians (890 healthy individuals and 1795 patients with various CCVDs and their combinations) from Kursk region was examined; written informed consent was obtained from all participants prior to entering the study [2, 3]. The study was approved by the Ethics Committee of Kursk State Medical University. When selecting genes and polymorphisms for genotyping, the main emphasis was placed on the genes of the antioxidant and detoxification systems, as well as known CCVDs candidate genes that showed an associations with the risk of hypertension, coronary heart diseases and stroke in previous studies; the amount of 34 genetic markers were genotyped. A log-additive regression model was used to interpret the results of the analysis of associations. The correction for multiple comparisons was carried out using an adaptive permutation test in the PLINK program. The Bonferroni correction was introduced for the number of comparison groups. DNA methylation analysis was performed in 84 healthy individuals and 128 patients with various CCVDs and their combinations. To assess the status of DNA methylation in peripheral blood cells, the promoter regions of the gene *TXNRD1* (3 CpG-site), *GSTP1* (2 CpG-site) and *GCLM* (4 CpG-site) and exons 4–6 *MPO* (3 CpG-site) were selected [4]. The Mann-Whitney test with the Bonferroni correction for multiple comparisons was used for comparative analysis of the methylation levels. The Complex-Traits Genetics Virtual Lab (<https://vl.genoma.io/>) was engaged for bioinformatic evaluation of genetic correlations between various cardio- and cerebrovascular phenotypes (by LD score method), and also to analyze the mechanisms of formation of genetic correlations (LCV, or analysis of models of the latent causal variable) [5].

*Results:* The arterial hypertension was associated with genetic variants *FGF* rs1800790 (OR = 0,81;  $P_{\text{perm}} = 0,007$ ) and *CYP1A2* rs762551 (OR = 0,81;  $P_{\text{perm}} = 0,005$ ). *PON2*



rs7493 showed an association exclusively with the comorbid phenotypes: AH+CHD (OR = 1,32;  $P_{\text{perm}} = 0,002$ ), AH+CS, AH+CHD+CS (OR = 1,76;  $P_{\text{perm}} = 1,00 \times 10^{-6}$ ). At the same time, *TXNRDI* rs1128446 was associated solely with cerebrovascular phenotype (OR = 1,83;  $P_{\text{perm}} = 1,00 \times 10^{-6}$ ). We observed an increase in the level of *MPO* gene methylation in healthy individuals (35,76 %) compared with patients suffering from AH (28,74 %), AH+CHD (26,29 %), AH+CS (24,27 %), AH+CHD+CS (22,1 %); P-levels were  $P_{\text{bonf}} = 8,56 \times 10^{-4}$ ,  $P_{\text{bonf}} = 1,39 \times 10^{-7}$ ,  $P_{\text{bonf}} = 4,32 \times 10^{-10}$  and  $P_{\text{bonf}} = 2,70 \times 10^{-10}$  consequently. To analyze the genetic correlations between different diseases as "input data", there were used the results of GWASs for three diseases (arterial hypertension, ischemic heart disease, cerebral stroke), presented in UKBB. Based on bioinformatic post-GWAS analysis of UKBB data in the Complex-Traits Genetics Virtual Lab statistically significant genetic correlations were established for the following phenotype pairs: AH vs CHD ( $rG = 0,3701$ ;  $P = 0,0012$ ); CS vs CHD ( $rG = 0,785$ ;  $P = 0,0005$ ). At the same time, the phenotypes of AH vs CS did not correlate with each other. Statistically significant latent causal variable model was discovered exclusively for the phenotypes CHD vs CS, supposing that the bigger part of genetic component of CHD causes CS ( $GCP = 0,62$ ;  $P_{\text{LCV}} = 0,03$ ), while genetic correlations between AH and CHD are mainly caused by horizontal pleiotropy.

*Conclusion:* Thus, the structural and epigenetic variability of the human genome is one of the most important factors in the formation of complex, comorbid cardio- and cerebrovascular phenotypes.

*Acknowledgements:* The study is supported by Kursk State Medical University.

### References

1. Read S.H., Wild S.H. Prevention of premature cardiovascular death worldwide. *Lancet*. 2020;395(10226):758-760.
2. Bushueva O.Y., Bulgakova I.V., Ivanov V.P., Polonikov A.V. Association of flavin monooxygenase gene E158K polymorphism with chronic heart disease risk. *Bull Exp Biol Med*. 2015;159(6):776-779.
3. Bushueva O.Yu. Single nucleotide polymorphisms in genes encoding xenobiotic metabolizing enzymes are associated with predisposition to arterial hypertension. *Res Results Biomed*. 2020;6(4):447-456.
4. Bushueva O., Barysheva E., Markov A., Belykh A., Koroleva I., Churkin E., ..., Nazarenko M. DNA Hypomethylation of the *MPO* Gene in Peripheral Blood Leukocytes Is Associated with Cerebral Stroke in the Acute Phase. *J Mol Neurosci*. 2021;71(9):1914-1932.
5. Cuellar-Partida G., Lundberg M., Kho P.F., D'Urso S., Gutiérrez-Mondragón L.F., Ngo T.T., Hwang L.D. Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. *BioRxiv*. 2019;518027.

## Functional enrichment analysis of cytochrome P450 genes differential expression in non-smoking lung cancer patients

Buslaev V.\*, Soboleva O.

*Institute of Human Ecology of FRC CCC, SB RAS, Kemerovo, Russia*

\* [vladislavbus94@gmail.com](mailto:vladislavbus94@gmail.com)

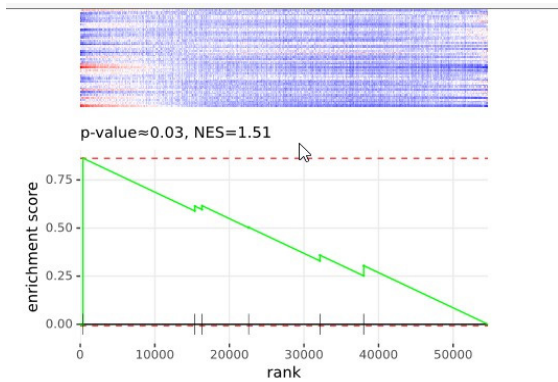
**Key words:** lung cancer, non-smokers, cytochromes P-450, GSEA-analysis, NTA-analysis

*Motivation and Aim:* Lung cancer is featured by high level of morbidity and mortality among the population [1]. It is characterized by presence of two main histopathological forms: small-cell lung cancer and non-small lung cancer. There are many determinants that can modulate individual predisposition to malignant disorders. Smoking is considered as the strongest life style risk factor for lung cancer development. However development of this pathology in non-smoking patients is still under discussion. Cytochromes of P450 family (CYP's) are actively considered as potential biomarkers for oncology development [2]. Normally, these enzymes are able to metabolize products of air pollution in to inactive compounds but also their performance may affect the occurrence of oxidative stress in cells. Active forms of oxygen can directly interact with DNA and initiate mutagenesis. Finally it can lead to uncontrolled cell divisions and cancerogenesis. This study was implemented to test potential contribution of CYP's in to lung cancer development in non-smoking patients.

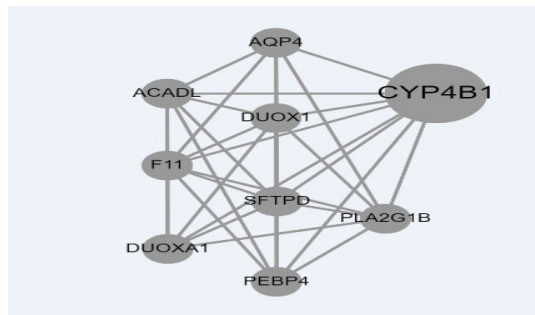
*Methods and Algorithms:* The most frequently studied genes from this group (*CYP2E1*, *CYP2A6*, *CYP4Z2P*, *CYP19A1*, *CYP4A22*, *CYP4B1*) were chosen for present analytic survey. Transcriptome dataset of non-smoking women with lung cancer obtained from NCBI GEO database was used for investigation of these genes contribution. Gene set enrichment analysis (GSEA) method was used to estimate level of cytochrome P450 genes association with lung cancer. Network topology analysis (NTA) and TCGA database were applied to reveal protein-protein interactions of cytochromes during cancerogenesis. This method along with Gene Ontology database application allowed to detect crucial biological processes (top enriched GO categories) in those these cytochromes are involved.

*Results:* As a result GSEA functional enrichment analysis showed significant association between cytochromes P450 enhanced expression and lung cancer development in non-smoking female patients. Along with GSEA-plot significant parameters of p-value and normalized enrichment score (NES) were registered (Fig. 1). Crucial enzymes that were involved in detoxification metabolic processes demonstrated significant interactions with CYP4B1 protein (Fig. 2). Among top 10 enriched GO categories signaling pathways associated with metabolism of benzene-containing compounds (GO:0042537) and drug catabolic process (GO:004737) were detected (Fig. 3).

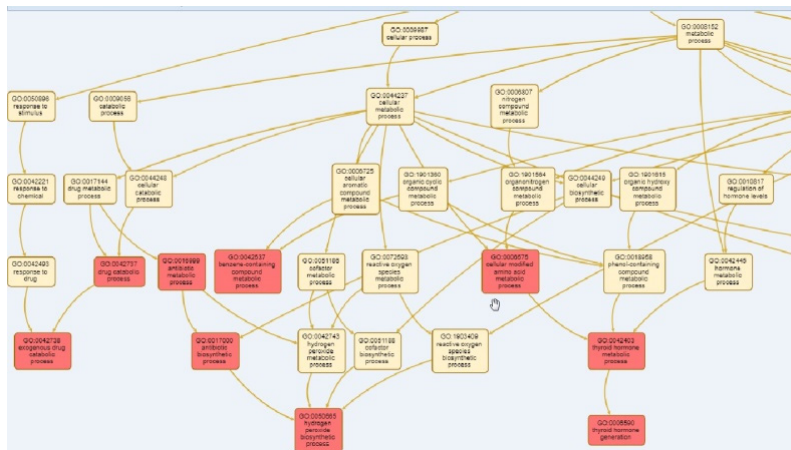
*Conclusion:* CYP genes that were tested via GSEA-analysis can be considered as potential biomarkers for lung cancer formation in non-smokers. They can act in gene networks and exhibit their features. Biological signaling pathways that can make potential contribution in this pathology were determined by NTA analysis. In this case lung cancer is possibly arises due to specificity of xenobiotics metabolism.



**Fig. 1.** GSEA functional enrichment analysis of CYP-genes



**Fig. 2.** Protein-protein interactions of CYP4B1



**Fig. 3.** Top enriched GO-categories (marked by red) associated with CYP's and lung cancer

**Acknowledgements:** The study is supported by state assignment (Project No. 0286-2022-0008).

### References

1. Bray F. et al. 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. Mittal B. et al. Cytochrome P450 in cancer susceptibility and treatment. *Adv Clin Chem.* 2015;71:77-139.

# Trans-biobank genome-wide association analysis of pregnancy complications

Changalidi A.I.<sup>1,2\*</sup>, Maksiukenko E.M.<sup>1</sup>, Tkachenko A.A.<sup>1</sup>, Barbitoff Y.A.<sup>1,2</sup>,  
Glotov A.S.<sup>1</sup>

<sup>1</sup> *Dpt. of Genomic Medicine, D.O. Ott Research Institute of Obstetrics, Gynaecology and Reproductology, St. Petersburg, Russia*

<sup>2</sup> *Bioinformatics Institute, St. Petersburg, Russia*

\* [anton@bioinf.me](mailto:anton@bioinf.me)

**Key words:** GWAS, SNP, meta-analysis, pregnancy, ukb, finngen

*Motivation:* For the last two decades an increasing number of genome-wide association studies (GWAS) were successfully conducted to identify the genetic variants associated with human traits and predispositions to different diseases. Since many biobanks and consortia are currently investigating identical or related traits in different populations, it is promising to aggregate the results of large-scale GWAS and conduct trans-biobank meta-analyses. It seems important to meta-analyze summary statistics from different studies to better understand genetics of complex diseases and traits in humans.

*Methods:* In our current study, we aggregated summary statistics of several GWASes of pregnancy-related traits from the UK Biobank and FinnGen projects. We used METAL and MTAG, Multi-Trait Analysis of GWAS, for meta-analysis of genome-wide summary statistics of 24 pregnancy-related traits (with different preprocessing approaches and settings). We also used HuGE Navigator Phenopedia tool to search for genes implicated in pregnancy complications in other gene-level association studies.

*Results and Conclusions:* 1 and 339 variants with genome-wide significant association with pregnancy complications were identified in UK Biobank and FinnGen data, respectively. When meta-analysis methods were used, results of the analysis with different meta-analysis tools tend to be highly similar (Pearson's correlation coefficient is almost 1). However, METAL was chosen as the best option as its results did not show over-inflation of the association statistics. Meta-analysis of pregnancy-related traits identified 6 variants with genome-wide significant association for three traits. Some of the identified loci overlapped with earlier genome-wide association efforts and gene-level association studies according to HuGE Navigator. The results point to a great role of immunity-related processes in driving pregnancy complications.

## **Ethnic assortativeness and endogamy index among Yakuts in the Republic of Sakha (Yakutia)**

Danilova A.\*, Sukhomyasova A., Yakovleva A., Maksimova N.

*M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia*

\* *ana-danilova@yandex.ru*

**Key words:** ethnic assortativeness, endogamy index, Yakuts

*Motivation and Aim:* From the point of view of determining the direction of changes in the genetic characteristics of populations, the description of the marriage structure of populations is important. Data on the uniqueness of the marriage structure are important for predicting the direction of development of the gene pool of populations, choosing a strategy for studying hereditary diseases, as well as for medical genetic counseling of families with hereditary and congenital diseases in a particular region and population [1]. The Republic of Sakha (Yakutia) is characterized by multinationality, where the Yakuts make up the majority of the population, more than 45 %. The representation of the Yakuts in the districts varies from 2.5 to 98.3 %. The genetic-epidemiological, clinical-genealogical and molecular-genetic studies carried out made it possible to identify ethno-specific "Yakut hereditary diseases" among the Yakuts, with foci of accumulation in the regions of the Republic. Based on this, the purpose of this study was to study the marital structure (ethnic assortativeness and endogamy index) among the Yakuts.

*Methods and Algorithms:* From 2001–2016 in the course of complex expeditions, records of acts on marriages were collected from the archive of the registry office in 10 districts of Yakutia: the Abyisky district – Belaya-Gora village (233 marriages), the Verkhnevilyuysky district – Verkhnevilyuysk and Khomustakh villages (115 marriages), Verkhnekolymsky district – Zyryanka village (200 marriages), Gorny district – Berdigestyakh village (352 marriages), Zhigansky district – Zhigansk village (253 marriages), Momsky district – Khonuu village (303 marriages), Olekminsky district – Olekminsk city (318 marriages), Lensky district – Lensk city (405 marriages), Churapchinsky district – Dyabyla and Kharbala-1 villages (95 marriages) and Ust-Aldan region – Borogontsy village (248 marriages). The surveyed areas differ in ethnic composition and geographical location. Ethnic assortativeness [2] and endogamy index were assessed: at the level of the district, at the level of the republic.

*Results:* The level of ethnic assortativeness depended on the representation of the Yakuts in the studied areas. In Churapchinsky, Ust-Aldansky and Verkhnevilyuysky districts (one-national Yakut regions), negative values were obtained, i.e. there is no mating assortativeness here, and high estimates of this indicator were found in those areas where the Yakuts are represented in a minority in terms of numbers (Verkhnekolymsky, Lensky and Olekminsky). Here we can talk about the effect of "national minority", i.e. when different ethnic groups live in the same territory, the "national minority effect" is revealed, which consists in the fact that marriage assortativeness is highest among representatives of small ethnic groups. Based on the assessment of the endogamy index at the district level, it can be said that the Yakuts have territorial differentiation and varied widely – from 24 % of the Zhigansky district to 90 % of the Verkhnekolymsky

district. At the level of the republic, the share of one-ethnic Yakut marriages between natives of Yakutia in the studied areas exceeds 97 %, i.e. among the Yakuts, migration processes are almost completely limited to the territory of the republic.

*Conclusion:* Thus, territorial differentiation of the studied indicators depending on representation of Yakuts in the studied areas was revealed. Population-genetic studies in the RS(Y) should be carried out at the level of the district, which is an elementary population. The results of this study will be used in planning medical genetic counseling and predicting the prevalence of hereditary pathology in the Republic of Sakha (Yakutia).

*Acknowledgements:* The work was supported by the State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. FSRG-2020-0014 "Genomics of the Arctic: Epidemiology, Heredity and Pathology".

### *References*

1. Nariman A. et al. The genetic background of Southern Iranian couples before marriage. *BJMG*. 2016;19(2):71-74.
2. Kurbatova O.L., Pobedonostseva E.Y. Genetic and Demographic Processes in Multinational Populations. *Adv Modern Genet*. 1996;20:38-61.

## Annotation of uORFs in the OMIM genes allows to identify disease-causing variants in 5'UTRs

Filatova A.<sup>1\*</sup>, Reveguk I.<sup>2</sup>, Piatkova M.<sup>3,4</sup>, Bessonova D.<sup>5</sup>, Kuziakova O.<sup>6</sup>, Demakova V.<sup>5</sup>, Romanishin A.<sup>6,7</sup>, Fishman V.<sup>8,9</sup>, Imanmalik Y.<sup>8</sup>, Chekanov N.<sup>8</sup>, Skitchenko R.<sup>10</sup>, Barbitoff Y.<sup>11,12,13</sup>, Kardymon O.<sup>8</sup>, Skoblov M.<sup>1</sup>

<sup>1</sup> *Research Centre for Medical Genetics, Moscow, Russia*

<sup>2</sup> *Laboratoire de Biologie Structurale de la Cellule, École Polytechnique, Paris, France*

<sup>3</sup> *Institute of Chemistry, FEB RAS, Vladivostok, Russia*

<sup>4</sup> *Institute of High Technologies and Advanced Materials, Far Eastern Federal University, Vladivostok, Russia*

<sup>5</sup> *Medical Center, Far Eastern Federal University, Vladivostok, Russia*

<sup>6</sup> *Institute of Life Sciences and Biomedicine, Far Eastern Federal University, Vladivostok, Russia*

<sup>7</sup> *Institute of Life Sciences, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

<sup>8</sup> *Bioinformatic group, AIRI, Moscow, Russia*

<sup>9</sup> *Molecular Mechanisms of Ontogenesis, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>10</sup> *Computer Technologies Laboratory, ITMO University, St. Petersburg, Russia*

<sup>11</sup> *Bioinformatics Institute, St. Petersburg, Russia*

<sup>12</sup> *Dpt. of Genomic Medicine, D.O. Ott Research Institute of Obstetrics, Gynaecology, and Reproductology, St. Petersburg, Russia*

<sup>13</sup> *Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia*

\* [maacc@yandex.ru](mailto:maacc@yandex.ru)

**Key words:** uORF, 5'UTR, translation initiation start, Mendelian disorders, hereditary diseases

**Motivation and Aim:** To date, the identification of disease-causing variants generally focuses on the coding part of the genes, although an increasing number of studies highlight the role of non-coding variants in diseases development. It was previously shown that variants in the 5'UTR could lead to hereditary diseases through disruption of upstream open reading frames (uORF). However, the identification of such variants in routine clinical genetic testing still remains a challenge since there is no precise annotation of uORFs in human disease-associated genes.

**Methods and Algorithms:** We used different types of publicly available data (Ribo-seq, mRNA-seq, CAGE data previously predicted by uORF) and self-generated data (Kozak scores) to manually annotate upstream translation initiation sites (TIS) in ~3,600 OMIM genes. Based on obtained data, we created a list of uORFs and investigated their various features. Experimental validation was conducted using dual luciferase assay. Two software tools were written in Python and are publicly available on the Github repository.

**Results:** We annotated about 5 thousand upstream TISs in 1835 OMIM genes. The major group of these TISs was related to uORFs. Comparison of our annotation with the previous studies produced a list of “high confidence” uORFs. In addition, we created a tool to assess the effects of nucleotide variants located in uORFs. This allowed us to identify variants from HGMD/ClinVar that disrupt uORFs and thereby could lead to hereditary diseases. We noticed that the distribution of uORFs-affecting variants differs between pathogenic variants and population variants from gnomAD. In addition, for

several genes, we performed luciferase experiments to test the translation of the predicted uORFs and to confirm the pathogenicity of the mutations identified in these uORFs. Finally, drawing on manually curated data, we developed a machine-learning algorithm that allows us to predict the TISs and uORFs in other human genes.

*Conclusion:* Thus, we attempted to obtain qualitative annotation of uORFs in the OMIM genes, which may be useful in interpreting the patient-derived variants in 5'UTRs.

*Acknowledgements:* This study was performed using the infrastructure of AIRI, Artificial Intelligence Research Institute (Moscow, Russia). Preliminary analysis of Ribo-seq data was carried out within the state assignment of the Ministry of Science and Higher Education of the Russian Federation for RCMG.



## Mutation c.396dupT in the *CLN6* gene – the main cause of neuronal ceroid lipofuscinosis in Yakutia

Golikova P.<sup>1\*</sup>, Sukhomyasova A.<sup>1,2</sup>, Nikolaeva I.<sup>2</sup>, Gurinova E.<sup>2</sup>, Petukhova D.<sup>1</sup>, Maksimova N.<sup>1</sup>

<sup>1</sup>Research Laboratory “Molecular Medicine and Human Genetics”, Medical Institute, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup>Medical Genetic Center, Republican Hospital No. 1 – National Medical Center, Yakutsk, Russia

\* Golikova2906@gmail.com

**Key words:** CLN6, neuronal ceroid lipofuscinosis, neurodegenerative disease

*Motivation and Aim:* The neuronal ceroid lipofuscinosis (NCLs) are neurodegenerative disorders, mostly of childhood onset [1]. They form a heterogeneous group of lysosomal storage diseases with a prevalence of 1:14000 to 1:1000000 worldwide, depending on the region [2]. The clinical features of NCL include visual loss, seizures, ataxia, epilepsy and both mental and motor deterioration. Today, at least 14 affected genes are implicated, from *CLN1* to *CLN14*, 13 of which have been identified (*CLN1-8* and *CLN10-14*) [3]. In the Republic of Sakha (Yakutia), a case of neuronal ceroid lipofuscinosis type 6 was previously identified, the cause of which was a mutation c.396dupT in exon 4 of the *CLN6* gene. Given this fact, a study was continued to search for molecular genetic causes in patients with similar clinical signs. First of all, a mutation in the *CLN6* gene was excluded. Considering the severity of the course of the disease, the disability of patients, the fatal outcome and the lack of specific therapy, the study of neuronal ceroid lipofuscinosis in Yakutia is both a fundamental scientific problem and a medical and social task for the diagnosis and timely prevention of this hereditary pathology.

*Methods and Algorithms:* A total of 33 DNA samples from patients registered at the Medical Genetic Center of the National Center of Medicine (Yakutsk) with suspected NCL6 were studied. 3 patients were subjected to molecular genetic testing using exome sequencing using the TruSight Inherited Disease panel (Illumina, USA) on the MiSeq genetic sequencer (Illumina, USA). The results were validated by Sanger direct sequencing. The remaining 30 patients were analyzed with commercial kits made to order (TestGene, Russia) by real-time PCR.

*Results:* Of 33 patients, 26 cases had of neuronal ceroid lipofuscinosis type 6, in 7 patients a mutation in the *CLN6* gene was excluded, and are currently registered with a diagnosis of leukodystrophy. In our patients we identified a homozygous frameshift variant c.396dupT (p.Val133CysfsTer18) in 4 exon of *CLN6*, parents are heterozygous carriers of the mutation. All patients of the Yakut ethnic group. The clinical picture: the onset of the disease at the age of 3-4 years, impaired coordination, frequent falls, regression of psychomotor development, seizures, atrophy or subatrophy of the optic nerves, the development of dementia in the later stages.

*Conclusion:* In Yakutia, the most common form of neuronal ceroid lipofuscinosis is type 6, caused by a mutation c.396dupT in 4 exon of *CLN6* gene. Perhaps the reason for the accumulation of this disease is also the founder effect, as in other monogenic diseases

common in Yakutia, this requires further study. Also, in 7 patients, the study will be continued to clarify the diagnosis.

*Acknowledgements:* Molecular genetic studies were carried out on the basis of the Collective Use Center of the Arctic Innovation Center of NEFU named after M.K. Ammosov. The study is supported by the Ministry Education and Science of Russian Federation (Project No. FSRG-2020-0014 “Genomics of Arctic: epidemiology, hereditary and pathology”).

### *References*

1. Simonati A., Williams R.E. Neuronal ceroid lipofuscinosis: the multifaceted approach to the clinical issues, an overview. *Front Neurol.* 2022;13:811686.
2. Chin J.J., Behnam B., Davids M. et al. Novel mutations in CLN6 cause late-infantile neuronal ceroid lipofuscinosis without visual impairment in two unrelated patients. *Mol Genet Metab.* 2019;126:188-195.
3. Nita D.A., Mole S.E., Minassian B.A. Neuronal ceroid lipofuscinoses. *Epileptic Disord.* 2016;18(Suppl. 2):S73-S88.

## Exo-C is a new Hi-C-based method for all type genome structural variants detection

Gridina M.<sup>1\*</sup>, Stepanchuk Y.<sup>1,2</sup>, Nuriddinov M.<sup>1</sup>, Mozheiko E.<sup>1</sup>, Valeev E.<sup>1,2</sup>, Nazarenko L.P.<sup>3</sup>, Lopatkina M.E.<sup>3</sup>, Markova Z.G.<sup>4</sup>, Yablonskaya M.I.<sup>5</sup>, Voinova V.Y.<sup>5</sup>, Shilova N.V.<sup>4</sup>, Lebedev I.N.<sup>3</sup>, Fishman V.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

<sup>4</sup> *Research Centre for Medical Genetics, Moscow, Russia*

<sup>5</sup> *Clinical Research Institute of Pediatrics Named After Acad. Y.E. Veltishev, Moscow, Russia*

\* *gridinam@gmail.com*

**Key words:** Chromosomal rearrangements, Hi-C, Chromosome conformation capture, Exome sequencing, Structural variation detection

*Motivation and Aim:* High-throughput methods of genome analysis allow efficient detection of point mutations. However, their sensitivity for detecting chromosomal rearrangements, especially balanced ones, remains rather weak. As a result structural variations are largely understudied, compared to single nucleotide variants (SNV). The Hi-C method is widely used to study the spatial chromatin architecture and to assemble genomes; moreover, it is a powerful tool to detect balanced and unbalanced chromosomal rearrangements [1]. However, Hi-C data are not completely suitable for single nucleotide variant (SNV) detection. First of all, it is limited by a nonuniform genomic coverage, which is result of using restriction endonucleases in Hi-C-protocol. Second, to the reliably detect SNV in exome, the sequencing have to be significantly deeper than for conventional *in situ* Hi-C. Thus, the aim of the research was developing new Hi-C-based method for simultaneous detection of SNV in exome and different type of chromosomal rearrangements at a genome-wide level.

*Methods and Algorithms:* We developed a new method Exo-C. The first step of Exo-C libraries preparation was DNase Hi-C [2]. The second step was enrichment with exome sequences, which was performed using the KAPA HyperExome kit. Sequencing of obtained libraries was performed on Illumina or MGI platforms, in PE150 paired read mode, with a depth of 30-60 million reads per individual library.

ATo detect chromosomal rearrangements, we derived statistics from scaling of chromatin 3D contacts with genomic distance. We used joint analysis of coverage and contacts scaling distributions to find significant outliers corresponding to the rearranged loci. To detect SNV, we used standard approach based on GATK guidelines.

*Results:* We modified the existed protocol of DNase Hi-C to robustly obtain the good quality Hi-C data. As a result of using sequence-agnostic exonuclease we obtained more uniform genomic coverage than it would be possible using restriction enzymes. Applying the developed Exo-C protocol to the collection of human blood samples, we obtained three-dimensional contact maps for 45 individuals among which were patients with translocations (13) and inversions (6) previously detected by cytogenetic analysis. As a result, we successfully confirmed all translocations and four inversions by Exo-C method. The rearrangement structure was significantly refined in some cases.

The one of representative example is P47, the patient with multiple exostosis. The patient karyotype was 46XY, t(8;11;21)(q23;q22;q21), however Exo-C data allowed us to determine the chromosome 20 also affected by the rearrangement. This result of Exo-C assay, was further confirmed by Multicolor FISH. Moreover one of the breakpoints was inside the first intron of *EXT1* gene. Mutations in the *EXT1* gene lead to the formation of multiple exostoses; this pathology is inherited in an autosomal dominant manner. Thus, the results of Exo-C analysis allowed us to fully explain the phenotype of the patient and establish the molecular mechanism of the development of the disease.

The SNV detection was the next important task, which we would like resolve. The comparison of the list of SNVs identified by the Exo-C method and classical exome sequencing showed 99 % overlap at a similar sequencing depth. Thus, the Exo-C method makes it possible to find single nucleotide variants with an accuracy comparable to classical exome sequencing.

*Conclusion:* We developed Exo-C method, which allowed to obtain the information about balanced and unbalanced chromosomal rearrangements as well as clinically relevant SNV from a single experiment and did not require extremely deep sequencing.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant No. 21-65-00017).

### *References*

1. Harewood L., Kishore K., Eldridge M.D., Wingett S., Pearson D., Schoenfelder S., Collins V.P., Fraser P. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* 2017;27(18(1)):125.
2. Gridina M., Mozheiko E., Valeev E., Nazarenko L.P., Lopatkina M.E., Markova Z.G., Yablonskaya M.I., Voinova V.Y., Shilova N.V., Lebedev I.N., Fishman V. A cookbook for DNase Hi-C. *Epigenetics Chromatin.* 2021;14(1):15.

## Detailed phylogeny of the Y-chromosome haplogroup N1a2 in the populations of Siberia and Eastern Europe

Kharkov V.\* , Stepanov V.

*Research Institute of Medical Genetics, TNMRC, Tomsk, Russia*

\* [vladimir-kharkov@medgenetics.ru](mailto:vladimir-kharkov@medgenetics.ru)

**Key words:** human population genetics, haplogroups, phylogeny and phylogeography of Y-chromosome lineages

*Motivation and Aim:* The modern level of solving the fundamental problems of general human genetics and population genetics – the study of genetic diversity and the structure of the gene pools of human populations, the mechanisms of its formation, the reconstruction of migrations and human settlement, is most thoroughly implemented in the aspect of analysis of data from sequencing of complete genomes and phylogeographic and phylogenetic analysis of Y-chromosome haplogroups, using the latest advances in the discovery and genotyping of new SNP markers by massive parallel sequencing.

*Methods and Algorithms:* We carried out a detailed analysis of phylogeny of the Y-chromosome haplogroup N1a2 in various populations of Siberia and the Volga-Ural region using bioinformatics processing of whole genome sequencing data. The samples represent all indigenous peoples living in these territories and belong to different subclades previously identified by YSTR haplotypes. Samples were genotyped from Siberian (Khanty, Nenets, Siberian Tatars, Southern Altaians, Kumandins, Tubalars, Chelkans, Khakases (Sagais and Kachins), Shors, Chulyum Turks, Tuvans, Buryats (eastern and western), Evenks, Yakuts) and Eastern European populations (Komi, Udmurts, Mari, Bessermen, Mordovians, Kazan Tatars, Bashkirs). After analyzing genome-wide data, hundreds of terminal SNPs were selected for the most detailed separation of subhaplogroups. Some of these SNPs were completely genotyped on the entire array of population samples.

*Results:* According to the results of genotyping, the age of separation of various branches of all subclades was calculated, their component contribution to the gene pools of all the studied peoples and their genetic continuity between different populations were assessed. The phylogeny of all ethnospecific sublines of the haplogroup N1a2 is described in detail. Median networks of YSTR haplotypes were constructed, reflecting the expansion of the number of ancestral populations, which coincide very well with the data on SNP markers. The initial territory for the emergence of all basic subclades N1a2 is the Sayan-Altai region. Further, their separation occurred during the migrations of various ancestral groups in different time periods. The main connection between the distribution of different N1a2 subclades can be traced with representatives of the Samoyedic and Finno-Ugric languages. The frequency of different sublines of this haplogroup in different populations correlates well with the proportions of different genetic components in their gene pools, calculated based on the analysis of whole genome sequencing data. Almost all sublines show significant population specificity. The results of a comparative analysis of the spectrum of haplotypes convincingly indicate the specificity of Y-chromosome sublines and clusters of haplotypes not only at

the level of ethnic groups and populations but also at the level of phratries, tribes, and individual family groups.

*Conclusion:* Genotyping of a wide range of selected terminal YSNPs and YSTRs on the entire array of population samples allowed for the most accurate and detailed separation of all sublines between populations. This further clarifies the relationship of subhaplogroups with migration processes, founder effects in different populations, genetic components, demographic growth, and the continuity of the genetic contribution between different populations.

*Acknowledgements:* The study is supported by The Research Institute of Medical Genetics, Tomsk National Research Medical Center (122020200083-8).

## Genes for aryl hydrocarbon receptor signaling are associated with uterine fibroids susceptibility

Kobzeva K.\*, Polshvedkina O., Bobyleva L., Barysheva E., Polonikov A.,  
Bushueva O.

*Kursk State Medical University, Kursk, Russia*

\**kseniya.kobzeva.02@mail.ru*

**Key words:** uterine fibroids, single nucleotide polymorphism, aryl hydrocarbon receptor pathway, xenobiotic biotransformation

*Motivation and Aim:* Uterine fibroids (UF) are one of the most common benign tumors in females worldwide. UF are associated with a significant impairment of reproductive health and can lead to impaired fertility, pregnancy complications, and adverse obstetric outcomes [1]. In recent years, evidence has emerged indicating that environmental pollutants such as dioxins may be a risk factor for the development of UF [1]. Dioxins contribute to the formation of tumors by affecting cell proliferation and differentiation, and their negative impact on reproductive function is associated with strong toxic effects. These compounds exert their biological and toxicological effects by binding to the aryl hydrocarbon receptor (AhR) [2]. AhR is a ligand-dependent transcription factor that regulates cell differentiation and the induction of xenobiotic biotransformation enzymes. AhR nuclear translocator (ARNT) and AhR repressor (AhRR) regulate AhR function. The ligand-bound AhR moves to the nucleus, where it heterodimerizes with ARNT. The AhR-ARNT heterodimer binds to xenobiotic response elements (XRE) and promotes the activation of target genes involved in the biotransformation of xenobiotics and included in the aryl hydrocarbon receptor signaling pathway. All aryl hydrocarbon receptor signaling pathway genes are expressed in the uterus and are involved in the local response to highly toxic environmental pollutants [3]. A comprehensive analysis of the involvement of polymorphic variants of the aryl hydrocarbon receptor signaling pathway genes in predisposition to UF has not been carried out.

*Methods and Algorithms:* A total of 1001 unrelated Russian females from Central Russia were examined: 603 patients with UF and 398 healthy females, without clinical and ultrasound signs of uterine fibroids. Written informed consent was obtained from all participants prior to entering the study. The study was approved by the Regional Ethics Committee of Kursk State Medical University. Genotyping of single nucleotide polymorphisms (rs2066853 *AHR*, rs2292596 *AHRR*, rs2228099 *ARNT*, rs1048943 *CYP1A1*, rs762551 *CYP1A2*, rs1056836 *CYP1B1*, rs1800566 *NQO1*) was performed by TaqMan-based PCR on a CFX96 amplifier (Bio-Rad). To analyze the associations of genotypes with the development of UF, we used a log-additive regression model calculated in the SNPStats program (<https://www.snpstats.net/start.htm>) with an adjustment for age and smoking. To study the regulatory potential of SNPs, the atSNP online resource was used, which assesses the binding ability of transcription factors to DNA regions in the SNP region depending on the carriage of the reference/alternative allele (<http://atsnp.biostat.wisc.edu/search>). To search for gene ontologies associated with the biological functions of transcription factors, the Gene Ontology tool was used (<http://geneontology.org/>).

**Results:** This study showed that rs2228099 *ARNT* (OR = 1.58; 95 %CI = 1.19-2.09; risk allele G) and rs1048943 *CYP1A1* (OR = 0.59; 95 %CI = 0.40-0.91; protective allele G) are associated with development of UF. SNP rs2228099 *ARNT* is characterized by high regulatory potential. According to GTEx Portal data, the risk allele G of rs2228099 *ARNT* has a pronounced cis-eQTL-mediated effect on the expression of *HORMAD1*, *CERS2*, *CTSS*, *ARNT*, *RP11-54A4.2*, *SETDB1* genes in adipose subcutaneous, *CTSS*, *ARNT*, *RP11-54A4.2*, *HORMAD1* genes in adipose visceral, *CERS2* gene in uterus, *CTSK*, *GOLPH3L*, *HORMAD1*, *CTSS*, *SETDB1* genes in whole blood. The risk allele G also creates sites for DNA binding to 8 transcription factors (SMAD3, FOXA1, RREB1, EGR1, HIF1A::ARNT, SMAD, CREB3L1, EGR2), which are jointly involved in the regulation of transforming growth factor beta 2 production (FDR = 0.002), regulation of cytokine production (FDR = 0.014), regulation of vascular endothelial growth factor receptor signaling pathway (FDR = 0.01), positive regulation of vascular endothelial growth factor production (FDR = 0.01), positive regulation of interleukin-1 beta production (FDR = 0.02), response to hypoxia (FDR = 0.002). Protective allele G of SNP rs1048943 *CYP1A1* binds to 20 transcription factors (HOXA2, BSX, NR2C2, CUX2, ATF3, DLX5, NR3C1, HOXB5, HSF, JUN, NKX1-1, SP2, BARX1, DLX2, YY1, PPARA, JUND (var.2), NANOG, PDX1, HOXD3) co-participating in cell differentiation (FDR =  $1.55 \times 10^{-5}$ ), steroid hormone mediated signaling pathway (FDR = 0.037), response to steroid hormone (FDR = 0.04). Thus, genetic variants associated with the development of UF have a high regulatory potential and mediate participation in a large number of biological processes involved in the pathogenesis of uterine fibroids.

**Conclusion:** Thus, for the first time in the world, we have obtained data on the involvement of the genes for the aryl hydrocarbon receptor signaling pathway in the predisposition to UF.

**Acknowledgements:** The study is supported by Kursk State Medical University.

### References

1. Pavone D., Clemenza S., Sorbi F., Fambrini M., Petraglia F. Epidemiology and risk factors of uterine fibroids. *Best Pract Res Clin Obstet Gynaecol.* 2018;46:3-11.
2. Khorram O., Garthwaite M., Golos T. Uterine and ovarian aryl hydrocarbon receptor (AHR) and aryl hydrocarbon receptor nuclear translocator (ARNT) mRNA expression in benign and malignant gynaecological conditions. *Mol Hum Reprod.* 2002;8(1):75-80.
3. Tsuchiya M., Katoh T., Motoyama H., Sasaki H., Tsugane S., Ikenoue T. Analysis of the AhR, ARNT, and AhRR gene polymorphisms: genetic contribution to endometriosis susceptibility and severity. *Fertil Steril.* 2005;84(2):454-458.



## Effect of estrogen and progesterone on the frequency and spectrum of karyotypically abnormal cells in cultured uterine leiomyomas

Koltsova A.S. \*, Efimova O.A., Baranov V.S., Yarmolinskaya M.I., Polenov N.I., Pendina A.A.

*D.O. Ott Research Institute of Obstetrics, Gynecology and Reproductology, St. Petersburg, Russia*

\* *rosenrot15@yandex.ru*

**Key words:** uterine leiomyoma, chromosomal abnormalities, 7q deletion, estrogen, progesterone

*Motivation and Aim:* Uterine leiomyomas (ULs), also called “fibroids”, are the most common tumors of female reproductive tract [1]. In spite of their benign nature, ULs can rapidly grow to a large size and seriously affect women’s health. UL symptoms include pelvic pain, excessive bleeding, infertility and recurrent pregnancy loss. ULs are clonal sex steroid hormone-dependent tumors. Up to 50 % of ULs have chromosomal abnormalities, which are considered to occur secondary during UL tumorigenesis [2]. We recently demonstrated that cytogenetically different UL cell clones have an unequal ability to grow *in vivo* and *in vitro* [3, 4]. In the present study, we aimed to check whether combined and isolated estrogen and progesterone treatment has an effect on both clonal spectrum and clonal proportion in ULs with an abnormal karyotype *in vitro*.

*Methods and Algorithms:* The experimental approach included conventional karyotyping of ULs cultured in the presence of  $10^{-8}$  M estrogen and/or  $10^{-6}$  M progesterone and without hormones (control) with subsequent verification of the detected chromosomal abnormalities by metaphase and interphase FISH.

*Results:* We detected karyotype abnormalities in 11 out of 15 ULs cultured without hormones (73 %). In four out of 15 ULs (27 %), no chromosomal abnormalities were found in all cultured samples. The revealed abnormalities were represented exclusively by chromosomal rearrangements, among which interstitial deletions in the long arm of chromosome 7 (del(7q)) were prevalent ( $n = 6$ ; 40 % of ULs). Rearrangements involving chromosomes 1, 2, 3, 4, 5, 6, 8, 10, 12, 13, 16, 17, 19, and 21 were also observed. In all ULs with abnormal karyotype, more than one clone was detected: two clones with abnormal karyotype were detected in one UL while in the remaining ten ULs clone(s) with abnormal karyotype were combined with 46,XX cells. The number of chromosomally different clones in ULs varied from two to five. A total of twenty different chromosomal abnormalities were detected in 11 ULs. In nine of 11 ULs (82 %) with abnormal karyotype, the spectrum of clones detected in hormone-treated samples differed from the controls. The detected changes were unique for each UL, however, the widest variety of chromosomal abnormalities was detected in control samples as well as in samples cultured using combined estrogen and progesterone supplementation.

We selected six ULs with the most common rearrangement – deletion in 7q – and analyzed the frequency of del(7q) in 1000 cells of each cultured sample using interphase FISH with DNA probes to chromosome 7 (7cen, 7q22, 7q31.2). All samples comprised cells both with and without del(7q). Significant changes in the frequency of del(7q) cells after hormone supplementation were detected in all ULs. Both isolated and combined

estrogen and progesterone supplementation had similar effect on the frequency (either up or down shift) of each clone. Among ULs with one del(7q) clone, two demonstrated significantly increased frequency of del(7q) cells (del(7)(q21q32), del(7)(q21q31),t(2;5)(p13;p13)) after hormone supplementation, while in the rest two ULs, del(7q) clones (del(7)(q21q31), del(7)(q21q31),inv(12)(q15q24)) significantly decreased in number ( $p < 0.05$ , Chi-square with Yates' correction). One UL with two unrelated deletions in 7q – del(7)(q21q31) and del(7)(q21q36) – showed significantly decreased frequency of both clones after hormone supplementation ( $p < 0.05$ , Chi-square with Yates' correction). In contrast, clone with del(7)(q11.23q31) decreased in number while clone with 46,XX,del(7)(q21q36) became significantly more numerous after hormone supplementation in other UL ( $p < 0.05$ , Chi-square with Yates' correction). Notably, for six out of eight del(7q) clones, the changes were more pronounced after isolated but not combined hormone supplementation.

*Conclusion:* Estrogen and progesterone treatment caused changes in the spectrum of chromosomal rearrangements detected by conventional karyotyping in 82 % of ULs. Interstitial deletions in the long arm of chromosome 7 were the most frequent chromosomal abnormalities ( $n = 6$ ; 55 %). Interphase FISH analysis showed that both isolated and combined estrogen and progesterone supplementation had the same effect on the frequencies of each del(7q) clone in cultured ULs. However, the frequencies of different del(7q) clones may shift up or down in cultured ULs. In most ULs, isolated hormone treatment has more pronounced effect on the frequency of del(7q) cells than combined one. An important finding in our study is that ULs consist of a variety of cell populations, which differ not only karyotypically, but also by their response to sex steroids.

*Acknowledgements:* The study is supported by RSF (grant No.19-15-00108). A.S. Koltsova is a grantee of RF President Scholarship (CII-3297.2022.4).

### *References*

1. Stewart E.A. et al. Epidemiology of uterine fibroids: a systematic review. *BJOG*. 2017;124(10):1501-1512.
2. Baranov V.S. et al. Pathogenomics of uterine fibroids development. *Int J Mol Sci*. 2019;20(24):6151.
3. Koltsova A.S. et al. Uterine leiomyomas with an apparently normal karyotype comprise minor heteroploid subpopulations differently represented *in vivo* and *in vitro*. *Cytogenet Genome Res*. 2021;161(1-2):43-51.
4. Koltsova A.S. et al. Cytogenomic profile of uterine leiomyoma: *in vivo* vs. *in vitro* comparison. *Biomedicines*. 2021;9(12):1777.

## Distribution of total cholesterol and glucose parameters in a sample of Tuvans in Kyzyl

Kovalev S.<sup>1\*</sup>, Churkina T.<sup>1</sup>, Tabikhanova L.<sup>1</sup>, Hantemirova M.<sup>1</sup>, Kawai-Ool U.<sup>2</sup>, Saryglar A.<sup>1</sup>, Osipova L.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Tuvan State University, Kyzyl, Russia*

\* *sergey.kovalev.1994@list.ru*

**Key words:** population genetics, dyslipidemia, hyperglycemia, multifactorial diseases

*Motivation and Aim:* Elevated blood cholesterol and glucose levels have a broad etiology and are one of the major causal risk factors for cardiovascular disease (CVD). Urbanization inherent in the Republic of Tyva, in particular the city of Kyzyl, is one of the main causes of chronic diseases in the population [1]. The interaction of genetic factors with lifestyle has a significant impact on the prevalence of obesity, diabetes, cardiovascular disease and other risk factors [2]. Based on the total mortality in the Republic of Tyva, the proportion of deaths from cardiovascular diseases is 32.6%. There is also an increase in the number of cases of coronary heart disease among the young population [3]. The study of the features of multifactorial diseases in the indigenous population and one of the risk factors predisposing to it, such as obesity, will determine the main ways to prevent these diseases and improve specialized medical care in the Republic of Tyva.

*Methods and Algorithms:* The study involved residents of the Republic of Tyva of both genders ( $n = 397$ ) with a female predominance of 72.29% ( $n = 287$ ). Men accounted for 27.71% ( $n = 110$ ). Age ranged from 11 years to 83 years. People were divided into 6 age groups with a predominance of people under the age of 21–30.23% ( $n = 120$ ). The average age was 37.2 years. Only persons who indicated the presence of relatives of Tuvan nationality up to the second generation during the questionnaire participated in the current research. The determination of cholesterol and glucose in the blood was carried out using portable express blood analyzer "CardioChek". The capillary blood was used as the research material. A statistical analysis of the income data was performed.

*Results:* When analyzing total cholesterol in the blood, an average value of 4.63 and 3.81 mmol/l was found for men and women respectively. The average for both sexes is 4.4 mmol/l. When determining hyper- and hypocholesterolemia, it was found that among women the proportion with hypercholesterolemia is 18 people (7%), and with hypocholesterolemia the proportion is 27 people (10%). For men, the number of persons with hypocholesterolemia is 41 people (37%). Cases of hypercholesterolemia in men group were not observed. As to hyper- and hypoglycosemia the results had shown its existence among 31 people (11%) and 3 people (1%) relatively in women group, and 8 (7%) and 1 (1%) in men representatives respectively. The number of people with a deviation in the level of glucose and cholesterol in the blood in the total score was 10.83 and 21.66% respectively.

*Conclusion:* Analysis of the data showed that the residents of the Republic of Tuva are more likely to have hyperglycosemia in women representative group, predominantly

over 46 years, and hypocholesterolemia in both genders, predominantly under 30 years. When comparing data with such northern ethnic groups as Evenks, Evens, Dolgans, Yukaghirs, Yakuts, it was found that the average level of total cholesterol in the blood of the inhabitants of Tuva (4.4 mmol/l) is lower [4]. Among other indigenous peoples of Siberia, the Khanty (4.46 mmol/l), as well as the Nenets and Komi (4.44 mmol/l), leading a sedentary, non-nomadic lifestyle, have a similar indicator [5, 6].

*Acknowledgements:* Research was supported by the budget project FWNR-2022-0021.

### *References*

1. Kavaj-ool U.N., Savost'yanov A.N., Seden A.N., SHozhul-Ool E.D. Food culture is the basis of human health in Tuva. *Bull Tuva State University Natural Agricultural Sci.* 2015;2:34-39.
2. Soyun Sh.Ch. The current demographic situation in Tuva. In: Proceedings of the conference Geography of Tuva education and science. 2016;101-104.
3. Shirizhik A.T., Balabina N.M. Risk factors for the development and prognosis of coronary heart disease in the indigenous population of the Republic of Tyva. *Sib Med J.* 2011;100(1):147-148.
4. Klimova T.M., Fedorova V.I., Baltahinova M.E., Krivoshapkin V.G. Lipid profile and dyslipoproteinemia in the indigenous rural population of Yakutia. *Sib Med J.* 2012;27(3):142-146.
5. Darenskaya M.A. Features of metabolic reactions in indigenous and alien populations of the North and Siberia. *Acta Biomed Sci.* 2014;2(96):97-103.
6. Sevost'yanova E.V. Features of human lipid and carbohydrate metabolism in the North. *Bull Sib Med.* 2013;12(1):93-100.

## Polymorphism in genes involved in glutathione metabolism is associated with the risk of uterine fibroids

Kudryavtseva O.<sup>1,2\*</sup>, Kobzeva K.<sup>1</sup>, Vdovina I.<sup>1,2</sup>, Ponomareva L.<sup>1</sup>, Polonikov A.<sup>1</sup>, Bushueva O.<sup>1</sup>

<sup>1</sup> *Kursk State Medical University, Kursk, Russia*

<sup>2</sup> *Kursk Regional Clinical Hospital, Kursk, Russia*

\* *kudrolga.75@mail.ru*

**Key words:** uterine fibroids, glutathione metabolism, candidate susceptibility genes, *GSR*, *SLC7A11*

*Motivation and Aim:* Uterine fibroids (UF) is a monoclonal benign tumor derived from the smooth muscle cells of the myometrium. UF is diagnosed in 35-70 % of women of reproductive age, while a third of patients with UF have severe clinical symptoms, which often leads to a significant deterioration in their quality of life and a decrease in reproductive potential [1]. Currently, a significant role in the development of UF is attributed to disturbances in redox homeostasis in myometrial tissues, leading to chronic oxidative stress [2]. One of the most powerful natural antioxidants is glutathione [3]. Depletion of its intracellular pool due to the inheritance of functionally defective alleles of genes encoding enzymes involved in its metabolism leads to free radical damage to the myometrium, initiation and maintenance of the activity of proliferative processes in the tissues of the uterus. The aim of our study was to analyze the role of polymorphic variants of genes involved in glutathione metabolism in the development of UF.

*Methods and Algorithms:* A total of 889 unrelated Russian individuals from Central Russia were examined: 552 patients with UF and 337 healthy females, without clinical and ultrasound signs of uterine fibroids. Written informed consent was obtained from all participants prior to entering the study. The study was approved by the Regional Ethics Committee of Kursk State Medical University.

Genotyping of single nucleotide polymorphisms of rs713041 *GPX4*, rs4902346 *GPX2*, rs41303970 *GCLM*, rs17883901 *GCLC*, rs1050450 *GPX1*, rs1799811 and rs1695 *GSTP1*, rs2551715 *GSR*, rs1801310 *GSS*, rs4820599 *GGT1*, rs7674870 *SLC7A11* was performed by TaqMan-based PCR on a CFX96 amplifier (Bio-Rad). Genotyping of deletion polymorphisms (+/0 *GSTMI* and +/0 *GSTTI*) was carried out by multiplex PCR. All the genes selected for the study are involved in the metabolism of glutathione and are expressed in myometrium. To analyze the associations of genotypes with the development of UF, we used a log-additive regression model calculated in the SNPStats program (<https://www.snpstats.net/start.htm>) with an adjustment for age and smoking.

*Results:* Comparative analysis of genotype frequencies revealed a protective effect of rs2551715 *GSR* on the risk of UF: OR = 0.83, 95 %CI = 0.70–0.99, P = 0.04. It is well-known that the *GSR* gene encodes the enzyme glutathione reductase involved in the reduction of the active form of glutathione [5]. As shown by functional studies, the C allele, compared with the minor T allele, reduces the activity of the *GSR* enzyme, which can lead to the accumulation of free radicals and increased oxidative stress in the tissues of the uterus due to the low rate of recovery of the active form of glutathione. We also found an association of rs7674870 *SLC7A11* with an increased risk of UF: OR = 1.25, 95 %CI = 1.03-1.50, P = 0.02. We found no functional studies of rs7674870 *SLC7A11*.

However, since this SNP is located in the 3'-untranslated region of the gene, it may affect transcriptional regulation, mRNA stability, or microRNA binding. For interpretation of the functional effects of rs7674870 *SLC7A11* bioinformatic analysis has been done. Analysis of expression quantitative trait loci (eQTL) using the GTEx Portal resource (<https://gtexportal.org/home/>) revealed that the risk allele G rs7674870 is associated with a decrease in the expression of *SLC7A11* in the blood ( $P = 0.012$ ,  $NES = -0.10$ ). Transcription factor binding analysis found that the protective allele A rs7674870 *SLC7A11* binds to 16 transcription factors (TFAP2A, NFIC, SP2, PAX9, RUNX1, RUNX3, RUNX2, PAX1, TFAP2, TFAP2C, NFIB, TFAP2A, DMRTC2, RAD21, RUNX3, E2F4), which are jointly involved in the regulation of cell population proliferation (GO:0042127;  $FDR = 6.57 \times 10^{-3}$ ). At the same time, the risk allele G binds to 8 transcription factors (MAFK, ATF1, AP1, SPDEF, SMAD3, MAFB, ATF1, AP1), however, no common biological processes were found for this transcription factors. *Conclusion:* Thus, we have carried out a comprehensive analysis of the involvement of glutathione metabolism genes in predisposition to uterine fibroids. An association of single nucleotide polymorphisms rs7674870 *SLC7A11* and rs2551715 *GSR* with the risk of uterine fibroids was found for the first time.

*Acknowledgements:* The study is supported by Kursk State Medical University.

### References

1. Stewart E.A., Cookson C.L., Gandolfo R.A., Schulze-Rath R. Epidemiology of uterine fibroids: a systematic review. *BJOG*. 2017;124(10):1501-1512.
2. Fletcher N.M., Abusamaan M.S., Memaj I. et al. Oxidative stress: a key regulator of leiomyoma cell survival. *Fertil Steril*. 2017;107(6):1387-1394.e1.
3. Bansal A., Simon M.C. Glutathione metabolism in cancer progression and treatment resistance. *J Cell Biol*. 2018;217(7):2291-2298.
4. Kudryavtseva O.K., Barysheva E.M., Sorokina M.V., Polshvedkina O.B., Ivanova N.V., Polonikov A.V., Bushueva O.Y. The analysis of relationship between the A/G (rs1801310) polymorphism of the GSS gene and the development of uterine myoma: a pilot study. *Med Genet*. 2017;16(2):37-39. (in Russian)
5. Kelner M.J., Montoya M.A. Structural organization of the human glutathione reductase gene: determination of correct cDNA sequence and identification of a mitochondrial leader sequence. *Biochem Biophys Res Commun*. 2000;269(2):366-368. doi: 10.1006/bbrc.2000.2267.

## Alternative splicing of human *FMRI* gene in peripheral blood: making a single-gene transcriptome

Lan F.H.<sup>1,2\*</sup>, Qiu P.<sup>3</sup>, Yan A.Z.<sup>1,2</sup>, Xu Y.J.<sup>1,2</sup>, Fu X.G.<sup>4</sup>, Yang W.J.<sup>5</sup>, Liao J.<sup>6</sup>, Guo X.Y.<sup>7</sup>, Zhang D.<sup>1,2</sup>, Zheng D.Z.<sup>1,2</sup>

<sup>1</sup> Center for Medical Genetics, Dongfang Hospital, Xiamen University School of Medicine, Fuzhou, China

<sup>2</sup> Fujian Key Laboratory of Transplantation Biology, Fuzong Clinical College, Fujian Medical University, Fuzhou, China

<sup>3</sup> Department of Laboratory Medicine, Yichang Central Hospital, Yichang, Hubei Province, China

<sup>4</sup> Central Laboratory, Ningde Municipal Hospital, Ningde, Fujian Province, China

<sup>5</sup> Department of Laboratory Medicine, Zhongshan Hospital of Fudan University, Shanghai, China

<sup>6</sup> Department of Laboratory Medicine, Fujian People's Hospital, Fujian University of Traditional Chinese Medicine, Fuzhou, China

<sup>7</sup> Department of Laboratory Medicine, Fuzhou Second Municipal Hospital, Xiamen University School of Medicine, Fuzhou, China

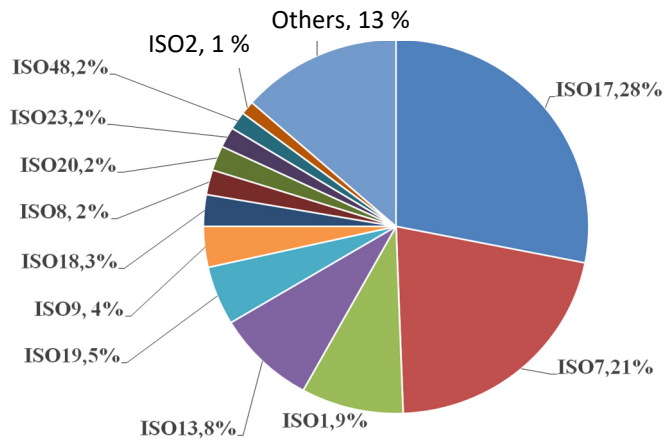
\* fhlana@xmu.edu.cn

**Key words:** alternative splicing, *FMRI* gene, transcriptome, fragile X syndrome

**Background and Aim:** Alternative splicing represents an important mechanism by which a eukaryotic organism diversifies the products from each single gene and aid a further layer to the network of gene regulation [1]. Fragile X mental retardation 1 gene (*FMRI*), the disease-determining gene of fragile X syndrome (FXS), which encodes the fragile X mental retardation protein (FMRP), is one of first human genes having the most abundant spliced products [2-6], with up to 20 FMRP isoforms being found before us [2]. The present study aims to further identify more spliced variants from *FMRI* gene in human peripheral blood, and the significance of single-gene transcriptome would be discussed. **Methods and Algorithms:** RT-PCR and clone-sequencing technology was employed to detect the types of alternatively spliced products in the peripheral blood from 10 normal individuals and their expression abundance was analyzed. Bioinformatics to analyze the splice sites and predict the structure of new splice isoforms was carried out by visiting the corresponding open-source servers. Potential functions of a splice isoform were initially studied by its lentiviral vector-based over-expression in HEK293T cells, followed by tracking the expression patterns of relevant genes.

**Results:** A total of 50 alternatively spliced products, including 27 novel ones, were identified from 505 recombinant plasmids of PCR product. The alternatively spliced products with the highest abundance are ISO 17 and ISO 7, both of which involve exon 12 skipping, followed by ISO 1 (containing the full-length sequence of the coding region) and ISO 13 (using the second acceptor site of exon 17), while the frequency of other types of spliced products is less than 5.0 % (Fig. 1). The novel alternatively spliced variants identified in this study involved the skipping of exons 3, 4, and 11, insertion of 30bp and 87bp sequences from the middle of intron 7, and 46 bp and 140bp sequences from the middle of intron 9. Bioinformatics analysis of the splice sites was performed, justifying the splicing of these variants. The structure of three novel spliced isoforms ISO51-ISO53 was predicted by I-TASSER server, showing that the intracellular location of truncated FMRP protein encoded by the novel transcripts was

altered. Over-expression of one novel FMRP isoform encoded by a spliced transcript with the insertion of the 46bp from amid intron 9 of human *FMR1* gene demonstrated that xCT might be a new interacting protein of FMRP.



**Fig. 1.** Expression abundance of spliced variants from human *FMR1* gene in peripheral blood

**Conclusion:** This study not only enriched our understanding of the complexity and diversity of alternative splicing of *FMR1* gene, but also laid the foundation for exploring the effects of alternative splicing on the function of FMRP and the pathogenesis of FXS.

**Acknowledgements:** This study was supported by Key Scientific Projects of Fujian Province, P. R. China (2010Y0045, 2016Y0071) and Fujian Provincial Natural Science Foundation of China (2018J01341).

### References

1. Blencowe B.J. Alternative splicing: new insights from global analyses. *Cell*. 2006;126(1):37-47.
2. Sittler A. et al. Alternative splicing of exon 14 determines nuclear or cytoplasmic localization of *fmr1* protein isoforms. *Hum Mol Genet*. 1996;5(1):95-102.
3. Fu X. et al. Alternatively spliced products lacking exon 12 dominate the expression of fragile X mental retardation 1 gene in human tissues. *Mol Med Rep*. 2015;12(2):1957-1962.
4. Tseng E. et al. Altered expression of the *FMR1* splicing variants landscape in premutation carriers. *Biochim Biophys Acta Gene Regul Mech*. 2017;1860(11):1117-1126.
5. Fu X.G. et al. Splicing of exon 9a in *FMR1* transcripts results in a truncated FMRP with altered subcellular distribution. *Gene*. 2020;731:144359.
6. Yang W.J. et al. Further identification of a 140bp sequence from amid intron 9 of human *FMR1* gene as a new exon. *BMC Genet*. 2020;21:63-72.



## Chronic disease markers analysis using high-throughput sequencing approaches

Li H.<sup>1\*</sup>, Chen W.-L.<sup>2</sup>, Sekacheva M.I.<sup>3</sup>, Cai G.<sup>4</sup>, Orlov Y.L.<sup>3,5\*</sup>

<sup>1</sup> Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup> Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China

<sup>3</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

<sup>4</sup> University of South Carolina, Columbia, USA

<sup>5</sup> Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia

\* [kaikaixinxin@sjtu.edu.cn](mailto:kaikaixinxin@sjtu.edu.cn); [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** disease markers, chronic disease, sequencing, bioinformatics, database

*Motivation and Aim:* With the recent development of sequencing technology and the rapid reduction of sequencing costs, high-throughput sequencing is revolutionizing basic life science research and clinical research from various aspects. High-throughput sequencing produces millions of sequencing reads at a time, and the alignment or assembly of these reads allows determination of various mutations at the genomic level, accurate gene expression quantification at the transcriptomic level, and identification of histone or DNA modification at the epigenomic level. The resulting accumulation of multi-omics information has opened up a new era of finding effective disease markers and studying their roles in disease occurrence and development [1, 5]. We aimed collect papers on biomarkers in a thematic journal issue to highlight recent findings in this field based on sequencing.

*Methods and Algorithms:* We have organized special journal issue (Research Topic) “High-throughput sequencing-based investigation of chronic disease markers and mechanisms” at *Frontiers in genetics* journal focuses on the genomics studies on cancer and chronic diseases (<https://www.frontiersin.org/research-topics/21036/high-throughput-sequencing-based-investigation-of-chronic-disease-markers-and-mechanism>). Using high-throughput sequencing, various markers of chronic diseases (such as cancer, heart disease, diabetes, and arthritis) have been developed at all omics levels, which have been used for diagnosis and classification of diseases, prediction of treatment effects, and prevention of diseases [6]. This research topic aimed at developing new chronic disease markers at four levels (i.e., genome, epigenome, transcriptome, and translome) with the help of high-throughput sequencing, and delineating potential marker-related mechanisms for chronic disease development. More specifically, this research topic contains contributions including:

1. Identification of novel biomarkers and prediction signatures for chronic disease detection (especially in early-stage) or prognosis prediction using high-throughput sequencing;
2. Analysis the possible pathological causes of markers as well as the potential roles they play in disease initiation and development;
3. Applications of new high-throughput sequencing techniques facilitating the development of more effective biomarkers of chronic disease;

4. New algorithms or tools for *in silico* identification of effective chronic disease markers based on high-throughput sequencing data.

**Results:** In this Research Topic a total of 14 papers could be arranged by two main areas – the cancer studies, and the works on the other chronic diseases such as allergy. Lung cancer is one of the leading causes of cancer-associated death in the world. This Topic complements recent Research Topics “Bioinformatics of Genome Regulation” and “Association between Individuals’ Genomic Ancestry and Variation in Disease Susceptibility” in *Frontiers in Genetics* considered more genetic background rather than molecular mechanisms of the human diseases [3, 4]. Non-small cell lung cancer comprises about 85 % of all lung cancers challenging treatment problems. We collected at the Research Topic the group of papers on lung cancer [2]. Traditional Chinese medicine formula Feiyanning has been clinically administered in China for more than a decade and raised attention due to its anticancer effect in lung cancer. Wang et al. (2021) revealed the migration-associated genes involved in antitumor effects of herbal medicine Feiyanning on lung cancer cells using RNA-Seq and ATAC-Seq analysis. Recent works presented novel biomarkers in gastric cancer [8].

**Conclusion:** Overall, we have compiled several special journal issued on biomedical problems [3, 4]. We hope that the readers will find these papers collections a stimulating reading and consider participation in the next conferences and journal issues in this area (“Medical Genetics, Genomics and Bioinformatics–2022” – [https://www.mdpi.com/journal/ijms/special\\_issues/Medical\\_Genetics\\_2022](https://www.mdpi.com/journal/ijms/special_issues/Medical_Genetics_2022); [https://www.mdpi.com/journal/genes/special\\_issues/Transcriptional\\_Regulation\\_Tumor](https://www.mdpi.com/journal/genes/special_issues/Transcriptional_Regulation_Tumor)).

**Acknowledgements:** This publication has been supported by the RUDN University Scientific Projects Grant System, project No. R.3-2022-ins).

## References

1. Anashkina A.A., Leberfarb E.Y., Orlov Y.L. Recent Trends in Cancer Genomics and Bioinformatics Tools Development. *Int J Mol Sci.* 2021;22:12146. doi: 10.3390/ijms222212146.
2. Chang Y., Wang Y., Li B. et al. Whole-Exome Sequencing on Circulating Tumor Cells Explores Platinum-Drug Resistance Mutations in Advanced Non-small Cell Lung Cancer. *Front Genet.* 2021;12:722078. doi: 10.3389/fgene.2021.722078.
3. Das R., Tatarinova T.V., Galieva E.R., Orlov Y.L. Editorial: Association between individuals' genomic ancestry and variation in disease susceptibility. *Front Genet.* 2022;13:831320. doi: 10.3389/fgene.2022.831320.
4. 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.
5. Orlov Y.L., Baranova A.V., Hofstaedt R., Kolchanov N.A. Computational genomics at BGRS\SB-2016: introductory note. *BMC genomics.* 2016;17(Suppl. 14):996. doi: 10.1186/s12864-016-3350-6.
6. Voronova V., Glybochko P., Svistunov A. et al. Diagnostic value of combinatorial markers in colorectal carcinoma. *Front Oncol.* 2020;10:832. doi: 10.3389/fonc.2020.00832.
7. Wang C., Li P., Peng Y. et al. Integrative RNA-Seq and ATAC-Seq analysis reveals the migration-associated genes involved in antitumor effects of herbal medicine feiyanning on lung cancer cells. *Front Genet.* 2021;12:799099. doi: 10.3389/fgene.2021.799099.
8. Wang Y., Fang Y., Zhao F. et al. Identification of GGT5 as a novel prognostic biomarker for gastric cancer and its correlation with immune cell infiltration. *Front Genet.* 2022;13:810292. doi: 10.3389/fgene.2022.810292.

## Screening of the pathogenic variant c.3751dupA of the *BRCA2* gene in women from the Republic of Bashkortostan

Mingazheva E.T.<sup>1\*</sup>, Valova Ya.V.<sup>1,2</sup>, Andreeva E.A.<sup>1</sup>, Korosteleva A.V.<sup>1</sup>,  
Harina O.K.<sup>1</sup>, Shaydullina A.D.<sup>1</sup>, Prokofieva D.S.<sup>1</sup>, Khusnutdinova E.K.<sup>1,3</sup>

<sup>1</sup> Bashkir State University, Department of Genetics and Fundamental Medicine, Ufa, Russia

<sup>2</sup> Ufa Scientific Research Institute of Occupational Medicine and Human Ecology, Ufa, Russia

<sup>3</sup> The Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Center, RAS, Ufa, Russia

\* Elvira.F91@mail.ru

**Key words:** ovarian cancer, targeted sequencing, candidate genes

**Motivation and Aim:** Ovarian cancer (OC) is one of the most common malignant neoplasms, the incidence of which has remained consistently high over the past decades. According to the World Health Organization (WHO), in 2020, 313,959 new cases of ovarian cancer were registered in the world, and the number of deaths from this pathology was 207,252 people [1]. According to Kaprin et al. (2021), more than 13,000 cases of this pathology are diagnosed annually in Russia, about 8,000 of which are fatal [2]. One of the important risk factors for OC is hereditary predisposition. The high risk of this pathology is primarily associated with mutations in the tumor suppressor genes *BRCA1* and *BRCA2*. As a result of previous targeted sequencing of DNA samples from patients with hereditary ovarian cancer in exon 11 of the *BRCA2* gene, we identified the pathogenic variant c.3751dupA, which leads to a frameshift (p.Thr1251AsnfsTer14).

The aim of this work was to screen the c.3751dupA variant of the *BRCA2* gene in a group of OC patients and individuals in the control group. **Materials and Methods:** DNA samples isolated from venous blood of patients with hereditary forms of OC ( $n = 70$ ), sporadic ovarian cancer ( $n = 167$ ) and women without cancer at the time of blood sampling ( $n = 322$ ) from the Republic of Bashkortostan were used as research material. Genotyping was performed by high resolution melting curve analysis (HRM).

**Results:** In our study among patients with OC and the control group, the c.3751dupA/*BRCA2* variant was identified in one patient with a diagnosis of hereditary ovarian cancer (0.01 %). By ethnicity, the carrier of the studied variant is a Tatar. The manifestation of the disease occurred in postmenopausal women. The carrier of the studied variant by ethnicity is a Tatar. A patient with this pathogenic variant was diagnosed with stage III adenocarcinoma. The tumor was detected in the postmenopausal period.

**Conclusion:** Thus, our data indicate a low frequency of occurrence of the c.3751dupA/*BRCA2* variant in women from the Republic of Bashkortostan.

**Acknowledgements:** The study was supported by the FASO Bioresource Collections Development Program. The work was supported by RFBR grants No. 20-34-90003; Republic of Bashkortostan No. 210/1; President of the Russian Federation No. MK-3208.2022.1.4; State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. 075-03-2021-193/5.

### 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. Kaprin A.D., Starinskiy V.V., Shakhzadova A.O. (Eds.). Malignant neoplasms in Russia in 2020 (morbidity and mortality). Moscow: Ministry of Health of Russia, 2021.

## Application of a novel k-mer primer design algorithm for detecting antibiotic resistance determinants

Nurmukanova V.\*, Stetsenko I., Say A., Ayginin A., Mikhaylov I., Matsvay A., Shipulin G.

*Centre for Strategic Planning and Management of Biomedical Health Risks of the Federal Medical Biological Agency, Moscow, Russia*

\* *VNurmukanova@cspmz.ru*

**Key words:** antibiotic resistance, drug resistance determinants, multiplex primer panels, k-mer algorithm, genome sequencing

*Motivation and Aim:* Last decades the constant expansion of multi-drug resistant microorganism strains became an important problem [2, 4, 7, 9]. As a result, antibiotic resistance leads to higher medical costs, prolonged hospital stays, and increased mortality due to reduced antibiotics effectiveness, including those against microorganisms causing common infectious diseases (such as pneumonia, tuberculosis etc.) [2, 4, 7, 9]. However, modern molecular methods of infectious disease diagnostics make it possible to provide timely and effective profiling of the resistance of an infectious agent, and thereby expand and enhance means of clinical analysis [2, 3, 5, 6, 8].

*Methods and Algorithms:* In the present work, a novel *k*-mer algorithm was used to generate a panel of degenerate multiplexable primers for NGS-based detection of known genetic determinants of drug resistance to colistin, polymyxin, and  $\beta$ -lactam antibiotics. The reference database, used for primer design, was based on available open sources (CARD, NDARO and MEGARes) and contained information on 6772 loci associated with the resistance of bacterial pathogens to these groups of antibiotics. We have designed a primer panel suitable for analysis using Illumina (Miseq) platform as well as Nanopore (ONT MinION) sequencing which allowed to identify any all known to date groups of genes of interest in a single study. We then analyzed 62 DNA samples isolated from antibiotic-resistant bacteria of the genera *Acinetobacter*, *Escherichia*, *Enterobacter*, *Pseudomonas*, *Serratia*, *Klebsiella*, found in clinical biological material. The studied isolates were obtained from hospitalized patients with laboratory-confirmed manifestations of infection in various localizations, including CNS infection, bloodstream infection, urinary tract infection, intra-abdominal infection, bone and joint infection, pneumonia, skin and soft tissue infection and primary bacteremia.

*Results:* Using the newly developed profiling method, in the studied set of biological samples we identified genes belonging to the following groups: *bla*OXA, *bla*TEM, *bla*GES, *bla*ADC, *bla*NDM, *bla*CTX-M, *bla*VIM, *bla*CMY, *bla*PER, *bla*SHV, *bla*KPC, *bla*IMP, *bla*DHA, *bla*CRT, *bla*SST, *bla*ACT, and *mcr*-1. All results were confirmed with available qPCR assays and/or whole genome sequencing.

*Conclusion:* A method based on multiplex PCR with subsequent next generation sequencing for identifying the genetic determinants of drug resistance was developed. This method is more informative than qPCR analysis, much cheaper and standardizable than whole genome sequencing.

## References

1. Blanquart F. Evolutionary epidemiology models to predict the dynamics of antibiotic resistance. *Evol Appl.* 2019;12(3):365-383. doi: 10.1111/eva.12753.
2. Didelot X., Bowden R., Wilson D.J., Peto T.E.A., Crook D.W. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* 2012;13(9):601-612. doi: 10.1038/nrg3226.
3. Franklin A.M., Brinkman N.E., Jahne M.A., Keely S.P. Twenty-first century molecular methods for analyzing antimicrobial resistance in surface waters to support One Health assessments. *J Microbiol Methods.* 2021;184:106174. doi: 10.1016/j.mimet.2021.106174.
4. Larsson D.G.J., Flach C.F. Antibiotic resistance in the environment [published online ahead of print, 2021 Nov 4]. *Nat Rev Microbiol.* 2021;1-13. doi: 10.1038/s41579-021-00649-x.
5. Leopold S.R., Goering R.V., Witten A., Harmsen D., Mellmann A. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J Clin Microbiol.* 2014;52(7):2365-2370. doi: 10.1128/JCM.00262-14.
6. Liotti F.M., Posteraro B., Mannu F., Carta F., Pantaleo A., De Angelis G., Menchinelli G., Spanu T., Fiori P.L., Turrini F., Sanguinetti M. Development of a multiplex PCR platform for the rapid detection of bacteria, antibiotic resistance, and candida in human blood samples. *Front Cell Infect Microbiol.* 2019;9:389. doi: 10.3389/fcimb.2019.00389.
7. World Health Organization report. Antimicrobial resistance: global report on surveillance. 2014.
8. Sigmund I.K., Renz N., Feihl S. et al. Value of multiplex PCR for detection of antimicrobial resistance in samples retrieved from patients with orthopaedic infections. *BMC Microbiol.* 2020;20:88. doi: 10.1186/s12866-020-01741-7.
9. Talebi B.A., Rizvanov A.A., Haertlé T. et al. World Health Organization report: Current crisis of antibiotic resistance. *BioNanoSci.* 2019;9:778-788. doi: 10.1007/s12668-019-00658-4.

## Possibilities of molecular genetic research in the diagnosis of MODY diabetes

Ovsyannikova A.\*, Rymar O., Ivanoshchuk D., Shakhtshneider E.

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

\* [aknikolaeva@bk.ru](mailto:aknikolaeva@bk.ru)

**Key words:** MODY diabetes, genetic research, mutations, young patients

*Motivation and Aim:* Most young patients with hyperglycemia are diagnosed with type 1 diabetes mellitus and type 2 diabetes mellitus but up to 10 % of all cases of the disease occur in MODY (Maturity-Onset Diabetes of the Young) diabetes [1]. Diagnosis of the correct type of diabetes mellitus (DM) leads to the appointment of pathogenetic effective therapy, adequate management of pregnancy and the prevention of specific complications. Currently 14 types of MODY are known which differ in clinical manifestations and therapy, so it is very important to understand which phenotypic manifestations correspond to a certain type of monogenic form of diabetes. The aim of the study was to explore the possibility of molecular genetic testing in determining the type of diabetes mellitus in young people.

*Materials and methods:* The Research Institute of Internal and Preventive Medicine – branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences examined 421 patients with diagnosed diabetes mellitus under the age of 45 years. Of these patients 387 met the inclusion criteria for targeted sequencing of genes associated with the development of MODY 1-14 diabetes: absence of antibodies to GAD, b-cells, insulin, normal or slightly reduced C-peptide level, normal body mass index.

*Results:* Different types of MODY were verified in 135 (34.9 %) patients out of 387. In the majority of patients as in the Russian Federation [2], GCK-MODY (MODY2) prevails (85 patients, 63.0 %). Also, many patients were diagnosed with HNF1A-MODY (MODY3) (30 people, 22.2 %) and ABCC8-MODY (MODY12) (10 people, 7.4 %). The following types of MODY were also identified: HNF4A-MODY (MODY1), NEUROD1-MODY (MODY6), CEL-MODY (MODY8), APPL1-MODY (MODY14) – each of the types for 1 patient (0.7 %); HNF1B-MODY (MODY5) – in 4 patients (3.0 %), PAX4-MODY (MODY9) – in 2 patients (1.5 %).

The digenic type of inheritance was verified in some patients, mutations were identified in two genes associated with MODY at once: in one patient in the genes associated with HNF1A-MODY and ABCC8-MODY and in one patient – in the HNF1B genes and predisposing to the development of diabetes mellitus type 1. The course of diabetes was more aggressive in patients with this type of inheritance, there was decompensation of carbohydrate metabolism for several months and insulin therapy was prescribed. Thus, if a patient is diagnosed with a mutation in a gene associated with MODY this does not exclude the presence of other types of diabetes mellitus. If a pathogenic mutation associated with MODY is detected and clinical manifestations are uncharacteristic for this type of MODY including a rapidly progressive course of DM, it is necessary to study other genes predisposing to diabetes.

*Conclusions:* 1. The frequency of mutations in MODY-associated genes among individuals with phenotypic features of this type of diabetes is 35 %.  
2. In case of decompensated progressive course of carbohydrate metabolism in MODY it is necessary to examine other genes that predispose to the development of other types of DM.

*Acknowledgements:* Abstract was written within the framework of the grant of the President of the Russian Federation for state support of young Russian scientists – doctors of sciences MD-3017.2022.3.

### *References*

1. Ovsyannikova A.K., Galenok R.B., Rymar O.D. The role of "latent autoimmune diabetes in adults" in rare cases of diabetes mellitus in young people. *Med Advice*. 2021;(21-1):150-155. doi: 10.21518/2079-701X-2021-21-1-150-155.
2. Zubkova N.A., Gioeva O.A., Tikhonovich Y.V., Petrov V.M., Vasiliev E.V., Tyulpakov A.N., Dedov I.I. Genotype-based personalized correction of glycemic control in patients with MODY due to mutations in GCK, HNF1A AND HNF4A genes. *World J Pers Med*. 2017;1(1):40-48. doi: 10.14341/WJPM9298.

## Bioinformatics search for genetic markers of socially significant diseases in promoters of human genes

Ponomarenko M.\*, Chadaeva I., Rasskazov D., Zolotareva K., Khandaev B., Oshchepkov D., Savinkova L., Sharypova E., Kolchanov N.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\*[pon@bionet.nsc.ru](mailto:pon@bionet.nsc.ru)

**Key words:** human, gene, promoter, TATA-box, SNP, expression alteration, disease, candidate SNP-marker, bioinformatics model, *in silico* prediction, *in vitro* verification

*Motivation and Aim:* The most important for diagnosing genetic susceptibility to human diseases within predictive-preventive personalized participatory (4P) medicine is a search for their genome-wide markers, among which single nucleotide polymorphisms (SNPs) are of the greatest value.

*Methods and Algorithms:* Using a computer-experimental approach, we created our free available Web service SNP\_TATA\_Comparator ([www.mgs.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl](http://www.mgs.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl)) [1].

*Results:* It allows to estimate how SNPs effect on the TATA-binding protein (TBP) binding affinity for the 70 bp proximal promoters in front of transcription start sites (TSS) of the human genes within our three-step model of TBP-promoter interaction [2] taking into account both nucleotide context and conformational together with physicochemical features of the B-helical double-stranded DNA [3]. Due to this we examined 15243 SNPs in question that yielded 3229 candidate SNP-markers aggravating or relieving the development of human diseases such as obesity [4], autoimmune disorders [5], subfertility [6], atherosclerosis [7], rheumatoid arthritis [8], hypertension [9], Alzheimer's disease [10], mental disorders [11] as well as SNP-markers of resistance to anticancer chemotherapy were also identified [12].

*Conclusion:* An independent selective experimental control verification [13, 14] shown a significant linear correlation between the *in silico*-estimated and *in vitro*-measured values of the TBP-promoter affinity rates ( $r = 0.77$ ;  $p < 0.000001$ ).

*Acknowledgements:* We thank the Multi-Access Center "Bioinformatics" for the use of computational resources supported by Russian government project FWNR-2022-0020.

### References

1. Ponomarenko 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. *Bio Med Res Int.* 2015;2015:359835.
2. Ponomarenko et al. 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.
3. Ponomarenko et al. Conformational and physicochemical DNA features specific for transcription factor binding sites. *Bioinformatics.* 1999;15:654-668.
4. Arkova et al. Obesity related known and candidate SNP markers can significantly change affinity of TATA binding protein for human gene promoters. *BMC Genomics.* 2015;16:S5.
5. Ponomarenko et al. Candidate SNP markers of gender-biased autoimmune complications of monogenic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front Immunol.* 2016;7:130.
6. Chadaeva et al. Candidate SNP markers of reproductive potential are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics.* 2018;9:0.



7. Ponomarenko et al. Candidate SNP markers of atherogenesis significantly shifting the affinity of TATA-binding protein for human gene promoters show stabilizing natural selection as a sum of neutral drift accelerating atherogenesis and directional natural selection slowing it. *Int J Mol Sci.* 2020;21:1045.
8. Klimova et al. Disruptive selection of human immunostimulatory and immunosuppressive genes both provokes and prevents rheumatoid arthritis, respectively, as a self-domestication syndrome. *Front Genet.* 2021;12:610774
9. Oshchepkov et al. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci.* 2022;23:2835.
10. Ponomarenko et al. Candidate SNP markers of familial and sporadic Alzheimer's diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front Aging Neurosci.* 2017;9:231.
11. Ponomarenko et al. Unannotated single nucleotide polymorphisms in the TATA box of erythropoiesis genes show *in vitro* positive involvements in cognitive and mental disorders. *BMC Med Genet.* 2020;21(Suppl. 1):165.
12. Turnaev et al. Hypothetical SNP markers that significantly affect the affinity of the TATA-binding protein to VEGFA, ERBB2, IGF1R, FLT1, KDR, and MET oncogene promoters as chemotherapy targets. *Mol Biol (Mosk).* 2016;50:161-73.
13. Savinkova et al. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. *PLoS One.* 2013;8:e54626.
14. Drachkova et al. The mechanism by which TATA-box polymorphisms associated with human hereditary diseases influence interactions with the TATA-binding protein. *Hum Mutat.* 2014;35:601-608.

## The genome-wide search of SNPs in 3'-UTR microRNA target sequences associated with individual drug susceptibility

Rykova E.<sup>1,2,3\*</sup>, Ershov N.<sup>1</sup>, Merkulova T.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State Technical University, Novosibirsk, Russia

\* rykova.elena.2014@gmail.com

**Key words:** SNPs, 3'-UTR, miRNA target sites, drug response

*Motivation and Aim:* The development of approaches to the large-scale identification of genetic determinants of individual sensitivity to drugs is one of the central tasks of fundamental medicine. Many SNPs in 3'-UTR microRNA (miRNA) target sequences are associated with various diseases [2] however, their role in response to therapy remains poorly understood. The aim of this study was the identification of SNPs in 3'-UTR miRNA target sequences, underlining individual drug susceptibility, based on available data on eQTL and allele-specific expression (ASE) analysis.

*Methods and Algorithms:* The search for SNPs in 3'-UTR miRNA target sites was carried out in 5402 eQTLs that appeared upon treatment of CD14+ monocytes with IFN-g [3] and in 253 ASE events identified upon exposure of 5 human cell types (LCLs, PBMCs, HUVECs, SMCs and melanocytes) to various drugs, nutrients and pollutants [4]. Data on the functional impact of SNPs on predicted miRNA sites were taken from the PolymiRTS database [1]. ASE events that showed inconsistent changes across replicates were filtered out additionally.

*Results:* Of the 5402 eQTLs (5270 SNPs and 4173 genes) identified under IFN-g treatment, 317 (307 SNPs and 298 genes) contained predicted miRNA binding sites affected by the nucleotide substitution. Since 317 eQTLs were shown not to be significantly enriched in GO/KEGG terms relative to 5402 eQTLs, it can be assumed that they all belong to the same functional groups. The highest enrichment in the group of 317 eQTLs was shown for the terms “cellular response to stress” (41 genes, FDR = 3.71 e-8) and “immune response” (36 genes, FDR = 5.59 e-6). According to the mSigDB database, the greatest enrichment (44 genes, FDR = 1.66 e-11) was shown for the genes up-regulated in PBMCs after exposure to FluMist live virus vaccine. In addition, predicted target genes of hsa-miR-8485 (28 genes FDR = 4.3e-7) were also enriched. Out of 253 ASE events, 58 coincided with the predicted SNPs in 3'-UTR miRNA target sequences; allele asymmetry in these sites in response to drugs was shown for 14 (Acetaminophen – 2 ASE; Triclosan – 1 ASE; Tunicamycin – 1 ASE; Dexamethasone – 5 ASE; Caff eine – 5 ASE). In most cases, reduced expression of the allelic variant corresponded to the appearance of a binding site for a particular Mir(s).

*Conclusion:* Both eQTL and ASE analysis are effective tools to identify SNPs located in 3'-UTR miRNA target sequences and associated with individual drug response.

*Acknowledgements:* RSF Grant No. 22-25-00255 supports the study.

## References

1. Bhattacharya A., Ziebarth J.D., Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 2014;42(Database issue):D86-D91. doi: 10.1093/nar/gkt1028.
2. Chhichholiya Y., Suryan A.K., Suman P., Munshi A., Singh S. SNPs in miRNAs and target sequences: role in cancer and diabetes. *Front Genet.* 2021;12:793523. doi: 10.3389/fgene.2021.793523.
3. Fairfax B.P., Humburg P., Makino S., Naranbhai V., Wong D., Lau E., Jostins L., Plant K. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science.* 2014;343(6175):1246949.
4. Moyerbrailean G.A., Richards A.L., Kurtz D., Kalita C.A., Davis G.O., Harvey C.T., Alazizi A., Watza D., Sorokin Y. et al. High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* 2016;26:1627-1638.

## DNA-microarray as an alternative diagnostic tool for targeted genetic carrier screening in population of Yakut ethnic group

Savvina M.<sup>1\*</sup>, Maksimova N.<sup>1</sup>, Danilova A.L.<sup>1</sup>, Sukhomyasova A.<sup>1,2</sup>, Lebedev I.<sup>3</sup>

<sup>1</sup> M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup> Republic Hospital No. 1 – National Center of Medicine, Yakutsk, Russia

<sup>3</sup> Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

\* mira@savv.in

**Key words:** DNA-microarray, population isolates, hereditary diseases, targeted genetic carrier screening

*Motivation and Aim:* There are populations that have kept apart for many centuries, show less genetic variability and they differ in their genetic structure from representatives of other groups – genetic isolates. One of its representative is the population of Yakut ethnic group that demonstrate a high prevalence of rare autosomal recessive disorders due to the founder effect [1]. We evaluated the heterozygous carrier frequencies of five monogenic disorders enriched in the population among healthy individuals using own developed low density DNA-microarray. The results were verified by genotyping using real-time PCR and PCR-RLFP assays. Testing using DNA-microarray showed 99 % sensitivity and 100 % specificity.

*Methods and Algorithms:* Oligonucleotide probes addressing point mutations 4582\_4583insT in CUL7, c.5741G > A in NBAS, c.806C > T in DIA1 c.1090G > C in FAH., c.-23+1G > A in GJB2 causing 3-M syndrome, SOPH-syndrome, Methaemoglobinaemia type 1, Tyrosinemia type 1, Nonsyndromic hearing loss and deafness (DFNB1) type 1A respectively and wild type probes were printed unto epoxy coated glass slide by non-contact manner using A Perkin Elmer Piezozarray (Perkin Elmer) microarrayer. The assay using developed microarray chip based on reverse hybridization which include two step multiplex PCR with production of one-stranded Cy5 labeled-PCR products following its hybridization with oligonucleotide probes [2]. *Results:* We developed population specific low-density microarray and performed genotyping of 120 healthy individuals for major mutations causing 3-M syndrome, SOPH-syndrome, Methaemoglobinaemia type 1, Nonsyndromic hearing loss and deafness (DFNB1) type 1A, Tyrosinemia type 1. The results of genotyping demonstrate a high frequency of all five genetic disorders tested and were: 3-M syndrome – 1:25, SOPH-syndrome – 1:50, Methaemoglobinaemia type 1 – 1:25, DFNB1 – 1:11, Tyrosinemia type 1 – 1:100.

*Conclusion:* Obtained results of high level of carrier frequencies of above genetic diseases indicates a high need of its prevention through implementation of carrier screening programs in population. The developed population specific low density oligonucleotide microarray can be considered as an alternative highly specific low-cost tool for heterozygous carrier screening and genetic diagnostics in Republic of Sakha (Yakutia). In addition, it can easily be modified to accommodate additional mutations.

*Acknowledgements:* The research is conducted under the state target program: project FSRG-2020-0014 “Genomic of Arctic: epidemiology, hereditary and pathology”.

### References

1. Puzyrev V.P., Maximova N.P. Hereditary diseases among Yakuts. *Russ J Genet.* 2008;44(10):1141-1147. doi: 10.1134/S1022795408100037.
2. Savvina M.T. et al. Oligonucleotide microarray based method for simultaneous diagnostic of 3-M syndrome, soph syndrome, tyrosinemia type 1, methaemoglobinaemia type 1, nonsyndromic hearing loss and deafness (DFNB1). *Med Genet.* 2019;18(9):24-33.

## Mutation landscape of non-small cell lung cancer with a high risk of metastasis and recurrence

Schegoleva A. \*, Gerashchenko T., Khozyainova A., Vorobev R., Rodionov E., Pankova O., Perelmuter V., Denisov E.

*Cancer Research Institute, Tomsk National Research Medical Center, Tomsk, Russian*

\* *shegolmay@gmail.com*

**Key words:** mutation, basal cell hyperplasia, squamous metaplasia, distant metastasis, relapse

*Motivation and Aim:* Previous studies have found that various combinations of premalignant changes in the small bronchi epithelium distant from the tumor are associated with the risk of progression of non-small cell lung cancer (NSCLC): patients with basal cell hyperplasia (BCH) more often have distant metastases, whereas patients with BCH combined with squamous metaplasia (SM) predominantly show locoregional relapses [1, 2]. The molecular mechanisms of this association are still not understood. Here, we aimed to assess somatic genetic alterations in NSCLC patients with different premalignant bronchial lesions.

*Methods and Algorithms:* The study included 59 patients with NSCLC divided into three groups: 1) cases with isolated BCH and a high risk of distant metastasis; 2) cases with the presence of BCH and SM (BCH+SM) and a high risk of locoregional recurrence; 3) cases without premalignant lesions. Premalignant lesions were identified in hematoxylin and eosin-stained sections of formalin-fixed paraffin-embedded samples of small bronchi distant (3–5 cm) from the tumor. DNA samples were isolated from fresh-frozen tumor and peripheral blood samples and sequenced on a NextSeq 500 (Illumina) instrument using SureSelect XT Human All Exon v7 kit (Agilent). Data were analyzed using the GATK pipeline, and genetic variants were annotated using the ANNOVAR. Copy number alterations (CNAs) were detected by CNVkit.

*Results:* Patients with BCH and a high risk of metastasis have alterations in genes involved in the regulation of interleukin synthesis, T-helper differentiation, and defense response to Gram-negative bacterium. Patients with BCH+SM and a high risk of locoregional recurrence demonstrate mutations in genes that are related to exocytosis and cell adhesion.

*Conclusion:* The mutational landscape of NSCLC differs between patients with a high risk of distant metastasis and locoregional recurrence. The detected genetic features confirm the stratification of NSCLC patients according to the type of premalignant changes in the bronchial epithelium and may underlie the NSCLC progression.

### References

1. Pankova O.V., Denisov E.V., Ponomaryova A.A., Gerashchenko T.S., Tuzikov S.A., Perelmuter V.M. Recurrence of squamous cell lung carcinoma is associated with the co-presence of reactive lesions in tumor-adjacent bronchial epithelium. *Tumour Biol.* 2016;37(3):3599-3607.
2. Pankova O.V., Tashireva L.A., Rodionov E.O., Miller S.V., Tuzikov S.A., Pismenny D.S., Gerashchenko T.S., Zavyalova M.V., Vtorushin S.V., Denisov E.V., Perelmuter V.M. Premalignant changes in the bronchial epithelium are prognostic factors of distant metastasis in non-small cell lung cancer patients. *Front Oncol.* 2021;11:771802.

## Pluripotent stem cell lines from two patients with *COH1* gene mutations as the valuable *in vitro* model of Cohen syndrome

Shnaider T.<sup>1</sup>, Khabarova A.<sup>1</sup>, Morozova K.<sup>1</sup>, Kiseleva E.<sup>1</sup>, Vladimirova E.<sup>2</sup>, Chechetkina S.<sup>2</sup>, Chvileva A.<sup>2</sup>, Smirnova A.<sup>3</sup>, Larina D.<sup>3</sup>, Pristyazhnyuk I.<sup>1\*</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Laboratory of clinical genomics and bioinformatics, Research and Clinical Institute for Pediatrics named after Academician Yuri Veltishev, Moscow, Russia*

\* [iprist@bionet.nsc.ru](mailto:iprist@bionet.nsc.ru)

**Key words:** COH1, VPS13B, intellectual disability, autism, induced pluripotent stem cells, Cohen syndrome, Golgi apparatus, vesicular transport

Cohen's syndrome (CS) is an autosomal recessive disorder caused by mutations in the *COH1* (*VPS13B*) gene. CS is characterized by growth and mental retardation, microcephaly, and the development of autism spectrum disorders.

The *COH1* gene has 62 exons and covers a region of approximately 864 kb on the human chromosome 8 [2]. For the moment, several hundred patients with CS have been described in the literature [1, 4], and according to the Cohen Syndrome Association (<http://www.cohensyndrome.org/>), there are about a thousand of CS patients all around the world.

The COH1 protein is one of the key participants of the intracellular membrane transport system, the central part of which is the Golgi apparatus (GA). Recent studies have shown that mutations affecting GA proteins also result in the postnatal microcephaly associated diseases. GA is extremely important for the differentiation and functioning of postmitotic neurons. It plays a major role in determining and creating the cellular polarity necessary for the formation of highly specialized neurites – axons and dendrites.

The product of the *COH1* gene is a peripheral membrane protein localized in GA and contributes to the structural maintenance and functioning of this complex. The supposed function of the COH1 protein is the regulation of vesicular transport and intracellular protein sorting [2, 5, 6]. Meanwhile, the molecular mechanisms of the *COH1* effects are still poorly understood.

Despite the similarity of the early stages of neurogenesis in all mammals, some of them cannot be reproduced in laboratory animals due to species-specific diversity. However, now the technologies of patient-specific induced pluripotent stem cells (iPSCs) have been developed and established the methods of their directed differentiation into neural cells. These technologies allow to overcome such limitations and carry investigations directly on the human cells. These approaches are already widely used for studying and modeling of the human neurodevelopmental and neurodegenerative disorders. However, in the literature there is only one report of CS iPSCs [3].

At the moment, we have obtained iPSC lines from the biopsy material of two patients with compound heterozygous mutations in the *COH1* gene:

*Patient 1*, (biopsy material – blood mononuclear cells) is a carrier of a compound heterozygous mutation (*COH1*-/-; chr8:g.100108653dup/chr8:g.100494031G > T).

*Patient 2*, (biopsy material – skin fibroblasts) is a carrier of a compound heterozygous mutation (*COH1*<sup>-/-</sup>; chr8:g.100514033T > C/ chr8:g.100844663\_100844664del). For the skin fibroblasts of this patient, the following was revealed: swelling and vacuolization of GA cisternas, as well as GA fragmentation, which is typical for the CS patients. In addition, in the CS fibroblast cytoplasm, there are a large number of electron-dense granules, vesicles, autophago- and lysosomes, as well as an arrangement of dense bundles of microfilaments associated with the nucleus.

The iPSC lines from both patients have morphology of pluripotent cell, normal karyotype (46, XX), express pluripotency markers (OCT4, NANOG), and surface markers SSEA-1 and TRA-1-60. During differentiation in embryoid bodies, they contribute to three germ layers (ectoderm, mesoderm, and endoderm).

iPSC lines with a compound-heterozygous mutation in the *COH1* gene will serve as the basis for studying the pathology of cortical neurons formation in CS patients. Their use will allow us to study the later stages of neurogenesis, in particular neuriteogenesis, axons and dendrites formation, as well as the features of the structural organization of the GA and vesicular transport in the cell. By studying the effect of mutations in the *COH1* gene that cause CS, we will be able to establish the role of this gene in the process of normal human neurogenesis.

*Acknowledgements:* This study was supported by the Russian Foundation for Basic Research (grant No. 19-29-04067). We also thank the Multiple-access Center for Microscopy of Biological Subjects (Institute of Cytology and Genetics, Novosibirsk, Russia), state project No. FWNR-2022-0019 for granting access to microscopic equipment.

We are grateful to Pomerantseva E., Kotov I. and Musatova E., from the Center of Genetics and Reproductive Medicine GENETICO, JSC, Moscow, Russia and Lagarkova M. and Anufrieva K. from the Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia, for the invaluable help in obtaining biopsy material from patients and fibroblast cell.

## References

1. Alipour N. et al. Mutations in the VPS13B Gene in Iranian Patients with Different Phenotypes of Cohen Syndrome. *J Mol Neurosci.* 2020;70(1):21-25.
2. Kolehmainen J. et al. Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. *Am J Hum Genet.* 2003;72(6):1359-1369.
3. Lee Y.-K. et al. Cohen syndrome patient iPSC-derived neurospheres and forebrain-like glutamatergic neurons reveal reduced proliferation of neural progenitor cells and altered expression of synapse genes. *J Clin Med.* 2020;9(6):1886.
4. Nasser F. et al., Ophthalmic features of retinitis pigmentosa in Cohen syndrome caused by pathogenic variants in the VPS13B gene. *Acta Ophthalmol.* 2020;98(3):e316-e321.
5. Seifert W. et al. Mutational spectrum of COH1 and clinical heterogeneity in Cohen syndrome. *J Med Genet.* 2006;43:e22.
6. Seifert W. et al. Cohen syndrome-associated protein, COH1, is a novel, giant Golgi matrix protein required for Golgi integrity. *J Biol Chem.* 2011;286(43):37665-37675.

# Investigation of the mechanisms of neurodevelopmental disorders caused by mutations in the *CNTN6* gene on the cerebral organoids model

Shnaider T.A.<sup>1\*</sup>, Pristyazhnyuk I.E.<sup>1</sup>, Yunusova A.M.<sup>1</sup>, Chechetkina S.A.<sup>2</sup>, Khabarova A.A.<sup>1</sup>, Belokopytova P.S.<sup>1,2</sup>, Smirnov A.V.<sup>1</sup>, Chivileva A.S.<sup>2</sup>, Serov O.L.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* shnayder.t@yandex.ru

**Key words:** cerebral organoids, *CNTN6*, iPSC, neural differentiation, neurogenesis, mental retardation, neurodevelopmental disorders, NOTCH pathway

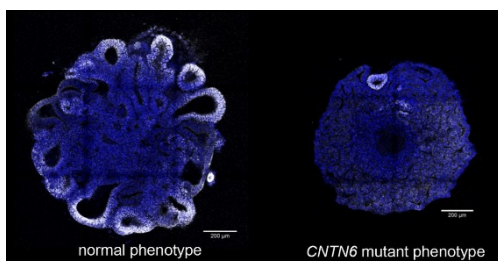
Hereditary intellectual disability occurs in 2-3 % of newborns. Up to 25 % of these neurodevelopmental disorders are associated with gene copy number variation. Several dozen patients have been described with CNVs that affect the *CNTN6* gene [4, 5]. This gene encodes the protein Contactin6, which is involved in the control of dendritogenesis and the synaptogenesis during the development of the mice nervous system [9]. Our attention in the last years is attracted by the patients – heterozygous carriers of the *CNTN6* gene deletion, who suffered from the severe neuropsychiatric symptoms and craniofacial dysmorphism. To study the effects of the *CNTN6* mutation, we obtained induced pluripotent stem cells (iPSCs) from skin fibroblasts of healthy donors and patients with *CNTN6* deletion. Additionally we obtained knockout lines based on iPSCs by applying the CRISPR/Cas9 system.

To simulate the early events of human neurogenesis, such as development of neuroepithelial structures and later ones (migration of neurons and the cortical plate formation), we used a unique 3D cellular model – human cerebral organoids (CO) [7]. This model system is currently used in the study of a wide range of diseases associated with brain development [3, 8].

For the first time, we have described a disturbance of neurogenesis in the *CNTN6* deletion patients that leads to malformations during cortical development. The main

features of the *CNTN6* mutant phenotype are the disturbance of the neuroepithelial structures formation and alteration of the radial glia proliferation (Fig.). In particular, knockout of the *CNTN6* gene leads to a disruption in the radial glial cells division and their transition from symmetrical proliferative to asymmetric neurogenic. The overexpression of *CNTN6* leads to the restoration of the mutant phenotype. In

particular, we have shown for the first time that *CNTN6* is expressed not only in the postmitotic deep layer neurons, but also in the radial glial cells. It turned out that the level of *CNTN6* expression in neural stem cells (NSC) is higher than in terminally





differentiated neurons, which may indicate an important role of the *CNTN6* gene in the maintenance of NSC, in particular, in the regulation of their proliferation.

Based on the transcriptome analysis of COs obtained from normal and mutant iPSCs, we identified four signaling pathways that may involve the Contactin 6 protein: NOTCH, WNT, TGF- $\beta$ , and BMP. In *CNTN6* knockout COs, the WNT, TGF- $\beta$ , and BMP signaling pathways were activated, while the NOTCH signaling pathway was downregulated.

In addition, we discovered, for the first time, that the process of PAX6 cytoplasm-to-nucleus translocation is disrupted during iPSC to NSC differentiation of the cells with *CNTN6* homozygous mutation. This violation may be the key molecular event leading to the mutant phenotype appearance. According to the literature data, it is known that the TGF- $\beta$  signaling pathway is involved in the process of the PAX6 translocation from the cytoplasm to the nucleus, in particular, through the direct binding of the SMAD3 protein to PAX6 [10]. On the other hand, the example of the TGF- $\beta$  and NOTCH signaling pathways cross-regulation through the interaction of the SMAD3 protein and the intracellular domain of the NOTCH1 receptor was described in the literature [1]. Recent publications point to the extremely important role of the NOTCH signaling pathway in the evolution of the human brain. In particular, new non-canonical ligands have been described that additionally activate this signaling pathway, thus affecting the human radial glial cells proliferation [3, 6]. Interestingly, the Contactin 6 is also a non-canonical ligand of the NOTCH signaling pathway, but this interaction has so far been described only for mouse oligodendrocytes [2].

Thus, our results in the context of the literature data mentioned above indicate a completely new previously undescribed function of the *CNTN6* in early human neurogenesis.

*Acknowledgements:* This study was supported by the Russian Foundation for Basic Research (grant No. 19-29-04067). We also thank the Multiple-access Center for Microscopy of Biological Subjects (Institute of Cytology and Genetics, Novosibirsk, Russia), state project No. FWRN-2022-0019 for granting access to microscopic equipment.

## References

1. Blokzijl A.C. et al. Cross-talk between the Notch and TGF-beta signaling pathways mediated by interaction of the Notch intracellular domain with Smad3. *J Cell Biol.* 2003;163(4):723-728.
2. Cui X.Y. et al. NB-3/Notch1 pathway via Deltex1 promotes neural progenitor cell differentiation into oligodendrocytes. *J Biol Chem.* 2004;279(24):25858-25865.
3. Fiddes I.T. et al. Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. *Cell.* 2018;173(6):1356-1369.e22.
4. Hu J. et al. CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *J Neurodevelop Disord.* 2015;7:26.
5. 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(1):97.
6. Lodewijk G.A. et al. Evolution of human brain size-associated NOTCH2NL genes proceeds toward reduced protein levels. *Mol Biol Evol.* 2020;37(9):2531-2548.
7. Lancaster M.A. et al. Cerebral organoids model human brain development and microcephaly. *Nature.* 2013;501(7467):373-379.
8. Mellios N. et al. MeCP2-regulated miRNAs control early human neurogenesis through differential effects on ERK and AKT signaling. *Mol Psychiatry.* 2018;23(4):1051-1065.
9. Sakurai K. et al. Synaptic formation in subsets of glutamatergic terminals in the mouse hippocampal formation is affected by a deficiency in the neural cell recognition molecule NB-3. *Neurosci Lett.* 2010;473(2):102-106.
10. Shubham K., Mishra R. Pax6 interacts with SPARC and TGF- $\beta$  in murine eyes. *Mol Vis.* 2012;18:951-956.

## Complex alleles of splicing variants in Mendelian diseases

Skoblov M.<sup>1\*</sup>, Gatupov M.<sup>2</sup>, Reveguk I.<sup>3</sup>, Marakhonov A.<sup>1</sup>, Bychkov I.<sup>1</sup>, Davydenko K.<sup>1</sup>, Filatova A.<sup>1</sup>

<sup>1</sup>Research Centre for Medical Genetics, Moscow, Russia

<sup>2</sup>Moscow Institute of Physics and Technology, Dolgoprudnyy, Moscow Region, Russia

<sup>3</sup>Laboratoire de Biologie Structurale de la Cellule, École Polytechnique, Paris, France

\*mskoblov@gmail.com

**Key words:** hereditary diseases, complex alleles, splicing variants

**Motivation and Aim:** It is well known that single nucleotide polymorphisms (SNPs) in the human genome are mostly neutral. However, if a mutation appears near to a frequent SNP their co-operative impact on splicing regulatory sequences can be very significant. At the same time, the effect of mutation could be different depending on the allelic states of SNP. For instance, the appearance of a mutation can lead to the formation of a splicing site in one allelic state, but not in another. Since SNPs occur in the human genome every 300 nucleotides on average, such events of complex genetic interaction between rare mutation and frequent SNP should be relatively frequent. However, today there are no algorithms to detect them.

**Methods and Algorithms:** We created an “alternative” version of the human genome (hg19 assembly) where every frequent SNP contains its alternative allele. Using the modified SpliceAI algorithm applied on both reference and “alternative” genome versions, we predicted how mutations can change the splicing pattern in one version of the genome but not in another. The minigene system used for experimental study of splicing in a complex alleles. All *in vitro* analysis performed in HEK293 cells.

**Results:** We performed genome-wide search for splice variants with different genetic backgrounds. We found 40'844 mutations which could form complex splicing alleles. About 20 % of these mutations locate in OMIM genes. A detailed analysis of these mutations has shown various mechanisms of disruption of normal splicing patterns. Using the minigene system, we carried out experimental validation of some cases that may be of great interest to the field of medical genetics. For example we demonstrated that some mutations in AR genes form complex pathogenic alleles and really exist in population. Another mutations in AD genes form protective complex allele.

**Conclusion:** Pathogenic complex alleles are very little studied. This work is one of the first attempts to search and analyze complex alleles that affect splicing.

### References

1. Filatova A. et al. Functional reassessment of PAX6 single nucleotide variants by *in vitro* splicing assay. *Eur J Hum Genet.* 2019;27(3):488-493.
2. Jaganathan K. et al. Predicting splicing from primary sequence with deep learning. *Cell.* 2019;176(3):535-548.e24.

## Ficolin-3 and MASP-2 gene variants in Russian Arctic populations

Smolnikova M.

Krasnoyarsk Scientific Centre, SB RAS, Scientific Research Institute of Medical Problems of the North, Krasnoyarsk, Russia

\* smarinv@yandex.ru

**Key words:** lectin pathway, gene polymorphism, Arctic population, MBL, FCN, MASP

*Motivation and Aim:* Ficolin-3 and MASP-2 are among the key participants in the lectin pathway of complement activation. Ficolins are pattern recognition receptors capable of binding to MASP and activating complement via the lectin pathway. Ficolin-3 is the most powerful activator of the lectin pathway, with its serum concentrations being significantly higher than MBL level. The *FCN3* rs28357092 mutation is associated with low ficolin-3 levels in plasma. MASP-2 is a protease that turns out to be complexed with MBL and ficolins, leading to complement activation. Mutation in *MASP2* rs72550870 is associated with impaired protein binding to lectins, and therefore complement activation does not occur. The purpose of the study was to reveal ethnic differences in the distribution of allelic gene variants for the lectin pathway components of complement activation between the indigenous populations of the Arctic territory of Siberia and Caucasoids.

*Methods and Algorithms:* Newborn children of indigenous Arctic ethnic groups (Nenets, Dolgan-Nganasans, mixed population) ( $n = 678$ ) and Russians ( $n = 302$ ) were included in the study. *FCN3* rs28357092 and *MASP2* rs72550870 genotyping was carried out by RT-PCR. Allele and genotype frequencies were compared using an online calculator, the  $\chi^2$  test.

*Results:* As the analysis result of the *FCN3* rs28357092 and *MASP2* rs72550870 genotype frequency, the prevalence of GG and AA homozygotes respectively, in all the studied populations according to the world data has been shown. The heterozygous genotype G/del rs28357092 was found to take place in only one Russian child. The heterozygous AG rs72550870 genotype occurs occasionally in the Arctic populations compared with Russians, i.e. the frequency of the AG genotype in Russians was 6.6 %, in Nenets – 0.3 %, in Dolgan-Nganasans – 1.4 %, and in the mixed population – 1.8 %. None of the groups were found to be homozygous for the minor allele G (Table 1). *Conclusion:* The genetic analysis results showed lower prevalence of genetic markers of ficolin-3 and MASP-2 deficiency in the indigenous populations of the Arctic territories compared to Russians.

*Acknowledgements:* Author expresses gratitude to the staff of the Krasnoyarsk Regional Medical Genetic Center for their help in collecting material and to the Professor S.Yu. Tereshchenko for discussing the results.

**Table 1.** The *FCN3* and *MASP2* genotype frequency in newborns of different ethnic populations, *n* (%)

Genotype /allele	Nenets, <i>n</i> = 323 (1)	Dolgan-Ngasans, <i>n</i> = 138 (2)	Mixed population, <i>n</i> = 217 (3)	Caucasoids, <i>n</i> = 302 (4)	<i>p</i>
<i>FCN3</i> rs28357092					
GG	292 (100.0)	128 (99.2)	199 (98.0)	291 (96.4)	1/3 = 0.02 1/4 < 0.001
G/del	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	–
del/del	0 (0.0)	1 (0.8)	4 (2.0)	10 (3.3)	1/3 = 0.02 1/4 = 0.002
del*	0 (0.0)	2 (0.8)	8 (2.0)	21 (3.5)	1/3 < 0.001 1/4 < 0.001 2/3 = 0.02 2/4 = 0.003
<i>MASP2</i> rs72550870					
AA	322 (99.7)	136 (98.6)	213 (98.2)	226 (93.4)	1/4 < 0.001 2/4 = 0.02 3/4 = 0.01
AG	1 (0.3)	2 (1.4)	4 (1.8)	16 (6.6)	1/4 < 0.001 2/4 = 0.02 3/4 = 0.01
GG	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
G*	1 (0.2)	2 (0.7)	4 (0.9)	16 (3.3)	1/4 < 0.001 2/4 = 0.03 3/4 = 0.01

## Clinical and molecular characteristics of ichthyosis in Yakutia

Sofronova V.<sup>1,2\*</sup>, Sukhomyasova A.<sup>1,3</sup>, Maksimova N.<sup>1</sup>

<sup>1</sup>Laboratory of Molecular Medicine and Human Genetics, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup>Department of Molecular and Genetic Medicine, Kawasaki Medical School, Kurashiki-shi, Japan

<sup>3</sup>Medical Genetics Center, Republican Hospital No.1 – National Medical Center, Yakutsk, Russia

\* viksofmax92@gmail.com

**Key words:** ichthyosis, genodermatosis, skin disorder, *ALOXE3*, *KRT10*

**Introduction:** Hereditary ichthyosis is common name for many dry and scaly skin disorders ranging in frequency from common to very rare [1]. Clinical genealogical analysis of hereditary forms of ichthyosis (from 1989 to 2018) are conducted in the Republic of Sakha (Yakutia), Russia.

**Materials and Methods:** Data was collected from the medical cards of patients of the “Republican Genetic Register of Hereditary and Congenital Pathology of the Republic of Sakha (Yakutia)” of the Medical Genetics Center of the Republican hospital No1 “National center of medicine”. Blood samples were drawn from 10 patients for clinical exome sequencing.

**Results:** The number of patients with ichthyosis is 126 patients (104 families) according to the genetic register of Yakutia according to date of 2018. The frequency of ichthyosis was about 1/7675, prevalence is 13,0 to 100 000. Gender: 94 men (74.6 %), 32 women (25.4 %). Ethnicity: Yakut – 109 (86.5 %), Russian – 17 (13.5 %). In four families affected patients of both genders and in ten cases only males. All patients had dryness, peeling of the skin. In some cases, more severe symptoms were noted in the form of pityriasis peeling, follicular keratosis. The manifestation varied from birth to adolescence. Results of clinical exome sequencing showed variants in *ALOXE3* c.1604A > G (4/10), *KRT10* c.458T > C (1/10), *STS* c.1334A > G (1/10), suspected deletion in X chromosome (2/10). Two patients were without found variants. Patients with the same c.1604A > G variant in *ALOXE3* demonstrated mild phenotype: mild/moderate erythema with fine white scale, palmoplantar hyperkeratosis as described in literature [2–6]. Patient with c.458T > C variant in *KRT10* demonstrated phenotype of collodion baby.

**Conclusions:** It is difficult to establish the type of inheritance and further medical and genetic counseling due to genetic heterogeneity and the similarity of the clinical features of ichthyosis. Thus, there is a strong need in a highly specific diagnostic method and in-depth research study for precise diagnosis, treatment and prevention of these diseases.

**Acknowledgements:** The study is supported by Grant FSRG-2020-0014 “Genomics of Arctics: epidemiology, hereditary and pathology”.

### References

1. Oji V., Tadani G., Akiyama M., Blanchet Bardon C., Bodemer C., Bourrat E. et al. Revised nomenclature and classification of inherited ichthyoses: results of the First Ichthyosis Consensus Conference in Sorèze 2009. *J Am Acad Dermatol.* 2010;63:607-641.
2. Yu Z., Schneider C., Boeglin W.E., Marnett L.J., Brash A.R. The lipoygenase gene *ALOXE3* implicated in skin differentiation encodes a hydroperoxide isomerase. *Proc Natl Acad Sci USA.* 2003;100:9162-9167.

3. Rodríguez-Pazos L., Ginarte M., Fachal L., Toribio J., Carracedo A., Vega A. Analysis of TGM1, ALOX12B, ALOXE3, NIPAL4 and CYP4F22 in autosomal recessive congenital ichthyosis from Galicia (NW Spain): evidence of founder effects. *Br J Dermatol.* 2011;165:906-911.
4. Eckl K.M., de Juanes S., Kurtenbach J., Nätebus M., Lugassy J., Oji V. et al. Molecular analysis of 250 patients with auto- somal recessive congenital ichthyosis: evidence for mutation hotspots in ALOXE3 and allelic heterogeneity in ALOX12B. *J Invest Dermatol.* 2009;129:1421-1428.
5. Jobard F., Lefèvre C., Karaduman A., Blanchet-Bardon C., Emre S., Weissenbach J. et al. Lipoxigenase-3 (ALOXE3) and 12(R)- lipoxigenase (ALOX12B) are mutated in non-bullous congenital ichthyosiform erythroderma (NCIE) linked to chromosome 17p13.1. *Hum Mol Genet.* 2002;11:107-113.
6. Eckl K.M., Krieg P., Küster W., Traupe H., André F., Wittstruck N. et al. Mutation spectrum and functional analysis of epidermis- type lipoxigenases in patients with autosomal recessive congenital ichthyosis. *Hum Mutat.* 2005;26:351-361.

## **Human genomics and evolutionary medicine**

Stepanov V.A.

*Tomsk National Research Medical Center, RAS, Tomsk, Russia*

\* *[vadim.stepanov@medgenetics.ru](mailto:vadim.stepanov@medgenetics.ru)*

The whole genome research on genetic diversity in human populations is accumulating significant impact for medical genetics and personalized medicine. The patterns of gene pool organization, current concepts of evolutionary medicine and search of genetic factors in human diseases within the framework of evolutionary genetic approaches are discussed.

## Discovery of somatic recombination events in APP and analysis of its occurrence

Sukhanova X.V.<sup>1\*</sup>, Danilov L.G.<sup>1,2</sup>, Komissarov A.S.<sup>1</sup>

<sup>1</sup> Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia

<sup>2</sup> Department of Genetics and Biotechnology, St. Petersburg State University, St Petersburg, Russia

\* [sukhanovaxeniav@gmail.com](mailto:sukhanovaxeniav@gmail.com)

**Key words:** amyloids, neurodegeneration, somatic recombination, Alzheimer

*Motivation and Aim:* Neurodegeneration is one of the most severe diseases. And it was distinguished that it is caused by protein misfolding into a specific amyloid structure and following aggregation and loss of function. Amyloid structure represents a cross- $\alpha$ -structure, possessing uniquely stable and resistant features. Several proteins are known to cause neurodegeneration, from which amyloid  $\alpha$  ( $A\alpha$ ) is the most crucial. Despite the urgency of disease, any effective therapy and even diagnostic system have not been created yet. Recent diagnostics system is based on already identified genetic markers – mutations in specific proteins-coding genes: amyloid precursor protein (APP), presenelin 1, presenelin 2. Which, indeed, lie in the protein misfolding processes and following aggregation and amyloidization. Though do not possess high specificity.

Recently, a new mechanism of APP amyloidization and, thus neurodegeneration development, was proposed [1, 2]. It was shown that neurodegeneration could be initiated by an event called somatic recombination. Commonly, this event is described as being responsible for antibody variation production and immune system reactivity. And is contributed by abnormal exon-exon recombination of specific V, D and J segments, leading to the formation of non-canonical splice variants. However, the exact mechanism and recombination positions have not been designated. In this work we created a new algorithm for somatic recombination searching and provided hypotheses of possible mechanisms.

*Methods and Algorithms:* We utilized open-access genomic sequencing data of human brain (SRR7905480 ID) to search in it all possible paired exon-exon combinations of APP gene. Before the search the filtration of raw data with 23-mers generated from APP sequences was implemented. After which the alignment of combinations on filtered genome by bwa algorithm was provided. From gained overlappings only with the highest coverage were taken into the following analysis. With obtained coordinates the insertion regions were determined. Further each insertion was analyzed on correlation with structural genomic elements, which was taken from annotation data. And the intersection of coordinates was applied as the method. For each insertion the area up and down ~100 bp was also searched for. Gained distributions then were compared with Kolmogorov-Smirnov test to define significant correlations, including the domain information of A $\beta$  protein from PDB database.

*Results:* We established that, indeed, amyloid precursor protein is affected by somatic gene recombination events in the human genome. From all possible pair-to-pair exons combinations only 5 % have a high raw reads coverage. From which combinations of exon 7 with exons 6, 9 and 10 appeared to be the most frequent. We also identified that recombinations were spread through the whole genome of chromosome 21 with a



tendency to the start of the APP gene. We additionally estimated the correlation between insertion positions distribution and structural genome elements which revealed the association between recombination insertions and such structural elements as: VDJ-segments, recombination patterns and mobile genetic elements.

*Conclusion:* Based on these findings, we proposed that indeed somatic gene recombination exists in the case of APP gene. And the whole genome of chromosome 21 is affected by APP exons recombinations. The most recombinant exon 7 could possess a sufficient role in recombination events and even disease development. And these events could be caused by sided VDJ-segment recombinations and mobile genetic elements transposition.

*Acknowledgements:* The study is supported by the laboratory of Applied Genomics SCAMT Institute, ITMO University and its collective.

### *References*

1. Gwendolyn K.E., Chun J. Mosaic somatic gene recombination as a potentially unifying hypothesis for Alzheimer's disease. *Front Genet.* 2020;11:390. doi: 10.3389/fgene.2020.00390.
2. Lee M.-H., Siddoway B., Kaeser G.E., Segota I., Rivera R., Romanow W.J., Liu C.S. et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature.* 2018;563(7733):639-645. doi: 10.1038/s41586-018-0718-6.

## Reproductive genetic screening for heterozygous carriage of hereditary diseases in the Republic of Sakha (Yakutia)

Sukhomyasova A.L.<sup>1,2</sup>, Danilova A.L.<sup>1</sup>, Grigoreva T.P.<sup>1,2</sup>, Fedorov A.I.<sup>1</sup>, Maksimova N.R.<sup>1</sup>

<sup>1</sup> M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup> Republican Hospital No. 1 – National Center of Medicine, Yakutsk, Russia

\* aitalinas@yandex.ru

**Key words:** reproductive genetic screening, autosomal recessive disease, heterozygous carriage, Republic of Sakha (Yakutia)

*Motivation and Aim:* On the territory of the Republic of Sakha (Yakutia), there is an accumulation of some hereditary diseases, their high frequency [1]. In this regard, genetic screening for heterozygous carriage of mutations in autosomal recessive diseases is of particular relevance in the RS(Y) in order to prevent them. The introduction of expanded screening requires a preliminary pilot study in populations, which makes it possible to analyze the merits and consequences of the program, for example, the clinical and cost effectiveness of screening, the psychological and social impact of the program [2]. The purpose of this study was to conduct a pilot project "Reproductive genetic screening for heterozygous carriage of hereditary diseases" to evaluate the effectiveness of genetic screening. The study included 7 autosomal recessive diseases that occur with a high frequency in Yakuts in the Republic of Sakha (Yakutia) – three M-syndrome 1, non-syndromic deafness type 1a, methemoglobinemia type 1, tyrosinemia type 1, SOPH syndrome, neuronal ceroid lipofuscinosis type 6 and mucopolysaccharidosis plus syndrome. For all these diseases, major mutations characteristic of the Yakut population have been established [1, 3].

*Methods and Algorithms:* During 2021, 915 women 11–13/6 weeks pregnant took part in the pilot project as part of early prenatal screening from 10 districts of Yakutia – Amginsky, Anabarsky, Verkhnevilyuisky, Megino-Kangalassky, Nizhnekolymsky, Nyurbinsky, Tattinsky, Tomponsky, Churapchinsky and Ust-Aldan. In terms of numbers, women of the Yakut nationality predominated – 847 (92.6 %), 52 (5.7 %) – representatives of the indigenous peoples of the North (Evenki, Evenki, Dolganki), 10 (1.1 %) – women of other nationalities. 6 participants (0.6 %) did not indicate their nationality. Genotyping was carried out using real-time PCR.

Before the start of the pilot project, a questionnaire survey was conducted to assess the attitude of the population towards genetic screening, in which 271 respondents over the age of 18 from different regions of Yakutia took part. The questionnaire included 21 questions to study the social status, family burden, awareness of the heterozygous carrier of hereditary diseases, and attitudes towards genetic screening of the respondents. Statistical analysis of the data was performed using Pearson's Chi-square test and hierarchical cluster analysis.

*Results:* According to the results of the pilot study, 193 heterozygous carriers were identified, the frequency of heterozygous carriage of 7 hereditary diseases among pregnant women from 10 districts was 21.1 %. 180 (93.3 %) women were of Yakut nationality, 9 (4.6 %) were representatives of the indigenous peoples of the North. 25

(13.0 %) women were carriers of several hereditary diseases: 3 – heterozygous carriers of 3 diseases, 22 – 2 diseases. Among the Yakuts, carriers for all 7 diseases were revealed, among the representatives of the indigenous minority of the North, carriers of the tri-M syndrome, type 1 methemoglobinemia and mucopolysaccharidosis-plus syndrome were identified.

The largest number of heterozygous carriers was found in areas with a predominance of the indigenous population – in the Amginsky, Nyurbinsky and Churapchinsky districts. Among them, the Nyurbinsky district stands out, where the maximum total number of heterozygous carriers of 7 hereditary diseases is recorded – more than 27 % of pregnant women.

The frequency of heterozygous carriage by disease was distributed as follows: non-syndromic deafness type 1a – 11 %, three M syndrome – 3 %, SOPH syndrome – 3 %, type 1 methemoglobinemia – 3 %, mucopolysaccharidosis-plus syndrome – 2 %, neuronal ceroid lipofuscinosis – 1 %, tyrosinemia type 1 – 1 %.

According to the results of a questionnaire survey, 89.3 % of respondents have a positive attitude towards genetic screening and agree to undergo DNA testing. At the same time, 93 % of rural residents, 100.0 % of respondents with a high level of awareness and planning to have children support screening. On the other hand, 22.5 % of respondents with children doubt their participation in DNA testing ( $p = 0.025$ ). According to 24 % of respondents, they have a family history of hereditary diseases.

*Conclusion:* Thus, the results on the assessment of the frequency of heterozygous carriage of seven frequent hereditary diseases among the population of Yakutia, especially in ethnically homogeneous regions, are of great practical and theoretical significance. The results obtained confirm the need for further genetic screening with an increase in the number of participants, expanding the geography of the study. The results of the questionnaire survey showed that the strengthening of explanatory work among the population about the heterozygous carrier of hereditary diseases in the media, the Internet and in antenatal clinics can help increase the effectiveness of genetic screening for prevention.

*Acknowledgements:* The work is done with the support of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. FSRG-2020-0014 "Genomics of the Arctic: Epidemiology, heredity and pathology".

### *References*

1. Genetic studies of the population of Yakutia. Puzyrev V.P., Tomskiy M.I. (Eds). Yakutsk, 2014.
2. Revazyan K.Z. Sociological aspects of genetic screening for the carriage of variants that cause the development of autosomal recessive diseases. *Med Genet.* 2020;19(10):86-88. doi: 10.25557/2073-7998.2020.10.86-88.
3. Maksimova N.R. Short Stature Diseases: 3-M syndrome and SOPH syndrome. M.: Nauka, 2021.

## Polymorphism *rs1800795* of the *IL6* gene in samples of Siberian indigenous ethnic groups

Tabikhanova L.E.<sup>1\*</sup>, Osipova L.P.<sup>1</sup>, Churkina T.V.<sup>1</sup>, Voronina E.N.<sup>2</sup>, Filipenko M.L.<sup>2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

\* *tabikhan@bionet.nsc.ru*

**Key words:** Buryats, Teleuts, Yakuts, Dolgans, Russians from East Siberia, genetic polymorphism, real-time PCR, *IL6* (*C-174G*, *rs1800795*)

**Motivation and Aim:** Investigation of the frequencies of functionally significant gene variants in the context of medical biology and gene geography is a relevant issue in studying of the population genetic structure of indigenous Siberian peoples. Genetic polymorphism of the interleukin *IL6* gene is of great importance for understanding the pathogenesis of cardiovascular, bronchopulmonary, oncological, autoimmune, non-infectious and infectious diseases.

**Methods and Algorithms:** The study was performed on the monoethnic samples of Buryats, Teleuts, Dolgans and Yakuts. Real-time PCR was used to determine the alleles of the single nucleotide polymorphism *C-174G*, *rs1800795*, in gene *IL6*.

**Results:** The highest frequency of the *-174G* allele was recorded in Yakuts (97.3 %), slightly less in Buryats, followed by Dolgans and Teleuts (87.1 %). The differences between Teleuts and Yakut populations are statistically significant. The results obtained were compared to frequencies identified for Russians from Eastern Siberia (56.5 %) and the values in other Caucasian groups (48–65 %) described in literature [1]. The studied samples of the indigenous populations showed a significantly higher occurrence of the *-174G* allele compared to Caucasians. At the same time, they were significantly lower than in several East Asian populations described in literature (Chinese, Japanese, Koreans and Vietnamese). Those Asian populations have this value close to 100 %.

**Conclusion:** Thus, an increase in the frequency of the *-174G* variant of the *IL6* gene in human populations is shown when moving along the Eurasian map from west to east. This in-between position of indigenous Siberian populations, as exemplified by Buryats and Teleuts, had been demonstrated earlier in the polymorphism frequencies of some other e genes [2]. Carriage of the *-174G* variant was shown to be associated with a high level of *IL6* expression and, as a consequence, with an enhanced inflammatory response to the pathogen encounter. The increased concentration of this allele in Siberian populations compared to Caucasian populations may indicate a selection pressure with increasing pathogen load due to environmental features.

**Acknowledgements:** Research was supported by the budget project FWNR-2022-0021 of the Institute of Cytology and Genetics, SB RAS.

### References

1. 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.
2. Tabikhanova L.E., Osipova L.P., Voronina E.N., Bragin A.O., Filipenko M.L. Polymorphism of lipid exchange genes in some populations of South and East Siberia. *Vavilov J Genet Breed*. 2019;23(8):1011-1019. doi: 10.18699/VJ19.578.

## Genomics and transcriptomics of preeclampsia

Trifonova E.<sup>1,2\*</sup>, Svarovskaya M.<sup>1,2</sup>, Gabidullina T.<sup>2</sup>, Serebrova V.<sup>1</sup>, Babovskaya A.<sup>1</sup>, Gavrilenko M.<sup>1</sup>, Stepanov V.<sup>1</sup>

<sup>1</sup> Tomsk National Research Medical Center, SB RAS, Tomsk, Russia

<sup>2</sup> Siberian State Medical University, Tomsk, Russia

\* [ekaterina.trifonova@medgenetics.ru](mailto:ekaterina.trifonova@medgenetics.ru)

**Key words:** preeclampsia, genome-wide analysis of expression profiles, placental tissue, tagSNPs

*Motivation and Aim:* Preeclampsia (PE) is a common pregnancy-specific disorder with unknown etiology. It is the leading cause of maternal and perinatal morbidity and mortality. It is generally accepted that the main pathogenetic mechanisms of PE are associated with impaired placentation processes and reduced remodeling of the myometrial spiral arteries. The study of molecular processes occurring in the placental tissue is considered as the most promising line of research for revealing the pathophysiological mechanisms of PE and the development of prevention and targeted therapy [1, 2]. The aim of our work was to identify placental tissue transcriptome patterns characteristic of women with PE and the physiological pregnancy and to assess the role of polymorphic markers of genes differentially expressed in pathological and normal pregnancy in the structure's of predisposition to PE.

*Methods and Algorithms:* We performed a genome-wide analysis of expression profiles of 24 placental tissue samples from patients with PE and women with physiological pregnancy using microarray technology. To search for differentially expressed genes (DEGs) in the case-control groups, the method of generalized linear models of the limma software package was used. A change in the level of gene expression (FC) by 1.5 times or more was considered significant at an adjusted significance level of  $p \leq 0.1$ . Functional annotation and functional cluster analysis of DEG groups were performed using the DAVID Web tool. Gene networks were constructed using the STRING 9.0 software.

The sample size for the analysis of DNA variability included more than 1000 women. They were divided into a group of patients with PE ( $N = 721$  people) and a control group ( $N = 552$  people) according to the course and outcomes of pregnancy. Genotyping of 85 polymorphic markers (tagSNPs) of differentially expressed genes was performed using MALDI-TOF mass spectrometry. To compare the frequencies of alleles and genotypes between the analyzed groups, Pearson's  $\chi^2$  test with Yates correction or two-tailed Fisher's exact test was used. To assess associations of SNP genes with the development of a pathological phenotype (PE), the odds ratio (OR) was calculated. The present study was approved by the ethics committee at the Research Institute of Medical Genetics of the Tomsk National Medical Research Center.

*Results:* Sixty three genes are statistically significantly differentially expressed between patients with PE and physiological pregnancy were found. The cluster of DEGs overexpressed in PE included not only known candidate genes previously identified in many genome-wide studies of placental gene expression profiles in preeclampsia (e.g., *LEP*, *BHLHB2*, *SIGLEC6*, *RDH13*, *BCL6*), but also new potential genes candidates (*CORO2A*, *SYDE1*, *PLIN2*, *CEBPA*, *HK2*, *NDRG1*, *ERRF1*, *EFNB1*, *GFOD2*, *NCOR2*,

*HMHA1, HERPUD1, KIF2A*), whose association with the development of PE has been established either in single studies or for the first time in the presented work [3].

The functional annotation of DEG indicates a number of biological processes that play an important role in the molecular pathogenesis of PE. These are reactions associated with the immune response, intercellular interaction, regulation of apoptosis, etc. An analysis of the metabolic pathways of DEG indicates a possible involvement in the pathophysiological mechanisms of PE cytotoxicity caused by NK cells, transendothelial migration of leukocytes and signaling pathways mediated by GTPase activators. 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. Analysis of this network identifies the cluster of co-expression genes including the *RAC2, CYBA, TYROBP, HMHA1, ITGB2, LYN*, and *LCPI* genes. In addition, pairs of *LEP*/its receptor *SIGLEC 6*, ephrin and its kinase *LYN* are of particular interest. Integration of DEG functional annotation data and analysis of network interactions of proteins encoded by these genes made it possible to identify a cluster of the most significant DEGs, whose hereditary variability was studied at the genomic level.

The association tagSNPs of 27 genes (*ANKRD37, BHLHE40, CORO2A, GPT2, HK2, INHA, LEP, LHB, NDRG1, PLIN2, PPP1R12C, SASH1, SIGLEC6, SYDE1*, and *ZNF175*) with PE was found. It should be noted that the majority of PE-associated tagSNPs, according to the results of analysis in the online resources HaploReg (v.4.1.) and RegulomeDB (<http://compbio.mit.edu/HaploReg>; <https://regulomedb.org/regulome-search/>) were localized at transcription factor binding sites and had a high regulatory potential.

**Conclusion:** We used a new approach to detecting genetic markers of preeclampsia based on a combination of genomic, transcriptomic, and bioinformatic approaches. We have discovered new genetic markers of PE and demonstrated a significant role of regulatory regions of the human genome in predisposition to this pathology. The functional annotation of DEGs demonstrated their overrepresentation in the processes associated with the immune response, intercellular interactions, regulation of apoptosis which probably play a key role in the molecular pathogenesis of PE.

**Acknowledgements:** The study is supported by the Russian Foundation for Basic Research (grant No. 20-34-90128).

### References

1. Orabona R. et al. The impact of preeclampsia on women's health: cardiovascular long-term implications. *Obstet Gynecol Surv.* 2020;75(11):703-709.
2. Staff A.C. The two-stage placental model of preeclampsia: An update. *J Reprod Immunol.* 2019;134-135:1-10.
3. Vennou K.E. et al. Meta-analysis of gene expression profiles in preeclampsia. *Pregnancy Hypertens.* 2020;19:52-60.

# Investigation of domain-specific functions and stability of VPS33A, responsible protein for mucopolysaccharidosis-plus syndrome

Vasilev F.<sup>1\*</sup>, Kawakami Y.<sup>2</sup>, Otomo T.<sup>2</sup>

<sup>1</sup> M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup> Kawasaki Medical School, Kurashiki, Japan

\*ff.vasilev@s-vfu.ru

**Key words:** lysosomal storage diseases, mucopolysaccharidoses, Yakuts

*Background and Objective:* In 2017, we reported a novel disease named mucopolysaccharidosis-plus syndrome (MPSPS, OMIM #617303). MPSPS is an autosomal recessive disorder caused by the mutation p.R498W in the *VPS33A* gene, a member of HOPS complex which coordinate fusion between endosome/autophagosome and lysosome [1]. Its clinical phenotype is similar to mucopolysaccharidoses or mucolipidoses, and 20 patients have been found among native population of Yakutia in Russia so far [2]. Previously, we showed that pathogenic mutation did not affect endocytic and autophagic pathways. Since detailed pathophysiology of MPSPS is not elucidated, we investigated stability of VPS33A protein and domain-specific functions.

*Methods:* The wild-type and mutant pcDNA3.1(-)-VPS33A (D2-1, p.R200P; D2-2, p.R498W; D3a, p.D251E; D3b, p.Y438D) expression vectors were prepared, and transfected by lipofection. Knock-down of *VPS33A* was performed with commercially available siRNAs. Cells were incubated in the presence or absence of cycloheximide (CHX), proteasome (MG132) and lysosome (BafA1, NH<sub>4</sub>Cl) inhibitors. We used the CRISPR-Cas9 genome editing technology for preparing cell lines containing different in-frame mutations of *VPS33A* gene in HeLa cells. Protein amount was analyzed by immunoblotting. Autophagic and endocytic functions were checked by LC3 expression and EGFR degradation assays respectively.

*Results:* We revealed that p.Y438D mutant, which disrupt HOPS/CORVET complexes formation, have no effect on VPS33A protein stability. Other mutations reduced the abundance of VPS33A protein. Autophagic flux and endocytosis were blocked in all clones with in-frame mutations in domains 2-2, 3a and 3b. Cells with mutations in domains 1 and 2-1 showed normal endocytic and autophagic function. Our study showed that ~20 % of normal protein is enough to maintain known functions. Results from knock-down of wild type *VPS33A* experiment suggested that reduction of VPS33A affects more on autophagic than endocytic function. Wild type or p.Y438D mutant VPS33A protein was maintained with MG132 treatment, suggesting that they are degraded in proteasome. However, other mutants were mostly maintained after treatment with lysosomal inhibitor, which suggests that they are degraded in lysosome.

*Conclusion:* Mutations in *VPS33A* gene may cause decrease in protein levels due to post-transcriptional modifications. We speculated that although p.R498W mutation is not disturb autophagic and endocytic functions, domain 2-2 of VPS33A might be involved in these intracellular pathways. In MPSPS patient's skin fibroblast, VPS33A protein with p.R498W is reduced into ~40 % of normal, and autophagic/endocytic function was not

affected. Our study showed that ~20 % of normal protein level does not affect autophagic/endocytic function so much, and supports our previous data from patient's skin fibroblast. Results of experiments with proteasomal and lysosomal inhibitors suggest that several mutants escape from ordinary proteasomal degradation pathway and go into lysosomal degradation, although detailed mechanisms including the involvement of autophagy are not clear. Investigation of the VPS33A protein will offer new insights and much more detailed molecular understanding of HOPS and CORVET complexes. Research is on-going.

*Acknowledgements:* This work was supported by the Ministry of Education and Science of Russian Federation (No. FSRG-2020-0014) and by the Grant for JSPS Research Fellow.

### *References*

1. Kondo H. et al. Mutation in VPS33A affects metabolism of glycosaminoglycans: a new type of mucopolysaccharidosis with severe systemic symptoms. *Hum Mol Genet.* 2017;26(1):173-183.
2. Vasilev F. et al. Mucopolysaccharidosis-Plus Syndrome. *Int J Mol Sci.* 2020;21(2):421.



## Epigenetic landscape in human aneuploid embryos

Vasilyev S.<sup>1\*</sup>, Tolmacheva E.<sup>1</sup>, Demeneva V.<sup>1</sup>, Zhigalina D.<sup>1</sup>, Vasilyeva O.<sup>1</sup>, Zuev A.<sup>1</sup>, Filatova S.<sup>2</sup>, Nikitina T.<sup>1</sup>, Sazhenova E.<sup>1</sup>, Markov A.<sup>1</sup>, Zarubin A.<sup>1</sup>, Kazhevarova A.<sup>1</sup>, Lebedev I.<sup>1</sup>

<sup>1</sup> Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia

<sup>2</sup> Biological Institute, National Research Tomsk State University, Tomsk, Russia

\* stanislav.vasilyev@medgenetics.ru

**Key words:** epigenetic landscape, DNA methylation, aneuploidy, spontaneous abortion, RRBS

*Motivation and Aim:* During gametogenesis and early embryogenesis parental genomes undergo two waves of DNA methylation reprogramming. Abnormal methylation due to errors in the reprogramming process can potentially lead to aneuploidy, which is the main cause of early fetal death in humans. On the other hand, the presence of an entire additional chromosome, or chromosome loss, can affect the global genome methylation level. Therefore, in this study, a genome-wide DNA methylation profile was analyzed in cells of human embryos with aneuploidy or normal karyotype at the first trimester of pregnancy.

*Methods and Algorithms:* As a marker of global genome methylation, the level of LINE-1 methylation was assessed using targeted bisulfite massive parallel sequencing in the chorionic trophoblast of 136 spontaneous abortions with normal karyotype, 96 spontaneous abortions with autosomal trisomy, 41 spontaneous abortions with monosomy X and 32 induced abortions. To analyze the effects of trisomy 16 we used samples of chorionic villi from 14 first-trimester miscarriages with trisomy 16 and from 6 induced abortions, for which a genome-wide analysis of methylation using the Infinium HumanMethylation27 BeadChip (Illumina, USA) was previously performed. Then the methylation index of individual genes was analyzed in chorionic villi of 29 first-trimester miscarriages with trisomy 16 using targeted bisulfite massive parallel sequencing. To analyze the effects of monosomy X, genome-wide methylation profiling was carried out using reduced representation bisulfite sequencing (RRBS) in spontaneous abortions with monosomy X with paternal (n = 4) or maternal (n = 4) X chromosome and induced abortions (n = 3).

*Results:* The LINE-1 methylation level was elevated statistically significant in chorionic villi of miscarriages with both trisomy of autosomes ( $44.2 \pm 4.7\%$ ) and monosomy X ( $42.1 \pm 5.1\%$ ) compared with that in induced abortions ( $39.9 \pm 2.4\%$ ) ( $p < 0.05$ ). The effects of the two most frequent aneuploidies (trisomy 16 and monosomy X) have been studied in more detail. In chorionic villi of miscarriages with trisomy 16, 97 differentially methylated sites in 91 genes were identified ( $FDR < 0.05$  and  $\Delta\beta > 0.15$ ) using DNA methylation arrays. Most of the differentially methylated genes were found to encode secreted proteins, signaling peptides, and receptors with disulfide bonds. Subsequent analysis using targeted bisulfite massive parallel sequencing showed hypermethylation of the promoters of specific genes in miscarriages with trisomy 16 but not in miscarriages with other aneuploidies. Some of the genes were identified as responsible for the development of the placenta and embryo (*GATA3-AS1*, *TRPV6*, *SCL13A4*, and *CALCB*) and the formation of the mitotic spindle (*ANKRD53*). RRBS analysis revealed 2,983

differentially methylated sites (DMS) (1,260 hypermethylated and 1,723 hypomethylated) in spontaneous abortions with monosomy X. Significant differences were found in the spectrum of DMS between groups with monosomy X of paternal and maternal origin. However, homeobox developmental proteins were enriched in both groups among the genes in which DMS were located.

*Conclusion:* Our results indicate that the pathogenic effects of aneuploidy in human embryogenesis can be supplemented with significant epigenetic changes in the repetitive sequences. Aberrant methylation of genes may lead to a decrease in their expression and to impaired trophoblast differentiation and invasion. Further research is needed to improve the understanding of the mechanisms for ensuring the genome stability in the early stages of human embryonic development and identify their possible contribution to human embryonic death.

*Acknowledgements:* The study was supported by the Russian Science Foundation (19-74-10026).

## The spectrum of mutations in the *EXT1* gene among patients with Multiple Hereditary Exostoses in the Republic of Sakha (Yakutia)

Yakovleva A.<sup>1\*</sup>, Danilova A.<sup>1</sup>, Petukhova D.<sup>1</sup>, Golikova P.<sup>1</sup>, Fedorov A.<sup>1</sup>, Nikolaeva I.<sup>2</sup>, Sukhomyasova A.<sup>1,2</sup>, Maksimova N.<sup>1</sup>

<sup>1</sup> M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup> Republican Hospital No. 1 – National Medical Center, Yakutsk, Russia

\* ae.iakovleva@s-vfu.ru

**Key words:** hereditary multiple exostoses, multiple osteochondromas, *EXT1*, massive parallel sequencing (MPS), Yakuts

*Motivation and Aim:* Multiple hereditary exostoses (MHE) or multiple osteochondromas (MO) (OMIM 133700, OMIM 133701) is a genetically heterogeneous disease with an autosomal dominant mode of inheritance. The most common cause is mutations in the *EXT* genes, which are responsible for at least 70 % of all cases of MHE. Three loci of the *EXT1*, *EXT2*, and *EXT3* genes have been described, mapped at 8q24, 11p12, and 19p respectively. Mutations in the *EXT2* gene are three times less common than in the *EXT1* gene; only a few descriptions are known for the *EXT3* locus [1]. MHE intensively develops during the period of physiological growth of the skeleton, in childhood and adolescence, it is represented by generalized forms of skeletal damage with numerous progressive deformities of bones and joints, shortening and secondary changes in bones, which can cause clinical, socio-psychological problems that affect the quality of life of the patient [2, 3]. Exostoses cause compression of large vessels and nerves, which requires repeated surgical intervention to remove them during the period of growth and development of the organism [4, 5].

On the territory of Russia, studies of the molecular genetic cause of MHE are rare, and in the Republic of Sakha (Yakutia) such studies have not been carried out. Therefore, the study of multiple hereditary exostoses in Yakutia is both a fundamental scientific problem and a medical and social task for diagnosing, preventing and predicting the course of this hereditary disease. This paper presents the results of a molecular genetic study to search for mutations in the *EXT1* gene among patients with MHE in Yakutia.

*Materials and Methods:* Molecular genetic study was carried out among 57 patients with MHE. For verification and validation of molecular genetic analyzes, DNA extraction, polymerase chain reaction, clinical exome sequencing on a high-performance MiSeq sequencer (Illumina, USA) using a Trusight One Sequencing panel (Illumina, USA) and Sanger sequencing were used. Variant Studio 3.0 programs, ExAC, 1000G and ESP6500 bases were used for bioinformatic analysis, and SIFT, PolyPhen2, MutationTaster software tools were used to check the functional significance *in silico*.

*Results:* Only 8 mutations were found in the *EXT1* gene, of which were previously published: in exon 1, c.854dupA (p.His285Glnfs\*4); in intron region 9, c.1883+2T > A; in exon 11, c.2101C > T (p.Arg701\*); previously unreported: in exon 2 – c.970delT (p.Tyr324Phefs\*35), c.1003delC (p.Leu335Trpfs\*24), c.1019G > A (p.Arg340His); in exon 4, c.1171delT (p.Ser391Leufs\*12), in exon 8, c.1703\_1713delCGGTGCTTTCA

(p.Thr568Asnfs\*16). According to the results of the study, mutations in the *EXT1* gene were found in 22.8 % of the examined patients.

*Conclusion:* The results of this study made it possible to expand the pathological spectrum of mutations in the *EXT1* gene and to establish the etiology of the MHE disease. New data have been obtained, which can be further used in the diagnostic program of patients with MHE.

*Acknowledgements:* Molecular genetic research was carried out at the Center for Collective Use of the Arctic Innovation Center "M.K. Ammosov North-Eastern Federal University".

The work was carried out within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation. (Topic of the scientific project: "Genomics of the Arctic: epidemiology, heredity, pathology", FSRG-2020-0014.)

### *References*

1. Lagunova I.G. Clinical and radiological diagnosis of skeletal dysplasia. Moscow: Medicine, 1989.
2. Shishkina N.S. et al. Combination of diabetes mellitus and syndrome of multiple exostoses osteochondral chondrodysplasia. *Diabetes*. 2004;4:34-36.
3. Fokhtin V.V. et al. Treatment results of the children with exostotic chondrodysplasia of complex anatomic localization. *Children Hospital*. 2014;1:13-16.
4. Chesnokova G.G. The Study of Structural Anomalies and Point Mutations of the *EXT1* and *EXT2* Genes in Multiple Exostoses Chondrodysplasia and Sporadic Malignant Neoplasms: Abstract of the thesis of a candidate of biological sciences: Moscow: Medical Genetic Scientific RAMS center, 1999.
5. Wang Y. et al. *EXT1* and *EXT2* Variants among 22 Chinese Families with Multiple Osteochondromas: Seven New Variants and Potentiation of Preimplantation Genetic Testing and Prenatal Diagnosis. *Front Genet*. 2020;11:607838.

## **Научно-техническая программа Союзного государства «ДНК-идентификация»: геногеографические и геномные технологии идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства**

Yankovsky N.K.<sup>1\*</sup>, Kilchevsky A.V.<sup>2</sup>

<sup>1</sup> *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

<sup>2</sup> *Институт генетики и цитологии НАН Белоруссии, Минск, Белоруссия*

\* *yankovsky@vigg.ru*

В 2017–2021 гг. в Российской Федерации и Республике Беларусь реализована и полностью выполнена программа Союзного государства «Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства». Разработаны отечественные импортозамещающие или не имеющие аналогов генетические технологии и методики, которые превышают или соответствуют по показателям зарубежные аналоги, а также наборы отечественных реагентов для применения на отечественном оборудовании для криминалистических и медицинских исследований.

В докладе на BGRS2022 будут представлены разработанные по программе «ДНК - идентификация» технологии определения по ДНК:

- вероятной внешности неизвестного индивида (цвет глаз и волос);
- возраста индивида по метилированию его ДНК;
- наиболее вероятного этногеографического происхождения индивида;
- популяционной принадлежности индивида к определенному народу, роду, семье;
- вероятного психоэмоционального статуса по сочетанию уровней метаболитов индивида;
- риска развития заболеваний (сердечно-сосудистые, онкология, аутоимунные, костные).

Разработаны отечественные импортозамещающие или не имеющие аналогов наборы реагентов, необходимые для применения разработанных или имеющихся ДНК-технологий или методик, в том числе для работы с предельно малым количеством ДНК и с глубоко деградированной ДНК, взятых с места преступления, аварии или катастрофы.

Разработана методика прогнозирования изменений этнонационального состава населения мегаполисов Союзного государства под воздействием миграционных процессов. Методики могут стать основой для построения моделей учета и динамики миграций в Союзном государстве как трудовой, так и теневой.

## RMetSeq package for processing targeted MRE-seq

Zarubin A.\*, Sivtsev A., Zhalsanova I., Skryabin N.

Tomsk National Research Medical Center, Research Institute of Medical Genetics, Tomsk, Russia

\* a.a.zarubin@gmail.com

**Key words:** MRE-seq, DNA methylation, NGS

*Motivation and Aim:* Methylation-sensitive restriction enzyme sequencing or MRE-seq is one of the methods for determining the level of DNA methylation. It usually involves comparing the results of sequencing samples after exposure to methyl sensitive and insensitive enzymes targeting the same restriction sites. This approach has some limitations, which leads to the fact that this method has not found wide application. We propose a new approach to the design of experiments and analysis of MRE-seq data, which allows integration of this analysis with standard NGS sequencing of exome and target panels with probe enrichment.

*Methods and Algorithms:* DNA isolated from 9 blood samples was sonicated in 300 and 500 bp modes. Thereafter, half of each sample was treated with the HpaII. All samples were enriched using SureSelect Custom DNA Target Enrichment Probes (which includes exomes of 37 genes of a total length of 168497 nucleotides) and sequenced using MiSeq. Genotyping analysis was performed using GATK4. The methylation level was assessed using the RMetSeq package published on [github.com/alekseizarubin/RMetSeq](https://github.com/alekseizarubin/RMetSeq).

*Results:* In the studied genomic regions, 159 restriction sites were found (with a coverage of more than 20). The methylation level ranged from 0.3 to 1. To assess the possibility of qualitative genotyping on MRE-seq data, modeling was performed on native DNA sequencing samples, DNA treated with restriction enzyme and a mixed sample. In the treated with restriction enzyme,  $\approx 3\%$  of the target area is not available for analysis. The mixed sample showed the quality of the native sample, but required 5–10 % more than the average coverage for this. On average coverage of mixed samples, with an average coverage of 70X, 97.7 % of the target coverage of the region is more than 10.

*Conclusion:* Targeted MRE-seq and the RMetSeq package make it possible to simultaneously determine the level of DNA methylation without compromising the quality of genotyping in standard sequencing using probe enrichment.

## 4.2 Section “Genome-wide association studies”



## Association of *PANX3* with chronic back pain discovered using gene-based analysis

Belonogova N.M.<sup>1\*</sup>, Kirichenko A.V.<sup>1</sup>, Freidin M.B.<sup>2</sup>, Williams F.M.K.<sup>2</sup>, Suri P.<sup>3,4,5</sup>, Aulchenko Y.S.<sup>1,6</sup>, Axenovich T.I.<sup>1</sup>, Tsepilov Y.A.<sup>1,6,7</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Twin Research and Genetic Epidemiology, King's College London, London, UK*

<sup>3</sup> *Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, USA*

<sup>4</sup> *Division of Rehabilitation Care Services, Seattle, USA*

<sup>5</sup> *Clinical Learning, Evidence, and Research Center, University of Washington, Seattle, USA*

<sup>6</sup> *Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia*

<sup>7</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *belon@bionet.nsc.ru*

**Key words:** back pain, gene-based analysis, GWAS summary statistics, intervertebral disc disorder

*Motivation and Aim:* Back pain is the leading cause of years lived with disability worldwide, yet surprisingly little is known regarding the biology underlying this condition. The impact of genetics is known for the chronic back pain: its heritability is estimated to be at least 40 % [1]. Large genome-wide association studies (GWAS) have shown that common variation may account for up to 35 % of chronic back pain heritability; rare variants may explain a portion of the heritability not explained by common variants [2]. In this study, we aimed to analyze the genes whose variants are directly responsible for differences in chronic back pain risk between individuals using data from UK Biobank.

*Methods and Algorithms:* For gene-based association analysis we used GWAS summary statistics for chronic back pain obtained on the UK Biobank imputed data including rare variants with moderate imputation quality,  $N = 439,831$  participants of European ancestry [3]. A sample of 265,000 randomly selected participants was used as the discovery sample. The remaining participants comprised a replication cohort. We used the sumFREGAT R-package [4] to perform three gene-based methods: SKAT-O, PCA, and ACAT-V. The results from these methods were combined using the aggregated Cauchy omnibus test, ACAT-O [5]. Genes having a combined ACAT-O p-value  $\leq 2.5 \times 10^{-6}$  (standard gene-based significance threshold) were considered as genome-wide significant genes. Genes with p-value  $\leq 2.5 \times 10^{-5}$  were considered as suggestively significant. For genes with genome-wide significant p-values, replication was carried out on the sample of remaining UK Biobank participants ( $N = 174,831$ ). An additional validation step was performed using FinnGen Biobank GWAS results version 6 (FinnGen-r6, <https://r6.finnngen.fi/>).

*Results:* We obtained 39 suggestively significant genes on discovery stage. Two of them, *SOX5* and *PANX3*, reached the genome-wide significance threshold ( $2.5 \times 10^{-6}$ ) and were further replicated with p-value  $< 0.025$  (0.05/2). The *SOX5* gene is well known back pain gene. The *PANX3* gene has not previously been described as having a role in chronic back pain. We showed that the association of *PANX3* with chronic back pain is



driven by rare non-coding intronic polymorphisms. Association of *PANX3* has been replicated on independent sample from UK Biobank and validated using similar phenotype, dorsalgia, from FinnGen Biobank. We also found that the *PANX3* gene is associated with intervertebral disc disorders on FinnGen sample. We can speculate that a possible mechanism of action of *PANX3* on the back pain is due to its effect on the intervertebral discs.

*Conclusion:* We found that *SOX5* and *PANX3* genes influenced chronic back pain. The *PANX3* gene has not previously been described as associated with chronic back pain. Its association was shown to be due to non-coding rare intron polymorphisms.

*Acknowledgements:* The study was conducted using the UK Biobank resource under application No. 59345. The work was supported by the Russian Foundation for Basic Research No. 20-04-00464, the budget project No. FWNR-2022-0020, the Russian Ministry of Education and Science under the 5-100 Excellence Programme, the Russian Science Foundation No. 22-15-20037, and NIH/NIAMS P30AR072572.

### *References*

1. Hartvigsen J. et al. Heritability of spinal pain and consequences of spinal pain: a comprehensive genetic epidemiologic analysis using a population-based sample of 15,328 twins ages 20-71 years. *Arthritis Rheum.* 2009;61(10):1343-1351.
2. Wainschtein P. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet.* 2022;54(3):263-273.
3. Tsepilov Y.A. et al. Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun Biol.* 2020;3(1):329.
4. Svishcheva G.R. et al. Gene-based association tests using GWAS summary statistics. *Bioinformatics.* 2019;35(19):3701-3708.
5. Liu Y. et al. 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.

## Genome-Wide detection of human regulatory SNPs based on analysis of ChIP-seq and RNA-seq complex data

Damarov I.<sup>1\*</sup>, Merkulova T.<sup>1,2</sup>, Bryzgalov L.<sup>1</sup>, Ustrokhanova D.<sup>2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* damarovis@bionet.nsc.ru

**Key words:** regulatory SNP, ChIP-seq, GWAS

*Motivation and Aim:* According to modern concepts, non-coding SNPs, including regulatory (rSNPs), involved in the control of gene expression, make the greatest contribution to the formation of the phenotype, accounting for more than 90 % of the SNPs identified in GWAS [1]. However, GWAS does not provide information about the functionality of the found variants, thus preventing progress in understanding the mechanisms underlying the development of certain diseases. Therefore, functional genomics data are widely used for the functional interpretation of the found variants. These same approaches are applied independently, regardless of the identified regulatory variants in the GWAS. The aim of this study is a genome-wide search for human regulatory SNPs based on the analysis of ChIP-seq and RNA-seq data.

*Methods:* To search for rSNPs, we used an approach based on the identification of allele-specific events in experimental data to determine the profile of histone modification of open chromatin H3K4me3 (ChIP-seq) and transcriptome (RNA-seq) obtained for peripheral blood lymphocytes from 20 donors. To determine the functional significance of potentially regulatory SNPs using GTEx (The Genotype-Tissue Expression) data [2], a random forest algorithm was developed, which, based on the specified parameters, is able to effectively classify the SNP according to whether it is eQTL / is not eQTL.

*Results:* ChIP-seq and RNA-seq libraries were prepared and sequenced, averaging 20 and 73 million pair-end reads per sample respectively. As a result of mapping data to the reference genome, a list of rSNPs was formed, demonstrating an allelic imbalance in the data on the representation of H3K4me3 in the promoter regions and the expression of the same genes. For each rSNP from this list, the probability of belonging to eQTL was predicted. Several rSNPs classified as eQTLs significantly change the sequence of transcription factor binding sites, which confirms their regulatory function. Connecting GWAS data provided the association of several identified eQTLs with phenotypic traits.

*Conclusion:* As a result of the analysis of the obtained experimental data, a panel of about 2 thousand rSNPs was formed, which with a high probability can be eQTL.

### References

1. GWAS Catalog [Electronic resource]. URL: <https://www.ebi.ac.uk/gwas/home>. Accessed 18.03.2022.
2. Genotype-Tissue Expression Project (GTEx) [Electronic resource]. URL: <https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project>. Accessed 18.03.2022.

# Bidirectional Mendelian randomization study reveals causal effects of psychosocial factors on chronic back pain and vice versa

Elgaeva E.<sup>1,2\*</sup>, Williams F.<sup>3</sup>, Freidin M.<sup>3</sup>, Zaytseva O.<sup>4</sup>, Aulchenko Y.<sup>1,5,6</sup>, Tsepilov Y.<sup>1,2</sup>, Suri P.<sup>7,8,9,10</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK

<sup>4</sup> Genos Glycoscience Research Laboratory, Zagreb, Croatia

<sup>5</sup> PolyOmica, 's-Hertogenbosch, Netherlands

<sup>6</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>7</sup> Division of Rehabilitation Care Services, VA Puget Sound Health Care System, Seattle, USA

<sup>8</sup> Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, USA

<sup>9</sup> Department of Rehabilitation Medicine, University of Washington, Seattle, USA

<sup>10</sup> Clinical Learning, Evidence, and Research (CLEAR) Center, University of Washington, Seattle, USA

\* elgaeva@bionet.nsc.ru

**Key words:** low back pain, epidemiology, prognosis, risk factor, causation

*Motivation and Aim:* Risk factors for chronic back pain (CBP) are difficult to study utilizing traditional approaches due to shared genetic background [1]. We performed a bidirectional Mendelian randomization (MR) study to examine causal relationships between risk factors (education, smoking, alcohol intake, exercises, sleep and depression) and CBP.

*Methods and Algorithms:* We identified instrumental variables for risk factors and CBP from the largest published genome-wide association studies of these traits conducted in Europeans. We carried out inverse weighted variance meta-analysis (IVW) [2], Causal Analysis Using Summary Effect (CAUSE) [3] and sensitivity analyses to investigate the evidence for causality. We interpreted the exposure-outcome associations as being consistent with a hypothesis of causality if results of IVW or CAUSE were statistically significant after correction for multiple testing ( $p < 0.003$ ), and the direction and magnitude of effect were concordant between IVW, CAUSE, and sensitivity analyses.

*Results:* We observed statistically significant causal associations between years of schooling (OR = 0.54 per increase of ~4 years of schooling), ever smoking (OR = 1.27), alcohol consumption (OR = 1.29) and major depressive disorder (OR = 1.41) and the CBP risk. We also detected a significant causal effect of CBP on greater alcohol consumption (OR = 1.19) and smoking (OR = 1.21).

*Conclusion:* Thus, fewer years of education, smoking, greater alcohol intake, and depression increase CBP risk. CBP increases the risk of greater alcohol intake and smoking.

*Acknowledgements:* Dr. Tsepilov and Ms. Elgaeva were supported by a grant from the Russian Science Foundation (RSF) No. 22-15-20037. Dr. Aulchenko was supported by

the Russian Ministry of Education and Science under the 5-100 Excellence Programme and by the budget project No. FWNR-2022-0020. Dr. Suri is an employee of the VA Puget Sound Health Care System and affiliated with the University of Washington Clinical Learning, Evidence and Research (CLEAR) Center, which was funded by NIAMS/NIH P30AR072572. The contents of this work do not represent the views of the US Department of Veterans Affairs, the National Institutes of Health, or the US Government. The study was conducted using the UK Biobank Resource under project 18219. We are grateful to the UK Biobank participants for making such research possible. We thank Dr. Shadrina A. for the help with the development of the MR protocol.

### *References*

1. Suri P., Boyko E.J., Smith N.L., Jarvik J.G., Williams F.M.K., Jarvik G.P. et al. Modifiable risk factors for chronic back pain: insights using the co-twin control design. *Spine J.* 2017;17(1):4-14.
2. Hemani G., Zheng J., Elsworth B., Wade K.H., Haberland V., Baird D. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7:e34408.
3. Morrison J., Knoblauch N., Marcus J.H., Stephens M., He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet.* 2020;52(7):740-747. doi: 10.1038/s41588-020-0631-4.

# A platform for case-control matching enables association studies without genotype sharing

Loboda A.<sup>1\*</sup>, Artomov M.<sup>1,2</sup>, Artyomov M.<sup>3</sup>, Daly M.<sup>2,4</sup>

<sup>1</sup>ITMO University, St. Petersburg, Russia

<sup>2</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

<sup>3</sup>Department of Immunology and Pathology, Washington University in St. Louis, St. Louis, MO, USA

<sup>4</sup>Institute for Molecular Medicine Finland, Helsinki, Finland

\* [aleks.loboda@gmail.com](mailto:aleks.loboda@gmail.com)

**Key words:** GWAS, Privacy Preserving, Database, Exome Sequencing

*Motivation and Aim:* Acquiring a sufficiently powered cohort of control samples can be time consuming or, sometimes, impossible. Accordingly, an ability to leverage control samples that were already collected and sequenced elsewhere could dramatically improve power in genetic association studies. Majority of the genotyped and sequenced human DNA samples to date are subject to strict data sharing regulations, large-scale sharing of, in particular, control samples is extremely challenging. We developed a method allowing selection of the best-matching controls in an external pool of samples that is compliant with personal genotype data protection restrictions.

*Methods and Algorithms:* We provide a web platform that stores 39,844 exome sequencing samples available for control selection and a complimentary R-package to be used on a user side for generation of anonymous data from case genotypes that will be uploaded to the web-platform.

*Results:* Our approach uses singular value decomposition of the matrix of case genotypes to rank external controls by similarity to cases. We demonstrate that this recovers an accurate case-control association results for both ultra-rare and common variants. Finally, we provide a free access to a database of 39,844 exomes to be used as external controls that enables association studies for case cohorts lacking control subjects.

*Conclusion:* We present a freely accessible resource with a large-scale control database enabling association studies for “case-only” sequencing data with carefully selected controls and controllable error rates.

## References

1. Bodea C.A., Neale B.M., Ripke S., Barclay M., Peyrin-Biroulet L., Chamaillard M., ..., Ferrier N. A method to exploit the structure of genetic ancestry space to enhance case-control studies. *Am J Hum Genet.* 2016;98(5):857-868.

# The influence of environmental and genetic factors on the portability of polygenic risk score models across diverse ancestry groups

Nostaeva A.<sup>1\*</sup>, Kuznetsov I.<sup>2</sup>, Zhaivoron A.<sup>3</sup>, Aulchenko Y.<sup>1,4</sup>, Karssen L.<sup>4</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Skolkovo Institute of Science and Technology, Moscow, Russia*

<sup>3</sup> *Novel Software Systems, Novosibirsk, Russia*

<sup>4</sup> *PolyKnomics, 's-Hertogenbosch, Netherlands*

\* *avnostaeva@gmail.com*

**Key words:** genome-wide association study, summary statistics, polygenic risk score, linkage disequilibrium, single-nucleotide polymorphisms

*Motivation and Aim:* The use of polygenic risk scores (PRS) has become widespread in biomedical and social science disciplines. Polygenic scoring studies have demonstrated reliable, though modest, prediction using straightforward scoring methods for many complex genetic phenotypes.

However, a historical tendency to use European ancestry samples hinders the use of PRS in non-European ancestry populations. Although there is good reason to anticipate reduced predictive power in non-European ancestry samples, the magnitude of performance decrease, as well as the factors affecting this decline, are not well known. In this work we try to answer questions about the predictive performance of PRS across diverse ancestry groups and how it depends on environmental and genetic factors.

*Introduction:* Many studies have shown that PRS is a powerful tool for prediction of the risk of future disease and values of quantitative traits [1, 2]. Through the initiatives of large consortiums and biobanks, sample sizes for genome-wide association studies (GWAS) have reached impressive levels, continuing to improve the predictive power of polygenic risk models [3, 4].

However, there is an over-representation of people of European ancestry in GWAS studies, which results in an under-exploration of non-European populations and a limited ability to use polygenic risk models for prediction in such populations. This problem has been widely recognized [5, 6].

Recent research has focused on the generalizability of polygenic scores to non-European ancestry populations [7]. Due to differences in variant frequencies and patterns of linkage disequilibrium, as well as in environment, predictive power in samples of non-European origin is reduced. However, the dependence of predictive quality on the listed factors is not well known. Thus, there is a need for systematic evaluation of PRS performance across multiple populations and phenotypes.

Here, we developed and validated genome-wide polygenic scores for three anthropometric traits: body mass index (BMI), height and weight, and analysed the influence of environmental and genetic factors on the portability of polygenic risk score models trained and calibrated in people of European ancestry to individuals of non-European descent.

*Methods and Algorithms:* We developed draft PRS models for each trait using the SBayesR [8] method that utilises summary statistics from GWAS. For GWAS we used a mixed linear model (MLM)-based approach, implemented in the GCTA software package [9], and a training dataset. We selected the best PRS model for each trait, using a validation dataset. Then, we used the selected model to calculate PRS for each person in the calibration dataset and estimated the effects of PRS, sex and age on the phenotype using a linear regression model. This calibrated predictive model was then transferred to other groups. We used not overlapping random samples of individuals of white British born in England from the UK Biobank data (UKB) [10] as the training, the validation, and the calibration datasets.

For quantitative traits, we characterised model performance by the proportion of variance of phenotype,  $R^2$ , explained by observed predictors. Estimation of the predictive performance of the model was made in a series of datasets. We used different cohorts from UKB, such as white British born in England, Wales and Scotland, Ireland, and outside of the UK (“other Europeans”), as well as samples of individuals of Pakistani, Indian, African, Caribbean, and Chinese ancestry. We also added in this analysis cohorts from other consortiums, which include individuals of British, American, Belgian, Croatian and Italian populations. To avoid “testimation” bias, we validated and estimated the models in independent data sets not overlapping with the samples used for GWAS.

As a genetic factor we used the fixation index ( $F_{st}$ ), which is a measure of population differentiation due to genetic structure. Regarding an environmental factor, we used the average value of a trait for each of the cohorts.

*Results:* Our results demonstrate a low portability of PRS from European ancestry to non-European for all considered traits. To illustrate, for BMI, it performed reasonably well for different groups of European ancestry ( $R^2$  varies from 9 to 15 %), but had a low predictive accuracy ( $R^2$  less than 6 %) for people of non-European descent.

These differences and a gradient in  $R^2$  have been studied in conjunction with genetic and environmental factors, where Pearson's correlation was observed. Results show, for example, that the correlation between the  $F_{st}$  parameter and the quality of BMI prediction was about  $-0.85$ , while for the mean value of this trait, the correlation was around  $-0.84$ .

*Conclusion:* We concluded that PRS models developed for European populations cannot be directly transferred to other populations. The model performance drops with increased genetic and environmental distances between the population where the model was developed (training and calibration) and the target population.

*Acknowledgements:* This research has been conducted using the UK Biobank Resource project 41601. This work was funded by PolyKnomics BV.

## References

1. Wray N.R., Goddard M.E., Visscher P.M. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev.* 2008;18:257-263.
2. Torkamani A., Wineinger N.E., Topol E.J. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19(9):581-590.
3. Lello L. et al. Accurate genomic prediction of human height. *Genetics.* 2018;210:477-497.
4. Lee J.J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* 2018;50:1112-1121.
5. Bustamante C.D., Burchard E.G., De la Vega F.M. Genomics for the world. *Nature.* 2011;475:163-165.
6. Petrovski S., Goldstein D.B. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* 2016;17:157.

7. Martin A.R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51:584.
8. Lloyd-Jones L.R., Zeng J. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun.* 2019;10(1):5086.
9. Jiang L., Zheng Z., Qi T. et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet.* 2019;51:1749-1755.
10. Bycroft C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203.



# New framework for gene-based association analysis using GWAS summary statistics and functional annotations

Svishcheva G.<sup>1,2\*</sup>, Belonogova N.<sup>1</sup>, Kirichenko A.<sup>1</sup>, Zorkoltseva I.<sup>1</sup>, Tsepilov Ya.<sup>1,3</sup>, Axenovich T.<sup>1,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [gulsvi@bionet.nsc.ru](mailto:gulsvi@bionet.nsc.ru)

**Key words:** gene-based association analysis, rare variants, summary statistics, genotypic correlations

**Motivation and Aim:** Gene-based association analysis using GWAS summary statistics is an effective tool for the mapping of common and rare genetic variants. However, none of the developed gene-based methods is a universal approach that would be effective for any trait with any genetic architecture. In this regard, the combination of different gene-based methods can be a universal and effective approach to the gene-based association analysis. Recently, several frameworks have been proposed to combine tests using various strategies for gene-based association analysis. They differ from each other in goals (one or more traits), input data (individual phenotypes and genotypes or the GWAS summary statistics) and ways to use functional annotations. However, these frameworks are not flexible, because they use a fixed set of methods, weights and combinations of functional annotations.

**Methods and Algorithms:** Here we propose a flexible framework for gene-based analysis using GWAS summary statistics. We implemented this framework in our sumFREGAT package [1], which now includes the aggregated Cauchy test, ACAT [2], and the probabilities of SNPs being causal [3] in Burden test, SKAT, SKAT-O, ACAT-V, PCA and FLM. For the new framework, the inputs are GWAS summary statistics, genotypic correlations, SNP-level weights defined by beta-distribution functions of MAFs and the probabilities of SNPs being causal, which are estimated by various functional annotations. First, the user arbitrarily defines a set of gene-based methods, a set of weights and a set of probabilities of SNPs being causal. Then all selected gene-based tests are run one-by-one, and the resulting test statistics are combined by ACAT.

**Results:** The new framework was successfully tested on neuroticism and coronary artery disease GWAS data from the UK Biobank and allowed us to identify several additional genes. The updated sumFREGAT is available at <https://cran.r-project.org/web/packages/sumFREGAT/index.html>.

**Conclusion:** The presented framework allows users to carry out a wide range of arbitrary defined gene-based association tests differing from each other by statistical method, weighting function, functional annotation, and then to combine these tests.

**Acknowledgements:** The study is supported by Russian Foundation for Basic Research (20-04-00464).

## References

1. 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.
2. 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.
3. Li X., Li Z., Zhou H., Gaynor S.M., Liu Y., Chen H., Sun R., Dey R., Arnett D.K., Aslibekyan S. Dynamic incorporation of multiple *in silico* functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet*. 2020;52(9):969-983.

## Defining the genetic control of N-glycosylation human blood plasma proteins using multivariate genetic association analysis

Timoshchuk A.<sup>1,2\*</sup>, Potapova N.A.<sup>3</sup>, Akmacic I.T.<sup>4</sup>, Lauc G.<sup>4</sup>, Sharapov S.<sup>1,3</sup>, Aulchenko Y.S.<sup>1,3</sup>

<sup>1</sup>MSU Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup>Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup>Genos Glycoscience Research Laboratory, Zagreb, Croatia

\*timoschukanny@gmail.com

**Key words:** human plasma N-glycome, GWAS, MANOVA approach, genetic control

*Motivation and Aim:* Biochemical pathways underlying protein N-glycosylation are well known, but the biological networks regulating these pathways remain poorly understood. Defining the genetic control of N-glycosylation of proteins of human blood plasma (“plasma N-glycome”) can shed light on the regulation of this complex process. To date, a series of genome-wide association studies (GWAS) of human blood plasma protein N-glycosylation have been published reporting more than 30 genetic loci affecting protein glycosylation [1-3]. These studies used a regression model under which glycome traits were analyzed independently of each other. However, it has been previously demonstrated that multivariate genetic association analysis of N-glycome, that is, a joint analysis of multiple glycomic traits, has higher power for loci detection [4, 5]. In this work we aimed to perform the first multivariate GWAS of total plasma N-glycome.

*Methods and Algorithms:* We utilized MANOVA approach for multivariate analysis of genetic associations. We defined a list of biochemically similar multivariate traits, grouping them on such characteristics as galactosylation and sialylation of antennary branches, core-fucosylation, bisecting GlcNAc, etc. After that, multivariate analysis was carried out using the summary statistics-based method described by Ning et al. [6]. This method enables multivariate analysis of variance using the results of multiple GWAS for individual traits. Moreover, our study included a replication experiment based on the protocol of Ning et al. [6] – reproduction of statistically significant associations in an independent set of samples. For loci that replicated in our study, we performed an *in silico* functional analysis.

*Results:* We identified 52 loci in total. Compared to results of previous studies we found 17 new loci associated with total plasma N-glycosylation and 7 loci that overlapped with those identified but not replicated previously. We replicated 8 loci and compiled a list of candidate genes for each of them.

*Conclusion:* Identification of new loci associated with total plasma N-glycosylation can provide new knowledge towards understanding the mechanisms of *in vivo* regulation of this process. Among the genes prioritized as possible candidates, several could be linked to N-glycosylation and/or immune function.

*Acknowledgements:* The work of NP, SS and YA was partly supported by a grant from the Russian Science Foundation (RSF) No. 19-15-00115.

## References

1. Lauc G. et al. The First GWAS Study of Human N-Glycome Identifies HNF1 $\alpha$  as a Master Regulator of Plasma Protein Fucosylation. *PLoS Genet.* 2010;6(12):e1001256.
2. Huffman J.E. et al. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet.* 2011;20(24):5000-5011.
3. Sharapov S.Z. et al. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum Mol Genet.* 2019;28(12):2062-2077.
4. Shen X. et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat Commun.* 2017;8(1):1-10.
5. Shadrina A.S. et al. Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. *Hum Mol Genet.* 2021;30(13):1259-1270.
6. Ning Z. et al. Nontrivial replication of loci detected by multi-trait methods. *Front Genet.* 2021;12:627989. doi: 10.3389/fgene.2021.627989.

# Association between polygenic risk scores for plasma protein N-glycosylation traits and 276 ICD-10 diseases

Zaytseva O.O.<sup>1\*</sup>, Nostaeva A.V.<sup>2</sup>, Sharapov S.Zh.<sup>2</sup>, Elgaeva E.E.<sup>2</sup>, Lauc G.<sup>1</sup>,  
Aulchenko Y.S.<sup>2</sup>, Tsepilov Y.A.<sup>2</sup>

<sup>1</sup> *Genos Glycoscience Research Laboratory, Zagreb, Croatia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [ozaitseva@genos.hr](mailto:ozaitseva@genos.hr)

**Key words:** plasma N-glycome, N-glycan, ICD-10 diseases, molecular marker, GWAS

*Motivation and Aim:* N-glycosylation, i.e., attachment of a carbohydrate to an asparagine residue in the peptide chain, affects the structure and function of proteins. N-glycosylation is involved in various biological processes, such as cell signaling, biomolecule trafficking, immune response, host-pathogen interactions, etc. It is also one of the most diverse post-translational modifications. N-glycosylation of a particular protein can vary greatly even within one organism and cell type. The spectrum of N-glycans found on human blood plasma proteins is altered in ageing, cancer, autoimmune, inflammatory, and other diseases, and plasma N-glycome is often regarded as a promising molecular marker of health status [1, 2]. N-glycosylation of proteins is defined by presence and activity of glycosyltransferases and glycosidases in the Golgi, the state of Golgi milieu, availability of activated sugars, as well as by the structure of glycosylated protein. Recent genome-wide association studies (GWAS) identified 15 genetic loci that are involved in regulation of the composition of the plasma proteins N-glycome, some of these loci are also associated with inflammatory, autoimmune, cancer and other diseases [3]. The aim of our study is to investigate the shared genetic architecture that underlies the plasma N-glycome composition and 276 diseases included in the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) list, GWAS summary statistic for which is available through the UK Biobank (UKB) database. To achieve that we have developed polygenic risk score (PRS) models for the 117 N-glycan traits and tested if these PRS are associated with the ICD-10 diseases in the UKBB cohorts.

*Methods and Algorithms:* PRSs for 117 plasma N-glycosylation traits were calculated using summary statistics of GWAS for each trait performed in a cohort of European descent (N = 7541) using the SbayesR software. The models were tested in independent validation data set of individual level data, sampled from the CEDAR cohort of European descent (N = 189). Then we calculated PRS values for every trait in a whole set of European ancestry individuals from UK Biobank (N<sub>total</sub> = 439,762).

To test the association between the N-glycosylation traits and ICD-10 diseases, we used logistic regression considering PRS for each glycan trait as a predictor for a disease phenotype. List of the diseases was taken from medical histories and questionnaires obtained from non-related UK Biobank participants of European descent (N = 374303). All medical codes were preliminary filtered by prevalence (> 0.5 % and < 99.5 %). To perform logistic regression analyses using standard glm() function in R and included sex, age, batch number and first ten principle components of the kinship matrix (PC1-10) as covariates in addition to the PRS predictor. Finally, we filtered out the results not passing

the significance threshold of  $p$ -value  $< 6.5 \times 10^{-6} = 0.05/(28 * 276)$ , where 28 is the number of plasma N-glycome principal components, and 276 is the numbers of ICD10 codes.

**Results:** We have found the PRS for 35 plasma N-glycosylation traits to be statistically significantly associated with at least one ICD-10 phenotype. Altogether, 18 ICD-10 phenotypes were statistically significantly associated with PRS for a least one plasma N-glycosylation trait. About a half of such ICD-10 phenotypes are related to the health of cardiovascular system (E78.0 Pure hypercholesterolaemia, E78.5 Unspecified hyperlipidaemia, I10 Essential (primary) hypertension, I20.9 Unspecified angina pectoris, I25.1 Atherosclerotic heart disease, I25.2 Old myocardial infarction, I25.8 Other forms of chronic ischaemic heart disease, I25.9 Chronic ischaemic heart disease, Z86.7 Personal history of diseases of the circulatory system, Z95.1 Presence of aortocoronary bypass graft). The lowest  $p$ -value of  $2.52 \times 10^{-48}$  for was observed for the association of PRS for the percentage of M9 glycan in the plasma N-glycome and pure hypercholesterolaemia (E78.0). The largest absolute effect size of 0.11 log odds per glycan PRS sd was observed for the effect of the PRS for the percentage of M9 glycan on the presence of aortocoronary bypass graft (Z95.1).

**Conclusion:** M9 N-glycan structure found to be associated with cardiovascular health phenotypes belongs to the high-mannose group of N-glycans and consists of the typical conserved chitobiose core with 9 mannose residues attached to it. Usually during N-glycan biosynthesis high mannose structures get converted to complex N-glycans. Therefore, high mannose structures on functional proteins are often regarded as products of incomplete N-glycan maturation in the ER and Golgi. Elevated levels of high mannose structures were previously observed in type 2 diabetes on urinary proteins [4], in the plasma of hypercholesterolemic patients [5] and in cancers [6, 7]. As a next step, we plan to elucidate the causal relationships in pairs of the statistically significantly associated plasma N-glycan traits and ICD-10 phenotypes using Two-sample MR.

**Acknowledgements:** The work of SZS and YSA was supported by a grant from the Russian Science Foundation (RSF) No. 19-15-00115.

## References

1. Juszczak A., Pavić T. et al. Plasma Fucosylated Glycans and C-Reactive Protein as Biomarkers of HNF1A-MODY in Young Adult-Onset Nonautoimmune Diabetes. *Diabetes Care*. 2019;42(1):17-26.
2. Callewaert N., Vlierberghe H.V. et al. Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics. *Nat Med*. 2004;10(4):429-434.
3. Sharapov S.Z., Shadrina A.S. et al. Replication of 15 loci involved in human plasma protein N-glycosylation in 4802 samples from four cohorts. *Glycobiology*. 2021;31(2):82-88.
4. Mise K., Imamura M. et al. Novel Urinary Glycan Biomarkers Predict Cardiovascular Events in Patients With Type 2 Diabetes: A Multicenter Prospective Study With 5-Year Follow Up (U-CARE Study 2). *Front Cardiovasc Med*. 2021;8:668059.
5. Bai L., Li Q. et al. Plasma High-Mannose and Complex/Hybrid N-Glycans Are Associated with Hypercholesterolemia in Humans and Rabbits. *PLoS One*. 2016;11(3):e0146982.
6. Bhat G., Hothpet V.R. et al. Shifted Golgi targeting of glycosyltransferases and  $\alpha$ -mannosidase IA from giantin to GM130-GRASP65 results in formation of high mannose N-glycans in aggressive prostate cancer cells. *Biochim Biophys Acta Gen Subj*. 2017;1861(11):2891-2901.
7. de Leoz M.L.A., Young L.J.T. et al. High-Mannose Glycans are Elevated during Breast Cancer Progression. *Mol Cell Proteomics*. 2011;10(1):M110.002717.

## New genes predisposing to coronary artery disease detected by genome-wide gene-based association analysis

Zorkoltseva I.<sup>1\*</sup>, Belonogova N.<sup>1</sup>, Shadrina A.<sup>1</sup>, Kirichenko A.<sup>1</sup>, Tsepilov Y.<sup>1,2</sup>, Axenovich T.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\*zor@bionet.nsc.ru

**Key words:** coronary artery disease, genome-wide gene-based association analysis, summary statistics

**Motivation and Aim:** Coronary artery disease (CAD) is one of the most common causes of death or disability worldwide, despite recent advances in its treatment and prevention. The etiology of CAD is complex and involves a genetic predisposition as well as various environmental factors. To date genome-wide association analyses (GWAS) have identified more than 300 single nucleotide polymorphisms at 163 genetic loci associated with CAD. However, there is no full understanding about the causal genes for CAD and the mechanisms of their action. We aimed to perform a detailed analysis of genes whose polymorphism may influence the risk of CAD.

**Materials and Methods:** Using the UK Biobank-based GWAS summary statistics (<https://www.leelabsg.org/resources>), we performed a three-stage gene-based association analyses of two CAD phenotypes, Ischaemic Heart Disease and Coronary atherosclerosis. We used four sets of genetic variants within a gene differing in their protein coding properties. We applied a combination of three methods, SKAT-O, PCA, and ACAT-V, implemented in sumFREGAT package (<https://cran.r-project.org/web/packages/sumFREGAT/index.html>) and used a ‘polygene pruning’ procedure to eliminate the influence of strong GWAS signals outside the gene.

**Results:** We found 123 genes significantly associated with CAD. Using the extended sample, we validated 63 of these genes. They contribute to CAD with their within-gene variants. Among 63 validated CAD-associated genes: 1) 26 genes did not contain any SNPs reaching the genome-wide significance threshold of  $5.0 \times 10^{-8}$  in GWAS; 2) The proportion of genes being closest to the top GWAS signals was 0.52; 3) 17 genes were identified using protein-coding SNPs sets. Ten of them included 25 potentially pathogenic variants, 4) 44 genes were identified using a set of non-coding within-gene variants. For 10 out of them we found evidence of pleiotropy with gene expression. Many of identified genes are well known. Some known CAD genes such as *FURIN* and *SORT1* did not show the gene-based association because their variants had low GWAS signals or gene-based association signals were inflated by the strong GWAS signal outside the gene. For several known CAD genes, we demonstrated that their effects can be explained not only or not at all by their own variants but by the variants within the neighboring genes controlling their expression. Three genes, *CDK19*, *NCALD* and *ARHGEF12* were not previously associated with CAD. **Conclusion:** We identified 63 genes that contribute to CAD with their within-gene variants, prioritized genes in known loci and identified three genes in the new CAD loci. The role of these genes should be clarified in further studies.

**Acknowledgements:** This work was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics, SB RAS (project 0259-2021-0009/AAAA-A17-117092070032-4) and the Russian Foundation for Basic Research (project 20-04-00464).

## 4.3 Section “Human origin and evolution”



## Genetic diseases and congenital genetic abnormalities in the context of ancient cultures research

Dobrovolskaya M.V.

*Institute of Archaeology, RAS, Moscow, Russia*

\* *mk\_pa@mail.ru*

**Key words:** ancient genomes, genetic diseases, cultural traditions

*Motivation and Aim:* Studies of ancient genomes have given rise to new perspectives for studying people of bygone eras. The greatest attention is currently paid to such issues as the history of formation of populations, the process of adaptation to food innovations, the coevolution of pathogens of infectious diseases and immunity in ancient populations. Meanwhile, there is a great variety of paleogenomic issues which need the interdisciplinary approach. One of them is the identification of individuals whose unusual behavioral traits and physical appearance were important in the context of the culture and beliefs of ancient societies. We can judge the attitude of the ancients towards "special" people from numerous data on dwarfs, people with an altered psyche, etc.

*Methods:* In analyzing the paleoanthropological materials, morphological and paleopathological manifestations of disorders can be revealed. However, similar morphological manifestations may have a different etiology. Finding this out and distinguishing between individual episodes and genetic disorders that could be repeated in a number of generations is very important.

*Hypotheses to test:* Here are two historical plots.

1. The myth of the Amazons. The frequency of *hyperostosis frontalis interna* – a pathology associated with a violation of the hormonal status – in burial mounds in the Middle Don region is more than 15 % (Buzhilova, Kozlovskaya, 2000). This poses a question whether this abnormality was caused only by an increased level of stress or there were a genetic predisposition?
2. The Enarei were fortune-tellers, the carriers of a sacred disease, passed down from generation to generation among the Scythian aristocracy. Presumably, the disease described by Herodotus and Hippocrates was caused by Forestier's disease, or rather, diffuse idiopathic hyperostosis, significantly associated with overweight and type 2 diabetes. Is there an inherited genetic predisposition to this pathology?

*Conclusion:* Testing these and other hypotheses using the methods of genome-wide genetic research is to allow us to interpret our hypotheses about the origins of a number of cult traditions that reflected the consciousness of humankind in the past.

*Acknowledgements:* This study is supported by a grant RFBR No. 20-29-01006.

### References

1. Buzhilova A.P., Kozlovskaya M.V. Were the Scythians Obese? (an anthropological analysis of cremated remains from a Scythian burial). In: Gulyaev V.I., Olkhovsky V.S. (Eds.). Scythians and Sarmatians in the 7th – 3rd centuries BC: paleoecology, anthropology and archeology. Moscow: IA RAS, 2000;36-38.



## Ancient DNA analysis of the North Caucasus populations from the III-XIII AD

Dzhaubermezov M.<sup>1,2\*</sup>, Gabidullina L.<sup>2</sup>, Ekomasova N.<sup>1,2</sup>, Atabiev B.<sup>3</sup>, Chagarov O.<sup>4</sup>, Korobov D.<sup>4</sup>, Khusnutdinova E.<sup>1,2</sup>

<sup>1</sup> *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre, RAS, Ufa, Russia*

<sup>2</sup> *Bashkir State University, Ufa, Russia*

<sup>3</sup> *Ltd. Institute for Caucasus Archaeology, Nalchik, Russia*

<sup>4</sup> *Institute of Archaeology, RAS, Moscow, Russia*

\* [murat-kbr@mail.ru](mailto:murat-kbr@mail.ru)

**Key words:** Ancient DNA, mtDNA, North Caucasus populations, Alans

*Motivation and Aim:* The earliest existing written evidence about Alans – people of Ponto-Caspian steppes and Caucasus region of I–XIII century AD – is found in the works of Chinese authors who mention the politonym "Yancai" in the "Records of the Grand Historian" and in the "History of the Former Han", and of European antique authors in the middle of the 1st century AD. Genetic data of the early Middle Ages Alans and their affinities to the Scythian-Sarmatian tribes, traditionally considered as their ancestors, as well as to the modern population of Europe and the Caucasus have not yet studied thoroughly, the whole genome data exists for only 6 individuals [1]. In this study we have analyzed ancient DNA of 27 individuals from 3 different archeological site belonging to the Alan culture from the III-XIII century AD, from the territory of the North Caucasus (Russian Federation).

*Methods and Algorithms:* DNA was extracted from the archaeological material (teeth, temporal bone) of 70 individuals. We produced low-coverage Illumina whole genome shotgun sequencing data for 63 individuals and will analyze these in a context of ancient and modern genetic variation of the region. The coverage of the genomes was 0.0035-0.2730X (median coverage 0.0638X and content of endogenous DNA 30.98 %).

*Results:* Radiocarbon analysis was carried out on 7 of the 27 samples studied. Radiocarbon dating was practically consistent with archaeological dating. The mtDNA palette of Alans consisting of 9 different haplogroups (F, K, T, U, D, H, X, HV, C) started to be similar to that of present day autochthonous North Caucasus populations.

*Conclusion:* The preliminary analysis allows us to assume close contacts of the Alans with the populations of the North Caucasus.

*Acknowledgements:* This study is supported by a grant Russian Science Foundation No. 22-24-00681 and the Ministry of Science and Higher Education of Russian Federation (No. 075-03-2021-193/5).

### References

1. Damgaard et al. *Nature*. 2018. doi: 10.1038/s41586-018-0094-2.

## Archaeology, anthropology and paleogenetics of the Fatyanovo culture – new results

Mustafin Kh.Kh.<sup>1</sup>, Alborova I.E.<sup>1\*</sup>, Engovatova A.V.<sup>2</sup>, Mednikova M.B.<sup>2</sup>,  
Chechyotkina O.Yu.<sup>1,2</sup>, Tarasova A.A.<sup>2</sup>

<sup>1</sup> *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

<sup>2</sup> *Institute of Archaeology, RAS, Moscow, Russia*

\* *ira\_teuchezh@mail.ru*

**Key words:** Bronze Age, Fatyanovo culture, craniology, paleopathology, paleogenetics, ancient *DNA*, *Y*-chromosomes, mitochondrial *DNA*, haplotypes, haplogroups, *STR*, *SNP*, *NGS*

*Goals:* The Corded Ware culture, one of the main archaeological traditions of Europe of the Late Neolithic, was widely distributed in the third millennium *BC*. The Fatyanovo archaeological culture, spread from Ilmen and Pskov lakes to Kama and Vyatka rivers and from the Vologda region to the Desna basin, can also be considered its derivative.

A long-standing comparative study of the craniological materials of the Upper Volga Fatyanovo culture included data of cultures of the battle axes and the Corded ceramics from the territories of Estonia, East Prussia, Saxony-Thuringia, the southwestern part of Germany, Bohemia, Moravia, Slovakia, and southeastern Poland. It was revealed the heterogeneity of the anthropological composition of representatives of this cultural and historical community [1]. According to the conclusions of this research, the Fatyanovo people of the Volga-Oka interfluvium formed a single group with the tribes of the Estonian battle axes culture and the Vistula-Neman culture in Poland.

A recent paleogenetic study included examination of samples from the Volga region and from Estonia. It was shown that the Fatyanovo people, being descendants of people of mixed origin (steppe dwellers and early European farmers [2]), were close to other carriers of the Corded Ware cultures, and were participants in a wide migration to the northeast from the Central Europe through the Baltics [3].

Within the framework of this work, the skeletal remains of 63 buried from the excavations of the Fatyanovo burial grounds (Volosovo-Danilovsky, Sushchevsky (Nikolo-Perevoz), Miloslavsky, Mytishchensky, Goretsky, Galuzinsky, Nikultsino, Voronkovo and Lovtsy) from the storage of the Institute of Archeology of the Russian Academy of Sciences were subjected to a comprehensive anthropological and genetic examination.

*Methods and Results:* Our anthropological examination of the materials of the Volosovo-Danilovsky burial ground, the largest monument of the Fatyanovo culture, investigated by D.A. Krainov in the Yaroslavl region, revealed rare and specific injuries of the left arm in the area of the elbow joint and hand in men. The closest analogies were previously described among representatives of the Corded Ware cultures in Germany (Eulau archaeological site) and in Poland (Kruszyn), where they were interpreted as military (defensive) wounds.

At the first stage of complex genetic studies of skeletal remains from the Fatyanovo burial grounds, samples (numbering more than 20) were selected, according to anthropological data, related to men. Ancient *DNA* was extracted from the samples, quantitative analysis of the quality of which in most cases showed high characteristics.

Microsatellite studies of *Y*-chromosomal *DNA* were carried out for all male samples – a fragment analysis was performed using a 27-locus *STR* panel. As a result, haplotypes were obtained, according to which haplogroups of the studied individuals were predicted. Then targeted *NGS* of *DNA* samples was carried out. A specially designed panel of probes targeting 700 *SNP* *Y*-chromosomes made it possible to check and refine the haplogroups of the studied individuals. A number of *SNPs* of nuclear *DNA*, on which a panel of probes was also designed, made it possible to identify phenotypic signs such as hair, eye, and skin color. Part of the probes of the developed panel completely covers mitochondrial *DNA*, therefore, as a result of *NGS*, mitochondrial haplogroups and haplotypes of the studied individuals were identified, and data were obtained that confirmed its antiquity by bioinformatic analysis of mitochondrial *DNA*.

Based on the results of genetic studies of the uniparental markers of a representative sample of individuals, new data on the possible origin of the population of the Fatyan culture were obtained.

*Conclusions:* The joint consideration of anthropological and genetic data gives evidence for hypothesis of migrations of the Central European inhabitants to the Russian Plain. Thus, injuries of the left arm in the area of the elbow joint and hand in the Volosovo-Danilovsky burial ground testify to similar battle tactics among the Fatyanovo people of the Upper Volga and the Corded Ware men of Germany and Poland. The features serve as an independent indicator of their belonging to a single cultural community. The paleopathological study finds confirmation in the results of paleogenetic analysis showing a connection with the Central European population of the Corded Ware culture.

*Acknowledgements:* The study was supported by the RFBR project No. 20-29-01002 "Migrations of the population of the Bronze Age in the forest belt on the Russian Plain according to paleogenetics and archeology".

### *References*

1. Denisova R.Ya. Anthropology of the ancient Balts. Riga: Znaniye, 1975. (in Russian)
2. Trifonov V.A., Prokhorchuk E.B., Zhur K.V. Genetic diversity of ancient populations in the Caucasus and the adjacent steppes during the eneolithic–bronze age (V–II mill. BC): main results and issues of cultural and historical interpretation. *Brif Commun Inst Archaeology*. 2021;262:95-114.
3. Saag L., Vasilyev S.V., Varul L., Kosorukova N.V., Gerasimov D.V. et al. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. *Sci Adv*. 2021;7:eabd6535.

# First ancient DNA analysis of mummies from the post-Scythian Oglakhty cemetery in South Siberia

Nedoluzhko A.<sup>1\*</sup>, Pankova S.<sup>1,2</sup>, Plotnikov N.<sup>3</sup>, Kim A.<sup>3</sup>, Shulpin M.<sup>3</sup>, Pogodina N.<sup>3</sup>, Nenasheva N.<sup>3</sup>, Rakitko A.<sup>3</sup>, Vergasova E.<sup>3</sup>, Poliakov A.<sup>4</sup>, Ilinskaya A.<sup>3</sup>, Ilinsky V.<sup>3</sup>

<sup>1</sup> European University at St. Petersburg, St. Petersburg, Russia

<sup>2</sup> The State Hermitage Museum, St. Petersburg, Russia

<sup>3</sup> Genotek Ltd., Moscow, Russia

<sup>4</sup> Institute for the History of Material Culture, RAS, St. Petersburg, Russia

\* nedoluzhko@gmail.com

**Key words:** Ancient DNA, Oglakhty cemetery, Tashtyk culture

*Motivation and Aim:* The Minusinsk Basin in Southern Siberia had unique conditions for the ancient human societies' development, because of its geographical location, favorable climatic conditions, and relative isolation. Located at the northern periphery of the eastern Eurasian steppe, surrounded by the Sayan Mountains this area witnessed numerous ancient human migrations with specific types of interaction between foreign and local groups.

The genomic diversity of the human population of Southern Siberia from the Chalcolithic to the middle Bronze Age is relatively well-studied due to the recent genome-wide studies, while the genetic ancestry of populations, represented by diverse archaeological cultures of the late Bronze and early Iron Age remains the white spot for modern paleogenomics [1–3].

Here, we present for the first time two ancient nuclear human genomes of the individuals buried in the Oglakhty cemetery (early Tashtyk culture, 2nd to 4th centuries AD). Our pilot study is undertaken within a multidisciplinary project on this distinguished site with well-preserved organic remains and provides new paleogenomic data of ancient societies of southern Siberia.

*Methods and Algorithms:* A dental drill was used for extremely careful sampling of bone powder from the proximal phalanx of the little finger of the left hand of the male mummy (KE9609), the middle phalanx of the little finger of the left hand of the female mummy (LC8544), and tooth (PI5222) and a metatarsal (QR6400) of a child from the Oglakhty grave 4 [4] in the extra-clean room DNA facilities of the Genotek Ltd. DNA-libraries were prepared from the ancient DNA extracts. Illumina HiSeq 2500 genome analyzer (Illumina, USA) with paired-end reads of 150 bp length was used for whole-genome sequencing (WGS). The number of well-established methods (haplogroup identification, PCA, ADMIXTURE, *f3*-statistics) were used for the bioinformatics analysis.

*Results:* Mitochondrial DNA and Y-chromosome haplogroup analysis of two adult individuals showed that they had a potential maternal kinship to each other and belonged to the same I4a1 subclade of mitochondrial haplogroup I. Previously, this subclade was only once observed in the Minusinsk Basin – in the sample of the Karasuk culture (RISE 493, Sabinka-2, grave 20) of the late Bronze Age. Y-chromosomal haplogroup was identified for the male individual (KE9609) as R1a1a subclade based on M738, CTS3230, M747 polymorphisms. R1a haplogroup was common for the Eurasian steppe's and Northern Caucasus' populations of the Bronze and early Iron Age [5, 6].

Based on genome-wide data we clearly showed that ancient Tashtyk specimens took place in close genetic proximity to the ancient individuals from the Siberian cultures.

*Conclusion:* We reconstructed genome-wide data from two – male (KE9609) and female (Ic8544) individuals from grave 4 whereas the third, child’s genome sequencing analysis proved unsuccessful. The autosomal DNA profiles of two Tashtyk individuals show their genetic connection with representatives of the preceded archeological cultures of the Minusinsk Basin.

### *References*

1. Allentoft M.E. et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015;522:167-172. doi: 10.1038/nature14507.
2. Haak W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015;522:207-211. doi: 10.1038/nature14317.
3. Damgaard P.B. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature*. 2018;557:369-374. doi: 10.1038/s41586-018-0094-2.
4. Pankova S.V. Mummies and mannequins from the Oglakhty cemetery in southern Siberia. Masters of the Steppe: the Impact of the Scythians and Later Nomad Societies of Eurasia. In: Proceedings of a Conference Held at the British Museum, 27-29 October. 2017;2020;373-396.
5. Wang C.C. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun*. 2019;10:590. doi: 10.1038/s41467-018-08220-8.
6. Boulygina E. et al. Mitochondrial and Y-chromosome diversity of the prehistoric Koban culture of the North Caucasus. *J Archaeol Sci-Rep*. 2020;31:102357. doi: 10.1016/j.jasrep.2020.102357.

## Reconstruction of genetic history in Southern Siberia by paleogenetic study of diachronic population samples

Pilipenko A.S.<sup>1\*</sup>, Molodin V.I.<sup>2</sup>, Trapezov R.O.<sup>1</sup>, Cherdantsev S.V.<sup>1</sup>, Pilipenko I.V.<sup>1</sup>,  
Tomilin M.A.<sup>1</sup>, Pozdnyakov D.V.<sup>2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Archaeology and Ethnography, SB RAS, Novosibirsk, Russia*

\* [alexpil@bionet.nsc.ru](mailto:alexpil@bionet.nsc.ru)

**Key words:** Ancient DNA, mtDNA, Y-chromosome, diachronic sample, Siberian populations

*Motivation and Aim:* Several areas in the south of Siberia, including the Altai, the Minusinsk Basin and the Baraba forest-steppe, have been well studied archaeologically. Archaeological cultures associated with populations chronologically covering the last few thousand years – from the Neolithic to the Late Middle Ages – are known for these regions. Representative collections of paleoanthropological material accumulated to date open up the possibility of a detailed study of the genetic history of Southern Siberia using paleogenetics methods, namely, the study of the genetic structure of populations within several local diachronic samples, identifying local features of the dynamics of the genetic composition of populations over time and comparing local models with each other to obtain a reconstruction of the main genetic events, influenced the formation of the genetic features of the indigenous population in this region and adjacent areas of Eurasia.

*Methods and Algorithms:* Representative diachronic samples of the ancient population of the Barabinsk forest-steppe, Mountain Altai, Steppe Altai, Minusinsk basin with a total number of samples of more than 2000 were formed. This unique sample bank is well characterized in terms of archaeological context and anthropological characteristics. Diachronic samples are subjected to sequential genetic analysis in several stages: 1. Analysis of the variability of mitochondrial DNA (mtDNA); 2. Determination of gender; 3. Analysis of the variability of the Y chromosome (for samples from men); 4. Analysis of individual autosomal markers (or small sets of markers); 5. Genome-wide methods of analysis. The combination of these methods makes it possible to gradually detail various aspects of the genetic history of the region in different chronological periods. *Results:* Data on the structure of more than 1,500 mtDNA samples and more than 300 Y-chromosome samples from our diachronic samples were obtained. These representative data allowed us to reconstruct the main stages of the formation of the population's gene pool in terms of maternally and paternally inherited genetic markers. We found correlations of the main changes in the structure of the gene pool with changes in the material culture of the population, including those caused by migration waves to the south of Siberia. Typical phenotypic features of their representatives (including features of pigmentation signs) have been determined for a number of populations. The study of large series of samples also allows us to study the local-territorial features of population processes in various areas of southern Siberia in detail.

*Conclusion:* Thus, the picture of the genetic history of the population in southern Siberia over the past few thousand years has been reconstructed. We continue to refine our reconstructions, including using genome-wide analysis methods.

*Acknowledgements:* This study is supported by grant from the Russian Science Foundation No. 17-78-20193 and grant from the Russian Foundation for Basic Research No. 20-29-01-024.

## Genetic history and culture of the oldest population of North Asia

Shunkov M.

*Institute of Archaeology and Ethnography, SB RAS, Novosibirsk, Russia*  
*shunkov77@gmail.com*

**Key words:** Ancient DNA, North Asia, Altai, Denisovan, Neanderthal

The oldest inhabitants of Northern Asia were most likely the late *Homo erectus*, who reached the foothill and low–mountain regions of Altai about 800 thousand years ago, where the most ancient Paleolithic objects were found – Karama and Ulalinka. Erectus came to the territory of Altai with the first migration wave from Africa, moving in a north-easterly direction through the Middle East, the Western Asian Highlands and the western regions of Central Asia. Penetration of erectus up to 52° north. It testifies to their significantly increased cognitive abilities, the development of effective adaptation strategies that allowed them to develop new areas in the northern part of Eurasia. The culture-bearing deposits of Karama indicate that representatives of the first migration wave with archaic pebble tools lived in Altai for almost the entire first half of the Middle Pleistocene. About 600 thousand years ago due to the significant cooling, early hominins apparently left this territory, migrating to regions with a more favorable climate, or died out, unable to adapt to the changed natural conditions.

The next stage of settlement of Northern Asia began about 300 thousand years ago with the appearance of the Denisovan in Altai, a carrier of Middle Paleolithic cultural traditions. The Denisovans brought to the newly developed territories more advanced methods of producing stone tools on deliberately shaped chips removed from well-prepared nuclei, i.e. techniques of parallel and Levallois splitting, as well as techniques of two-sided stone processing.

The formation of Denisovans, as well as Neanderthals, apparently occurred on the ancestral basis of the late *H. heidelbergensis* in the chronological interval of 400–200 thousand years ago. Sequencing of the nuclear genome of Denisovan showed that the separation of Denisovans and Neanderthals populations took place 430–380 thousand years ago [1]. The late *H. heidelbergensis*, settling from the Levant in the Iranian Highlands and further into Central and Northern Asia, gave rise to a new taxon – the Denisovan, the remains of which were first discovered in the Denisova Cave. The industry of Denisovans in the lower cultural layers of the Denisova Cave is technically and typologically close to the Ashelo-Yabrudien materials of the Levant.

Anthropological and paleogenetic data indicate that Denisovans lived in the cave for a long time, which is consistent with cultural continuity in the development of stone industries and suggests that Denisovans were an autochthonous population, with which the development of Middle Paleolithic and early Upper Paleolithic cultural traditions is associated. At the same time, remains of Altai Neanderthals were found in the Middle Paleolithic layers of Denisova Cave, but no vivid manifestations of Mousterian traditions were revealed, which is probably explained by the interbreeding of Denisovans and Neanderthals. Evidence in favor of this assumption is the bone of a hybrid girl found here, whose mother was a Neanderthal and her father was a Denisovan.

From the early stage of the Middle Paleolithic, the Denisovans developed an original stone industry with the periodic participation of Altai Neanderthals, which served as the basis for the formation of the Upper Paleolithic culture on an autochthonous basis in the interval of 50–40 thousand years ago with sets of stone and bone tools, as well as objects of symbolic activity, indicating a sufficiently high level of material and spiritual culture of the inhabitants of Altai at the beginning of the Upper Paleolithic. At the same time, the methods and techniques of Denisovans life activity were not inferior to the behavioral characteristics of early *H. sapiens* that appeared in the west of Siberia no later than 45 thousand years ago [2]. Denisovans, like Neanderthals, participated in the formation of a modern human species, transferring about 6 % of their genome to modern inhabitants of Australia, Oceania and the island part of Southeast Asia [3].

Although the remains of Denisovans have so far been found only in two caves (Denisova in the Altai and Baishiya Cave on the Tibetan Plateau), the level of their genetic diversity was higher than that of seven Neanderthals from different regions of Western and Central Europe, for whom complete mtDNA sequences were obtained, but lower than that of modern humans [4]. These data indicate that the population of Denisovans could be larger in number and more diverse than the population of Neanderthals, and also that Denisovans apparently had an exceptionally wide geographical distribution from Northern to tropical Asia [5].

*Acknowledgements:* This study is supported by a grant Russian Science Foundation No. 22-28-00049 and the Russian Foundation for Basic Research No. 20-29-01011.

### *References*

1. Meyer et al. *Nature*. 2014;505.
2. Fu et al. *Nature*. 2014;514.
3. Reich et al. *Am J Hum Genet*. 2011;89(4).
4. Sawyer et al. *PNAS*. 2015;112(51).
5. Prüfer et al. *Nature*. 2014;505.



# 5 Symposium “Biotechnologies: computational and experimental approaches”



- 5.1 Section “Biotechnology through the prism of microbiome” [489](#)
- 5.2 Section “Microbial communities of natural and anthropogenic habitats” [509](#)
- 5.3 Section “Industrial biotechnologies: creation of producer strains” [542](#)
- 5.4 Section “Modeling and computer analysis of microbiological systems and processes” [572](#)

# **Study of human and animal microbiome as genetic and pharmacological resources for the development of innovative biotechnologies for medicine, animal husbandry and agro-industrial complex**

Danilenko V.N.<sup>1\*</sup>, Yunes R.A.<sup>1</sup>, Ilyasov R.A.<sup>1</sup>, Kovtun A.S.<sup>1</sup>, Yanenko A.S.<sup>2</sup>, Kozlovsky Yu.E.<sup>2</sup>, Sidorenko A.V.<sup>3</sup>, Kolomiets E.I.<sup>3</sup>

<sup>1</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

<sup>2</sup> *National Research Centre "Kurchatov Institute", Moscow, Russia*

<sup>3</sup> *Institute of Microbiology of the National Academy of Sciences of Belarus, Minsk, Belarus*

\* *valerid@vigg.ru*

**Key words:** pharmacobiotics, antioxidant potential, probiotic potential, microbiota

Today, everyone is familiar with the stages of development of antibacterial drugs, the successes and problems of their use in medicine. Initially, the efforts of scientists around the world helped to create antibiotics based on natural producers of soil origin. However, the large-scale and sometimes irrational use of antibacterial drugs has contributed to the emergence and wide spread of pathogens resistant to most antibiotics. A similar situation occurred in the development of antiviral and immunomodulatory drugs. It became obvious that humans are losing the race to diseases and pathogens. Today, new approaches and paradigms are needed in the development technologies as well as in the search for new pharmaceutical substances. The advances of the past decade in the field of human microbiota have opened up many opportunities for us. Contemporary ideas about the structure and functions of the human intestinal microbiota and the conceptual integration of knowledge about commensal microorganisms, including probiotics obtained using molecular genetics, transcriptomic, proteomic, and metabolomic techniques, made it possible to forge the concept of "pharmabiotics". The creation of pharmabiotics based on microorganisms inhabiting various cavities of the human body, including the intestines, is a new and rapidly developing area of pharmacological science. The tasks set in this area can be solved if a number of fundamental tasks are integrated to ensure the creation of effective and safe pharmaceuticals and ingredients for them. The role of the microbiome and pharmabiotics as selective modulators of the human immune system, primarily cellular immunity, as well as the antioxidant system, is becoming increasingly clear. Over the past decade, Russia and the Republic of Belarus have been intensively studying the human intestinal microbiome and developing probiotics and pharmabiotics with neuromodulatory, immunomodulatory, antioxidant and pro-inflammatory potential.

## Technological aspects of the search, cloning and production of enzymes for industrial use

Goryachkovskaya T.<sup>1,2\*</sup>, Korzhuk A.<sup>1,2</sup>, Bochkov D.<sup>1,2</sup>, Romanov V.<sup>1,2</sup>, Zadorozhny A.<sup>1,2</sup>, Voskoboev M.<sup>1,2</sup>, Bogacheva N.<sup>1,2</sup>, Shlyakhtun V.<sup>1,2</sup>, Meshcheryakova I.<sup>1,2</sup>, Bannikova S.<sup>1,2</sup>, Antonov E.<sup>1,2</sup>, Shedko E.<sup>1,2</sup>, Sergeeva S.<sup>1,2</sup>, Bryanskaya A.<sup>1,2</sup>, Uvarova Y.<sup>1,2</sup>, Peltek S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [tanago@bionet.nsc.ru](mailto:tanago@bionet.nsc.ru)

**Key words:** thermostable enzymes, extremotolerant bacteria

*Motivation and Aim:* The search for enzyme genes and the production of enzyme preparations based on them is an urgent topic of our time: enzymes are in demand both in industry and in scientific research. Our approach is based, firstly, on the large-scale collection of microorganisms from extreme environmental conditions – thermal springs, salt and soda lakes, industrial landfills and others. Secondly, on the screening of both the physiological characteristics of individual microorganisms and the search for homologues of known enzymes in genomes by bioinformatics methods [1, 2].

*Methods and Algorithms:* By screening methods on Petri dishes, we determine the presence of enzymatic activity in a particular microorganism, by biochemistry methods we test preferred carbon sources for biomass production, by bioinformatics methods we investigate the presence of the desired gene in the genome of the microorganism. The next stage is the production of a sufficient amount of culture fluid with sufficient content of the target enzyme for analysis, isolation and chromatographic purification of the target enzyme, its mass spectrometric identification. The data obtained in this way allow us to assess the prospects of an enzyme preparation before the laborious process of cloning a gene in the yeast producer *Komagataella phaffii*.

*Results:* Among the natural strains, we have identified producers of secreted enzymes of various classes: amylases, proteases, cellulases, effectively producing enzymes to the culture medium at pH 9–10 at 55 °C. 32 cultures of amylolytics were obtained, the best activity is demonstrated by representatives of the species *Bacillus licheniformis*. In the genomes of this species there are two amylases, intracellular maltogenic amylase and alpha-amylase. pH and temperature optimums of the chromatographic purified alpha-amylase 47018 were characterized. The adapted to expression in yeast *K. phaffii* enzyme gene sequence was chemically synthesized.

The search for thermotolerant destructors of protein (protease producer), cellulose (cellulase producers) and olive oil (lipase producers) was carried out at different pH values (7, 8, 9) and a temperature of 50 °C. 20 strains have been isolated that can use olive oil as the only source of carbon and energy. Three thermotolerant strains of *Brevibacillus thermoruber* and a strain of *Thermomonas* with secreted cellulases have been isolated. Isolated strain 6008 (*Bacillus* sp.) had high secreted proteolytic activity at 50 °C. 31 protease genes were found in the genome of strain 6008.

*Conclusion:* Among the natural strains, we identified producers of secreted enzymes of various classes (amylases, proteases, cellulases) and obtained chromatographically pure enzyme preparations. The developed preparations were studied according to biochemical parameters – operating temperature ranges and pH, resistance to metal ions. The genomes of the isolated producers were sequenced. Enzymes of the corresponding classes were found in the genomes of the strains.

*Acknowledgements:* This research was funded by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2019-1662 (Kurchatov Genomic Center, federal research center ICG SB RAS, Novosibirsk).

The study was funded by the Ministry of Science and Higher Education of the Russian Federation as a part of the comprehensive project on high-tech industry “Creation of high-tech production of high-quality plant food proteins” (The agreement on the provision of subsidies from the federal budget for the development of cooperation of a state scientific institution and organization the real sector of the economy in order to implement a comprehensive project for the creation of high-tech industry No. 075-11-2020-036 from 15.12.2020) in the framework of the Decree of the Government of the Russian Federation of April 9, 2010 No. 218 on the basis of the ICG SB RAS.

We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education grant No. FWNR-2022-0022) at the Federal Research Center ICG SB RAS.

### *References*

1. Beklemishev A.B., Pykhtina M.B., Kulikov Y.M., Goryachkovskaya T.N., Bochkov D.V., Sergeeva S.V., Vasileva A.R., Romanov V.P., Novikova D.S., Peltek S.E. Creation of a recombinant *Komagataella phaffii* strain, a producer of proteinase K from *Tritirachium album*. *Vavilov J Genet Breed.* 2021;25(8):882-888.
2. Bryanskaya A.V., Shipova A.A., Rozanov A.S., Volkova O.A., Lazareva E.V., Uvarova Y.E., Goryachkovskaya T.N., Peltek S.E. Metagenomics dataset used to characterize microbiome in water and sediments of the lake Solenoe (Novosibirsk region, Russia). *Data Brief.* 2020;34:106709.

## Metagenomic studies of the honey bee gut microbiome: prospects for the development of pharmabiotics

Ilyasov R.A.<sup>1,2\*</sup>, Guzenko E.V.<sup>3</sup>, Danilenko V.N.<sup>1</sup>, Kilchevsky A.V.<sup>3</sup>

<sup>1</sup> Vavilov Institute of General Genetics, RAS, Moscow, Russia

<sup>2</sup> Bashkir State Agrarian University, Ufa, Russia

<sup>3</sup> Institute of Genetics and Cytology, National Academy of Sciences of Belarus, Minsk, Belarus

\* [apismell@hotmail.com](mailto:apismell@hotmail.com)

**Key words:** microbiome, honey bee, lactobacteria, *Apis mellifera*, *Lactobacillus*, pharmabiotics, metagenomic studies

The honey bee *Apis mellifera* is a social insect species with a unique complex of adaptations to adverse environmental conditions. The unique characteristics of the honey bee colony are the result of the symbiotic relationship of the honey bee organism with the gut microbiome through chemical signals. The gut microbiome is an additional organ of the honey bee, which consists of a unique small anaerobic community of gram-negative (genera *Apibacter*, *Bartonella*, *Commensalibacter*, *Gilliamella*, *Parasaccharibacter*, *Saccharibacter*, *Serratia*, *Snodgrassella*, *Frischella*) and gram-positive (genera *Lactococcus*, *Lactobacillus*, *Bifidobacterium*) species of bacteria. These types of intestinal bacteria have undergone a long co-evolution with bees and have accumulated unique genes in their genomes that provide immunity, adaptation, and detoxification of xenobiotics and pesticides. The gut microbiome of honey bees contributes to the acceleration of growth, an increase in longevity, an increase in the absorption of food and an increase in the body weight of individuals. The microbiome is involved in the growth, development and reproduction of honey bees and contributes significantly to their metabolism. The microorganisms in the gut of honey bees synthesize biologically active substances, increase the efficiency of digestion and help to assimilate nutrients. The microbiome modulates the behavior of honey bees, sensing, memory and learning by influencing the brain and nervous system. The genome of the honey bee and the metagenome of the gut microbiome are combined into a single, functionally related complex genome, which should not be considered separately. Some strains of lactobacilli from the honey bee gut microbiome may be potential pharmabiotics for humans. Lactobacilli of the honey bee microbiome are believed to be the host's basic line of defense against pathogens using antimicrobial peptides and inhibitors [1, 2].

*Acknowledgements:* The study is supported by the grant of the Russian Foundation for Basic Research (RFBR) (No. 19-54-70002).

### References

1. Ilyasov R.A., Gaifullina L.R., Saltykova E.S., Poskryakov A.V., Nikolenko A.G. *J Apicultural Sci.* 2012;56(1):115-124. doi: 10.2478/v10289-012-0013-y.
2. Danilenko V.N., Devyatkin A.V., Marsova M.V., Shibilova M.U., Ilyasov R.A., Shmyrev V.I. *J Inflamm Res.* 2021;14:6349-6381. doi: 10.2147/JIR.S333887.

## Fundamental and applied aspects of microbiome genetics

Kilchevsky A.V., Mikhalenka A.P., Andreeva I.N., Mazur A.Ch., Shulinski R.S.

*Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus*

\* [Kilchev@presidium.bas-net.by](mailto:Kilchev@presidium.bas-net.by)

**Key words:** aging, gut microbiome, metagenomic analysis

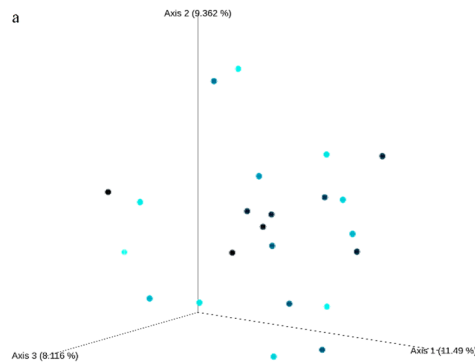
*Motivation and Aim:* Over the past twenty years, the scientific community has actively been involved in the study of microbiota characteristics of various organisms. The microbiome is involved in paramount ecosystem processes, contributing both to host metabolism on a local scale and biogeochemical nutrient cycling globally. It was recognized that the microbial community is a key factor in determining the physiological state of the host organism. In turn, the genetic features of the host affect the microbial community, which confirms the hypothesis that the microbiome depends on the genetics of the host [1]. Cognizance of interaction mechanisms between the genome and the microbiome will make it possible to use this knowledge for the practical application both in medicine and other branches. Investigation of the interaction between the genome and the microbiome has become an overriding area of scientific interest of the Republican Centre for Microbiome Study, the Institute of Genetics and Cytology, NAS of Belarus. First of all, such a study aims to investigate the human microbiome in normal and pathological conditions, its footprint on the implementation of certain molecular genetic characteristics of the human genome, and its role in maintaining health and the pathogenesis of diseases.

*Methods and Algorithms:* Genetic testing of patients' biological material was performed by metagenomic analysis using the Illumina high-throughput sequencing technology. After the total DNA isolation and preparation of libraries for sequencing, V3-V4 variable regions of the 16S rRNA gene were sequenced using the MiSeq Reagent Kit v3 (600 cycles) and MiSeq Illumina (USA) according to the manufacturer's recommendations. At the initial stage, the data obtained were analyzed and visualized using the 16S Metagenomics program (Version 1.1.0) provided by Illumina as part of the BaseSpace package (<https://basespace.illumina.com>). Evaluation of the quality of reads and trimming of low-quality bases were carried out using FastQC and Trimmomatic programs respectively. The V3-V4 regions were analyzed using the qiime2 program. When analyzing the data, the main parameters of the patients were considered and downloaded as a file with metadata in the TVS format. The final pipeline version is available at [https://github.com/IGC-bioinf/metagenome\\_scripts/blob/main/qiime2.sh](https://github.com/IGC-bioinf/metagenome_scripts/blob/main/qiime2.sh).

*Results:* In recent years, a growing body of research papers has come to light indicating that the dynamics of gut microbiome modification is associated with many age-related changes and influences longevity in various species [1]. In this regard, it is of considerable interest to study the characteristics of the human intestinal microbiota, which plays an important role in the formation of healthy aging and active longevity.

A survey of 23 people (an elderly group (mean age  $-81 \pm 0.2$ )) was carried out. The most represented components in the studied microbiome samples included *Ruminococcaceae* and *Verrucomicrobiaceae* representatives (Table 1). In some samples, representatives of the same family (*Ruminococcaceae*, *Verrucomicrobiaceae*) consisted of more than 50 % of the detected reads. Hierarchical clustering of samples identified a cluster that included the samples with the dominant bacterium species of the *Akkermansia* genus

(from 5.89 to 52.86 %). When decreasing the dimension of the Bray-Curtis dissimilarity, a tendency for clustering the samples of people over 82 was revealed, and a high degree of divergence among the remaining samples was noted (Fig. 1). The results obtained indicate a more homogeneous composition of the microbiome in people over 80.



**Fig. 1.** Clustering of samples with colouring according to the patient's age. A darker tone of the dot indicates a higher age value

In addition to the fundamental research, metagenomic analysis in practical medicine is getting more and more widespread. Work on the identification of pathogenic and potentially pathogenic microorganisms in children with chronic inflammatory bowel diseases (CIBD) is underway at the Republican Centre for Microbiome Study. A combination of CIBD and other bacterial and viral bowel infections may significantly change a clinical pattern, characteristic of each infection separately, which complicates disease diagnosing and choosing a therapy [2]. Metagenome sequencing of DNA from colon biopsy material makes it possible to determine the totality of the microbiome community, which cannot be detected by classical methods currently used in medicine. We present a clinical case of a 2-year-old female patient with a provisional diagnosis of "Non-infectious gastroenteritis and colitis, unspecified. A risk of CIBD development". *Escherichia marmotae*. *E. marmotae* was identified using 16S rRNA sequencing of colon biopsy material, a fairly recently described pathogen phenotypically indistinguishable from *E. coli* isolated only with the advent of new metagenomic technologies [3]. The results obtained allowed us to correctly adjust a therapy regimen for the patient.

**Conclusion:** The Republican Centre for Microbiome Study, the Institute of Genetics and Cytology, NAS of Belarus, is a multifaceted Centre, which together with the organizations concerned of the Ministry of Health and the National Academy of Sciences allows addressing fundamental and applied problems of various branches.

### References

1. Santoro A. et al. Gut microbiota changes in the extreme decades of human life: A focus on centenarians. *Cell Mol Life Sci.* 2018;75:129-148.
2. Vindigni S.M. et al. The intestinal microbiome, barrier function, and immune system in inflammatory bowel disease: A tripartite pathophysiological circuit with implications for new therapeutic directions. *Therapeutic Adv Gastroenterol.* 2016;9:606-625.
3. Denamur E. et al. The population genetics of pathogenic *Escherichia Coli* *Nat Rev Microbiol.* 2021;19(1):37-54.

## Bioinformatics approaches and methods for analysis of gut microbiota

Kovtun A.S.<sup>1,2\*</sup>, Danilenko V.N.<sup>1</sup>

<sup>1</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

<sup>2</sup> *Skolkovo Institute of Science and Technology, Moscow, Russia*

\* *kovtunas25@gmail.com*

**Key words:** microbiota, metagenome, metagenomic signature

Active development of methods for the analysis of microbial communities called microbiota became possible with introduction of the next generation sequencing (NGS) platforms. Nowadays, growing number of scientific groups worldwide contribute to the knowledge about various microbial communities, such as those inhabiting soils, wastewater and natural waters, human organism, including gut microbiota, etc. Generally, the bioinformatics analysis of microbiota can be divided into taxonomic identification and functional annotation of metagenome (the overall set of genetic information contained in microbiota).

The key goal of the taxonomic analysis is to describe, which organisms constitute the microbiota and what their abundance is. This kind of analysis can be conducted with two types of data: 16S rRNA and whole metagenome. Each of them has its advantages and disadvantages and requires specific algorithms. The 16S rRNA data can give the information about taxonomy of a microbial sample and is still widely used, thanks to its cost-effectiveness and relatively modest requirements for computational resources. Although the genes encoding the 16S rRNA are good taxonomic classifiers in theory, the short length of read sequences obtained from the NGS imposes limitations on the resolution of the method – the taxonomic abundance can be well-resolved mostly on the ‘Genus’ level [1]. Another issue is the generation of chimeric reads. The actively developing single-cell sequencing techniques, which allow the whole-length sequencing of the 16S rRNA gene, partly solve the resolution problem with exchange on still significantly lower quality of the data [1]. Despite these difficulties, the number of publications in metagenomics using the single-cell sequencing is growing and the corresponding algorithms are being developed and improved constantly. The whole metagenomic data gives better resolution and allows the researchers to analyze the taxonomy up to the ‘Strain’ level, as the taxonomy is often assigned by implementation of the whole genomes of microorganisms. However, such approach requires deeper sequencing and, thus, is way costlier, and more resource-demanding. It is also more prone for contamination with eukaryotic DNA.

Unlike the taxonomy analysis, the functional annotation of a metagenome should be conducted on the whole metagenome data. One can either conduct the assembly recreate (usually partially) the genomes, or simply align the reads against the reference database of genes and metabolic pathways. The first approach requires significantly more resources, can generate chimeric assemblies and collapses the information on relative abundances of each gene, thus, the further mapping of reads is still required. But it allows a more straightforward annotation of the genes, as they are being recreated during the assembly. The second method shows more difficulties in correct normalization, since



not only depth, but also breadth and uniformity of coverage of the reference genes should be accounted for to correctly tell, if the gene really exists in the metagenome.

The key disadvantage of the traditional metagenomic analysis is that the taxonomic and the functional results are reported separately. The unification of the information on the identified genes, their abundance and bacterial source can be implemented through metagenomic signatures. One of existing algorithms the metagenomic isoforms' recreation has been introduced by us in Kovtun et al., 2018 [2]. It has been successfully used for description of gut microbiota of healthy adults [2], its development in children of different age [3], and also for the analysis of changes in gut microbiota in children during autism spectrum disorder [4]. Currently, another research is being conducted by us, which uses the similar approach to identify the changes in gut microbiota of people with depression from Moscow region. The preliminary results demonstrate statistically significant decrease of abundance of genes involved in synthesis of butyric acid, spermidine and estradiol in *Faecalibacterium prausnitzii* and *Roseburia intestinalis*, which also correlates with the severity of the disease.

### References

1. Johnson J.S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun.* 2019;10(1):5029.
2. Kovtun A.S. et al. *In silico* identification of metagenomic signature describing neurometabolic potential of normal human gut microbiota. *Russ J Genet.* 2018;54(9):1101-1110.
3. Averina O.V. et al. Altered neurometabolic potential of gut microbiome in healthy children of different age. *Bulletin of RSMU.* 2020;6:68-79.
4. Averina O.V. et al. The bacterial neurometabolic signature of the gut microbiota of young children with autism spectrum disorders. *J Med Microbiol.* 2020;69(4):558-571.

# Synthetic community analysis by the trait-based method of quantitative assessment of ecological functional groups

Kropochev A.<sup>1,2\*</sup>, Lashin S.<sup>1,2,3</sup>, Klimenko A.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [kropochev@bionet.nsc.ru](mailto:kropochev@bionet.nsc.ru)

**Key words:** trait-based approach, functional groups, microbial ecology, human gut microbiome

*Motivation and Aim:* The human gut microbiome is a complex, ever-changing ecosystem of trillions of bacteria. It consists of a large number of species, some of which are not well characterized. In addition, for bacterial communities, horizontal gene transfer events are not rare. Consequently, it can be difficult for researchers to predict for certain taxonomic groups their functions in the community.

We believe that an attempt to reveal the ecological relationships in the microbial community by shifting the focus from taxonomy to ecological function could be very promising. We have developed the method that allows us to conduct a functional analysis of the community implementing a trait-based approach. Using this method to analyze a synthetic community – a community where the number of species is strictly fixed – helps to uncover the problem of the relationship between taxonomy and function. The developed method was also verified using independent taxonomic-based data.

*Methods and Algorithms:* In our work, we use publicly available data on synthetic communities of the human intestine from the study of the effect of diet on the stability of the mucin layer [1]. The community functional analysis was carried out by the trait-based method of quantitative assessment of ecological functional groups.

*Results:* Verification showed the concordance of the patterns of changes in the abundance of the functional groups between the estimates derived by the developed method and independent taxonomic-based estimates. Comparison of the results with another widespread functional analysis method – Human N 3 – demonstrated the applicability of the developed method and its analytical power for investigating such processes as acetogenesis, butyrate production, sulfate reduction and mucin decomposition in human gut.

*Acknowledgements:* The study was supported by Russian presidential grant for support of young scientists No. MK-3363.2022.1.4.

## References

1. Desai M.S., Seekatz A.M., Koropatkin N.M. et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*. 2016;167:1339-1353.e21.

## Metatranscriptomic analysis of the stomach antrum vs corpus biopsy microbiota in patients with *Helicobacter pylori* infection

Kupriyanova E. \*, Abdulkhakov S., Markelova M., Grigoryeva T.

*Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

\*[fewrandomletters@gmail.com](mailto:fewrandomletters@gmail.com)

**Introduction:** Gastric, non-*H. pylori* microbiota is supposed to play an important role in progression of inflammation and pathogenesis of upper GI disorders. However, to this date the issue remains poorly understood. The aim of this study was to characterize the microbiota of the gastric mucosa in *H. pylori*-positive and *H. pylori*-negative patients in biopsy samples from the antrum and corpus of the stomach.

**Methods:** Seventy two mucosal biopsy samples were obtained from 36 patients – one sample from the antrum and one from the corpus of the stomach respectively. According to the rapid urease test combined with cytology results, 13 *H. pylori*-positive and 23 *H. pylori*-negative patients were detected. Total RNA from each sample were extracted with TRIZOL reagent. Fraction of RNA was then used to synthesize cDNA. Amplicon libraries of V3-V4 variable region of the 16S rRNA gene were sequenced on the Illumina MiSeq platform.

**Results:** There is a significant difference in bacterial composition between *H. pylori*-positive and *H. pylori*-negative samples. *H. pylori*-positive patients showed a dominance of the *Helicobacter* genus up to 90 %. Also, there is a noticeable difference between the antrum and the corpus of the stomach, which is especially noteworthy for *H. pylori*-positive patients. Apart from *H. pylori*, the predominance of *Streptococcus*, *Prevotella* and *Enterobacteriaceae* was observed as well as the presence of the *Rothia*, *Gemellaceae*, *Fusobacterium* and *Neisseria* genera, which are usually recognized as the normal gastric microbiota.

**Conclusion:** *H. pylori* tends to prevail in the stomach of *H. pylori*-positive patients, especially when the body of the stomach is analyzed compared to the antrum. The use of a biopsy from the corpus of the stomach is likely to have a greater predictive value when diagnosing *H. pylori* infection. Using cDNA for 16S rRNA gene libraries preparation allows detecting a greater variety of bacteria.

**Acknowledgements:** This research was funded by the subsidy allocated to Kazan Federal University for the assignment in the sphere of scientific activities (project No. 0671-2020-0058).

## Omix technologies and bioinformatics: an example of use in creating a pharmabiotic based on the *Limosilactobacillus fermentum* U21 strain

Poluektova E.U.<sup>1\*</sup>, Mavletova D.A.<sup>1</sup>, Odorskaya M.V.<sup>1</sup>, Marsova M.V.<sup>1</sup>, Klimina K.M.<sup>1,2</sup>, Koshenko T.A.<sup>1</sup>, Yunes R.A.<sup>1</sup>, Vatlin A.A.<sup>1\*</sup>, Danilenko V.N.<sup>1</sup>

<sup>1</sup> Vavilov Institute of General Genetics, RAS, Moscow, Russia

<sup>2</sup> Federal Research and Clinical Center for Physical and Chemical Medicine of the Federal Medical and Biological Agency of Russia, Moscow, Russia

\*epolu@yigg.ru vatlin; alexey123@mail.ru

**Key words:** Pharmabiotics, antioxidants, omics technologies, lactobacilli

**Motivation and Aim:** Today's world is undergoing revolutionary changes in the development and use of pharmacological preparations based on bacteria and their biologically active ingredients. Such preparations are increasingly referred to as pharmabiotics, as opposed to probiotics, used primarily as dietary supplements and consumed by healthy people. Pharmabiotics are live biotherapeutic drugs and/or their metabolites and components with established pharmacological ingredients, mechanism of action and intended for treating specific diseases. In order to develop pharmabiotics, aside from traditional microbiological and biotechnological approaches, a complex of omics technologies can substantially facilitate the process. Lactobacilli are the important component of the human microbiota and, due to the active synthesis of biologically active compounds and bidirectional communication with the host organism, they can affect the state and antioxidant (AO) status of the macroorganism. The strain of *Limosilactobacillus fermentum* U-21 showed high AO activity on *in vitro* and *in vivo* models [1, 2]. In this work, we used a combination of genomic, transcriptomic, and proteomic technologies to identify genes and proteins that potentially determine the unique AO properties of the strain.

**Methods and Algorithms:** The DNA of the *L. fermentum* U21 strain and the comparison strains *L. fermentum* 103 and *L. fermentum* 279 was sequenced and deposited in the NCBI GenBank (WGS PNBB01, PGGI01, PGGE01). The search for genes exhibiting AO properties in the genome of *L. fermentum* U-21 was carried out using the reference catalog of genes of antioxidant proteins found in various species and strains of lactobacilli [3; <https://github.com/Alexey-Kovtun/Catalog>] and the developed algorithm for their search [4]. RNA was isolated on a King Fisher automated station with the MagMAX™ mirVana™ Total RNA Isolation Kit (Thermo Fisher Scientific). Ready libraries were sequenced on an Illumina HiSeq 2500. Kallisto v0.46.0 software was used to map reads and evaluate transcript abundance. Differential expression analysis was performed using the edgeR v3.26.8 package integrated into the Degust v4.1.1 web tool. The proteins of the cell-free culture supernatants of *L. fermentum* strains were separated in SDS-PAGE. For mass spectrometric analysis, the colored protein bands were excised and the peptides were separated using the Ultimate 3000 Nano LC system connected to a Q Exactive HF mass spectrometer through a nanoelectrospray source (Thermo Fischer Scientific) based on the mass spectrometry group of the Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS. LC-MS/MS data analysis was performed using PEAKS Studio 8.0 build 2016-0908 software. The primary structures of the generated peptides were analyzed based on the UNIPROT KB protein sequence database (07.2016).

**Results:** Genomic analysis of the strain made it possible to identify 29 genes, the products of which are likely to exhibit antioxidant properties. The most important genes are those encoding proteins of the thioredoxin complex and those encoding the metabolism and transport of heavy metals. Hydrogen peroxide was chosen as a stress inducer. We used a peroxide concentration that did not affect the viability of the cells of the strain in a given period of time (10 mM, 30 min). 380 genes showed an increase in expression and 370 genes showed a more than twofold decrease in expression upon exposure to hydrogen peroxide. The greatest increase in expression was observed in genes involved in the production of ammonium as a source of nitrogen (14–24 times), various oxidoreductases, catabolism of disaccharides, as well as enzymes that protect proteins and nucleic acids from oxidation and stress proteins common to different types of stress. The expression of genes encoding mobilome elements, IS elements, type II toxin-antitoxin systems also increased. The expression of genes encoding subunits of ATP-binding cassette transporters (ABC-transporters), the phosphotransferase system for sugar transport and fatty acid biosynthesis was sharply reduced. Among the antioxidant protein genes, the greatest increase in gene expression was noted for the operon containing the genes encoding thioredoxin reductase and cysteine synthase and the operon containing the genes for the P-type ATPase heavy metal transporter and the protein containing the domain associated with heavy metals. Proteomic analysis of the exoproteome of the strain enabled the identification of the ClpB protein belonging to a chaperone complex, which can play a key role in refolding the misfolded proteins as a result of oxidative stress in various tissues and organs of the animal body.

**Conclusion:** The use of a complex of omics technologies to characterize the therapeutic properties and mechanism of action of the *L. fermentum* U21 strain is one of the first attempts in this field of research. The data of genomic and transcriptomic analysis indicate that the main defense against oxidative stress in this strain is the thioredoxin system, which functions through the regulation of the dithiol/disulfide balance. The process of metal chelation also makes a significant contribution to the antioxidant properties of the strain. The results of transcriptomic analysis of a strain of lactobacilli under conditions of oxidative stress, as well as the data of genomic and proteomic analysis, make it possible to significantly expand the range of bacterial enzymes involved in the response to stress. These data can be used to select probiotic strains with antioxidant activity, as well as to develop additional components in probiotic preparations aimed at treating diseases associated with oxidative stress, primarily inflammatory diseases.

**Acknowledgements:** The study is supported by State task No. 0092-2022-003.

## 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. Danilenko V.N., Stavrovskaya A.V., Voronkov D.N., Gushchina A.S., Marsova M.V., Yamshchikova N.G., Ol'shansky A.S., Ivanov M.V., Illarioshkin S.N. The use of a pharmabiotic based on the *Lactobacillus fermentum* U-21 strain to modulate the neurodegenerative process in an experimental model of Parkinson's disease. *Ann Clin Exp Neurol.* 2020;14(1):62-69. doi: 10.25692/ACEN.2020.1.7
3. 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
4. Kovtun A.S., Averina O.V., Zakharevich N.V. et al. *In silico* identification of metagenomic signature describing neurometabolic potential of normal human gut microbiota. *Russ J Genet.* 2018;54(9):1101-1110. doi: 10.1134/S1022795418090089.

# Investigating bacterial and yeast diversity of spontaneous fermentation beer and cider using Hi-C metagenomics

Sonets I.<sup>1\*</sup>, Ivanova V.<sup>1</sup>, Vasiliev P.<sup>1,3</sup>, Ulianov S.<sup>1,2</sup>, Solovyev M.<sup>2</sup>, Tyakht A.<sup>1</sup>

<sup>1</sup> *Institute of Gene Biology, RAS, Moscow, Russia*

<sup>2</sup> *Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *Research Centre for Medical Genetics, Moscow, Russia*

\* *ignatsonets@gmail.com*

**Key words:** Hi-C metagenomics, microbiome, comparative genomics, chromosome conformation capture, foodomics, yeast, spontaneous beer, probiotics, plasmids

*Motivation and Aim:* Microbiome of fermented foods and beverages is a contributor to their organoleptic properties and health-related effects. Health claims must be based on scientific evidence supported by robust and detailed omics-based analysis of food products. The classic method for investigating taxonomic and functional profile is high-throughput shotgun sequencing, but metagenome-assembled genomes (MAGs) yielded from pure WGS data are often fragmentary and consist of a set of unordered scaffolds. Augmentation of metagenomics with chromosome conformation capture methods (like Hi-C/3-C) provides more accurate profiles of microbiome structure and functions. Combined investigation of food community structure along with the functional potential of each of its members will allow better understanding of microbial ecology and contribute to the improved quality control and innovative formulation development of beneficial products including low-alcohol/alcohol-free beverages. Apart from improving genome assembly, usage of Hi-C methods and its modifications allow revealing the chromatin structure and genome structural organization (i.e. number of chromosomes, folding properties, localisation in nucleus etc.).

*Methods and Algorithms:* We present a Hi-C metagenomics-based pipeline for exploring food microbial communities – with a particular focus on beverages yeast identification and comparative genomics. This pipeline includes Hi-C metagenomic library preparation and sequencing; reads preprocessing, Hi-C-aided reconstruction of metagenome-assembled-genomes (MAGs) from sequencing data and taxonomic profiling, and yeast comparative genomic analysis as a separate module. Yeast comparative genomics module includes the following steps: *de novo* gene annotation, comparative genomics with published reference genomes, phylogenetic analysis, average nucleotide identity analysis and functional enrichment analysis. In our pilot study, we applied Hi-C and common WGS metagenomic library preparation and sequencing for each of the 12 spontaneous fermentation beverages based on a larger collection analyzed using high-throughput 16S rRNA and ITS sequencing (<https://biota.knomics.ru/beer-cider-2020>) [1, 2]. We recovered multiple microbial genomes mainly represented by the prokaryotes from the *Acetobacteraceae*, *Enterobacteriaceae* and *Lactobacillaceae* families. Among the yeasts, the major species detected were *Brettanomyces/Dekkera bruxellensis* (a non-conventional yeast considered a spoilage organism in products like wine but an essential component in beers like lambics and American wild ales) and *Saccharomyces cerevisiae* (well-known yeast

prevalent in foods and beverages). We focused on recovered MAGs (bacteria and yeasts) as a showcase for validating the developed pipeline.

*Results:* Up to 6 bacterial Hi-C MAGs per sample was recovered; yeast MAGs were found in 5 samples. Recovered MAGs tended to have high completion ( $\geq 80\%$ ) and low contamination ( $\leq 5\%$ ). Features of phylogeny and gene content of the identified *Lactobacilli* genomes reflected their ability to survive in sour beers (like presence of hop resistance genes). For one *Lacticaseibacillus casei* MAG recovered from a mixed-fermentation ale sample, we found enrichment of genes possibly reflecting its high probiotic potential. As a showcase of Hi-C metagenomic aid in linking plasmids to bacterial hosts, we explored *Pediococcus damnosus* plasmidome. By utilizing Hi-C reads to improve genome assembly of *B.bruxellensis* MAGs, we achieved significant improvement in comparison with WGS-only genome assembly (up to 95 % completion and 4 % contamination). The obtained chromosome contact maps demonstrated chromosome conformation structure similar to *S. cerevisiae* Rabl-like conformation. Yeast comparative genomic analysis clearly demonstrated clustering by niche, and functional enrichment analysis showed prevalence of specific metabolic modules which are likely responsible for adaptation to niche conditions such as SO<sub>2</sub> tolerance (applied in winemaking as a preservative agent) or reoxidation of pyruvate (thus outperforming *S. cerevisiae* in beer production).

*Conclusion:* Hi-C metagenomics is a promising technique for in-depth microbiome analysis of spontaneously fermented foods and beverages. It allows improved reconstruction of bacterial and yeast genomes compared to conventional metagenomics, thus facilitating downstream analysis steps like comparative genomics. In the future, the approach might aid in developing advanced food quality control as well as in discovery of novel strains with high biotechnological potential.

*Acknowledgements:* This work was supported by the Russian Science Foundation (grant 19-74-10092).

## References

1. Tyakht A., Kopeliovich A., Klimenko N., Efimova D., Dovidchenko N., Odintsova V. et al. Characteristics of bacterial and yeast microbiomes in spontaneous and mixed-fermentation beer and cider. *Food Microbiol.* 2021;94:103658.
2. Efimova D., Tyakht A., Popenko, A. et al. Knomics-Biota - a system for exploratory analysis of human gut microbiota data. *BioData Mining.* 2018;11:25.

## Commensal microbiota: realities and vaccine making potential

Suvorov A.N.

*Institute of Experimental Medicine, St. Petersburg, Russia*

*Alexander\_suvorov1@hotmail.com*

**Key words:** Microbiota, immunity, vaccine, mucosa

Intestinal microbiota in terms of diversity and quantitative content of microorganisms (bacteria, fungi, protozoa and viruses) significantly exceeds all other loci of microbial localization in the human body. This is due both to the peculiarities of the intestinal tube and to the special role of the intestinal microbiota in the functioning of human organs and systems. It is no coincidence that the intestinal microbiota is actively involved in the metabolic function of the intestine, determines the functioning of endocrine organs, displaces exogenous pathogens, and regulates the immune functions of the body.

Under the conditions of a healthy organism, microbiota components provide the right balance of response to exogenous pro- and anti-inflammatory stimuli by regulating the production of cytokines and secretory immunoglobulins (mainly IgA). Components of the microbiota are capable of stimulating the immunological apparatus of the intestine, in particular, a complex of specific cells of the small intestine, mainly located in Peyer's patches.

A feature of the immune response to bacterial stimuli in reality triggers a response not only in the intestine itself, but also on other mucous membranes of the body: the nasopharynx, oral cavity, and urogenital system. This feature of mucosal immunity opens up opportunities for the formation of specific immune responses to an orally administered antigen not only in the intestine, but also in other parts of the body, which are most often the "gates" of infection caused by pathogenic viruses and bacteria.

In this respect, commensal intestinal microorganisms may function as vectors for presenting vaccine antigens to the immune system, and there is a real opportunity to create new vaccine preparations based on bacteria. It is known that probiotic strains are able to successfully maintain viability, passing through the gastric barrier and colonizing the small and large intestine for a long time. IEM has developed an approach that allows, by introducing the genes of pathogenic bacteria and viruses into the genome of lactic acid bacteria, to form a specific immune response against various pathogens. Features of the construction and functioning of mucosal vaccines against various pathogens, including the SARS-Cov-2 virus, discussed.

*Acknowledgements:* Scientific and educational center «Molecular bases of interaction of microorganisms and human» of the world-class research center «Center for personalized Medicine» FSBSI «IEM».



# Gene networks underlying the mechanisms of resistance to inflammatory factors in *Bifidobacterium longum* and *Lacticaseibacillus rhamnosus*

Veselovsky V.A.<sup>1\*</sup>, Dyachkova M.S.<sup>2</sup>, Olekhnovich E.I.<sup>1</sup>, Klimina K.M.<sup>1,2</sup>

<sup>1</sup> Department of Molecular Biology and Genetics, Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia

<sup>2</sup> Department of Biotechnology, Vavilov Institute of General Genetics, RAS, Moscow, Russia

\* djdf26@gmail.com

**Key words:** *Bifidobacterium longum*; *Lacticaseibacillus rhamnosus*, RNA sequencing; inflammatory process; pro-inflammatory cytokines; transcription start site; transcriptome

**Motivation and Aim:** Bifidobacteria and lactobacillus are a key component of the commensal gut microbiota conferring considerable health benefits and supporting the normal functioning of the host organism. As permanent residents of the normal gut microbiota, bifidobacteria and lactobacillus have evolved to adapt the host's immune response whose priority is to eliminate pathogenic agents. The mechanisms that ensure the survival of commensals during inflammation and maintain the stability of the core component of the normal gut microbiota in such conditions remain poorly understood. We performed transcriptome analysis of the *B. longum* GT15 and *L.rhamnosus* K32 strains to investigate the mechanisms lying behind the response of bifidobacteria and lactobacillus to signaling molecules of the immune system.

**Methods and Algorithms:** *B. longum* GT15 cells were collected in the middle of the log-phase in both experimental (with IL-6 and TNF $\alpha$ ) and control conditions. *L.rhamnosus* K32 cells were collected in the middle of the log-phase in both experimental (with IL-6, IL-8, IL-10 and TNF $\alpha$ ) and control conditions. Total RNA extraction and purification was performed using the RNeasy Mini Kit. Ribosomal RNA was removed using the RiboZero rRNA Removal Kit (Bacteria) and libraries were prepared using the NEBNext® Ultra II Directional RNA Library Prep Kit. The Kallisto v0.46.0 software was used for mapping of reads and estimation of transcript abundance. Differential expression analysis was performed using edgeR v3.26.8 package. The plots were generated using the ggplot2 package in R. Differentially expressed genes (DEGs) were then subjected to enrichment analysis of COG functions and KEGG pathways. The construction of PP was performed using the program OrthoFinder v.2.3.7. Identification of TSS for *B. longum* GT15 and subsequent analysis were performed from .sam files with BAC-Browser.

**Results:** Addition of IL-6 [0,1 ng/ml] and TNF $\alpha$  [10 ng/ml] on the growth medium led to moderate but statistically significant changes in the growth rate of the culture *B. longum* GT15. Genes that were significantly differentially expressed in response to the addition of cytokines were singled out: a total number of 128 DEGs in the sample of *B. longum* GT15 grown with IL-6 compared to the control sample, including 68 downregulated and 60 upregulated genes and 987 DEGs for the sample of *B. longum* GT15 grown with TNF $\alpha$  compared to the control sample, among which 477 genes were downregulated and 510 genes – upregulated. For *B. longum* GT15 the analysis of DEGs

showed, TNF $\alpha$  altered the expression of a higher number of genes than IL-6. The observed effects of cytokines on *B. longum* GT15 were cumulative and prolonged, features that are conducive for the adaptation to the new conditions. Gene (BLGT\_RS00625) encoding *hsp20* was upregulated 5-fold exposed to TNF $\alpha$ . Since the response of *B. longum* GT15 could be directed at bringing down TNF $\alpha$  levels, we suspect this mechanism to be one of the many ways used by bifidobacteria to dampen inflammation. As predicted by KEGG, 34 pathways were altered by IL-6 and 87 pathways were altered by TNF $\alpha$ . We used phylogenetic profiling to identify putative operons among the DEGs. The 410 potential TSS were identified for *B. longum* GT15. From them, 38 TSS confirm predicted DEGs operons for TNF $\alpha$  treatment and 7 operons for IL-6 treatment. According to functional analysis (COG), exposure to TNF $\alpha$  altered mainly the expression of genes included in energy pathways. This may be explained by the accumulation of resources caused by inflammation and activation of ATP-dependent ABC transporters to maintain homeostasis. Exposure to TNF $\alpha$  has also stimulated the expression of genes involved in propanoate metabolism is potentially one of the main metabolites accounting for the probiotic action of bifidobacteria. Exposure to TNF $\alpha$  can signal to bifidobacteria an increase in the concentration of ammonia leading to increased expression of genes, involved in nitrogen metabolism.

Interleukins do not affect the growth of the strain *L. rhamnosus* K32. But the analysis of DEGs showed IL-8 and IL-10 altered the expression of a higher number of genes than TNF $\alpha$  (507, 317 and 77 DEGs respectively). IL-6 doesn't affect gene expression *L. rhamnosus* K32. Exposure to cytokines IL-8 and IL-10 at various concentrations (0,1 ng/ml, 1 ng/ml and 10 ng/ml) affects the amount of DEGs, but does not affect their expression. Phylogenetic profiling predicted evolutionarily stable groups of DEGs combined into operons. 17 potential groups were predicted for IL-8 and 10 groups for IL-10. Common groups were found for IL-8 and IL-10 containing 22 DEGs. DEGs of these common operons are involved in the protection and maintenance of cell homeostasis. Most DEGs are ABC transporters. Also 2 DEGs (LRK\_05050 and LRK\_08445) refer to the XRE family transcriptional regulator that regulate transcription during stress. Functional analysis showed that most DEGs are involved in carbohydrate metabolism and transport. This may be due to intense energy-consuming processes activated by the addition of interleukins and associated inflammation.

**Conclusion:** We demonstrated how IL6 and TNF $\alpha$  activate their defense mechanisms, and also showed how bifidobacteria can exhibit probiotic effects in conditions of inflammatory processes. There is a single mechanism involved in the formation of resistance to immune response factors in the conditions of the inflammatory process in lactobacillus. Thus, in this study, we predicted the genes involved in a putative signaling pathway underlying the mechanisms of resistance to inflammatory factors in bifidobacteria and lactobacillus. Since bifidobacteria and lactobacillus are a major component of the human intestinal microbiota exhibiting pronounced anti-inflammatory properties, this study is of great practical and scientific relevance.

**Acknowledgements:** The study is supported by the Presidential Bursaries CII-4751.2021.4.

# PFNA operon of bifidobacteria: Role of bioinformatics in the discovery, structural and functional characterization, and possible application in biotechnology

Yunes R.A. \*, Nezametdinova V.Z., Alekseeva M.G., Danilenko V.N.

Vavilov Institute of General Genetics, RAS, Moscow, Russia

\*romanyunes@gmail.com

**Key words:** bifidobacteria, host-microbe interaction, fibronectin domains, PFNA cluster

*Motivation and Aim:* PFNA (The acronym is formed by the initial letters of the three genes: *pkb2*, *fn3*, *aaa-atp*) cluster is widespread among Bifidobacterium strains [1–4]. The PFNA cluster, which was found to function as an operon in the strain *B. longum subsp. longum* GT15, is made up by genes encoding species-specific proteins, which diverge in homology between different bifidobacterial species [1–4]. It has been shown that this divergence resulted from positive selection likely exerted by the host immune system [3]. One of the genes of the species specific PFNA operon, *fn3*, encodes a unique FN3 protein. The expression of this gene (as well as other genes of the PFNA operon) increased as the strain *B. longum subsp. longum* GT15 was grown in the presence of TNF $\alpha$  [5]. The FN3 protein is found in most bifidobacteria and it contains a signal peptide, a transmembrane domain, and two FN3 domains [1–4]. The presence of two FN3 domains in this protein containing motifs of cytokine receptor implied that this protein might be involved in binding cytokines produced by the host organism. An experiment devised and conducted previously demonstrated the ability of the protein fragment FN3 of *B. longum subsp. longum* GT15 to specifically bind to the human cytokine TNF $\alpha$  [6]. In this study, we further study the motifs of cytokine binding receptors in different species of bifidobacteria. Furthermore, non-synonymous substitutions (NSNP) in the sequenced genomes of strains of the species *B. longum longum* and *B. longum infantis* isolated from different populations were analyzed. The impact of the detected NSNPs on the 3D structures of the studied proteins was investigated.

*Methods and Algorithms:* The NCBI (<http://www.ncbi.nlm.nih.gov/>) and UniProt (<http://www.uniprot.org/>) databases were used to search and identify genes/proteins; the trRosetta (<https://yanglab.nankai.edu.cn/trRosetta/>) and IntFOLD (version 5.0) ([https://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD5\\_form.html](https://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD5_form.html)) programs to predict protein structures; the ProtParam (<http://web.expasy.org/protparam/>) program to calculate the molecular weight and isoelectric point of proteins.

*Results:* The gut-dwelling species of bifidobacteria were found to display strong differences in the sequences of the studied motifs [7]. Analysis of NSNPs made it possible to divide them into several groups according to the geographical location of the human populations from which the strains were isolated. The detected NSNPs lead to significant changes in the 3D structures of the studied proteins [8].

*Conclusion:* It can be assumed that the revealed structural characteristics of the cytokine receptor motifs are related to the species-specific adaptations employed by bifidobacteria inhabiting different hosts and to the interspecies competition of different species of

bifidobacteria that takes place in the intestines of the host organism. The study of the cytokine-binding properties of proteins of bifidobacteria is a promising field of research given the current epidemiological situation. The most lethal outcome of COVID-19 is disproportionate inflammation, otherwise known as cytokine storm. Selective binding of TNF- $\alpha$ , one of the key factors of inflammation, by the FN3 protein fragment of *Bifidobacterium longum* opens up prospects for the development of new drugs aimed at keeping this cytokine in check in target diseases.

*Acknowledgements:* This research was partially funded by the State program “Genetic technologies in biology, medicine, agricultural and environmental activities” No. 0112-2019-0002.

### References

1. Nezametdinova V.Z. et al. Identification and characterization of the serine/threonine protein kinases in *Bifidobacterium*. *Arch Microbiol.* 2014;196:125-136.
2. Nezametdinova V.Z. et al. Species-specific serine-threonine protein kinase Pkb2 of *Bifidobacterium longum* subsp. *longum*: Genetic environment and substrate specificity. *Anaerobe.* 2018;51:26-35.
3. Dyachkova M.S. et al. Positive selection in bifidobacterium genes drives species-specific host-bacteria communication. *Front Microbiol.* 2019;10:2374.
4. Zakharevich N.V. et al. Complete Genome Sequence of *Bifidobacterium angulatum* GT102: Potential Genes and Systems of Communication with Host. *Russ J Genet.* 2019;55(7):847-864.
5. Veselovsky V.A. et al. Gene networks underlying the resistance of *Bifidobacterium longum* to inflammatory factors. *Front Immunol.* 2020;11:595877.
6. Dyakov I.N. et al. FN3 protein fragment containing two type III fibronectin domains from *B. longum* GT15 binds to human tumor necrosis factor alpha in vitro. *Anaerobe.* 2020;65:102247.
7. Nezametdinova V.Z. The Role of the PFNA Operon of Bifidobacteria in the Recognition of Host's Immune Signals: Prospects for the Use of the FN3 Protein in the Treatment of COVID-19. *Int J Mol Sci.* 2021;22:9219.
8. Danilenko V.N. et al. Species-forming PFNA operon in Bifidobacteria: modules of sensor proteins Pkb2 and *fn3*, structure and distribution among different species and strains of Bifidobacteria derived from the human intestinal microbiome. *Genetika.* (in press).

## 5.2 Section “Microbial communities of natural and anthropogenic habitats”



## Comparative and evolution genomics of somatic antigens of Oxalobacteraceae family

Afonnikova S.D.

Applied Genomics Laboratory, SCAMT, ITMO University, St. Petersburg, Russia

\* [svetaafonnikova@gmail.com](mailto:svetaafonnikova@gmail.com)

**Key words:** O-antigen, somatic antigens, Oxalobacteraceae family

Many plants and animals have symbiotic relationships with bacteria. Interactions between bacteria and their hosts result in a different outcome for a host organism (neutral, harmful or have beneficial effects). Remarkably, these relationships are dynamic, as they change throughout an organism's lifetime and on evolutionary scale.

Investigating the association between bacteria lifestyle and their genomes would simplify the prediction of the possible evolution of bacteria. O-antigen is a key component in Gram negative bacteria's outer membrane. It facilitates interaction between bacteria and host immune system. High variability of the structure reflects the genomic variability of genes coding O-antigen synthesis components. There are two O-antigen processing pathways: *wzm/wzt*-dependent and *wzx/wzy*-dependent.

In our previous study an association between operon composition of O-antigens genes and bacteria lifestyles in genus *Herbaspirillum* (Oxalobacteraceae family) was found. The purpose of the current study was to explore, whether such association occurs in other genera of this family. Oxalobacteraceae family members can be found in soil, water, glaciers. Some species are described as plant and mushroom pathogens, opportunistic human pathogens. Initial data included genomes of 20 species.

Assembly quality was assessed by the QUAST tool. Then assemblies were annotated using Prokka and EggNOG. Search for genes of interest was conducted with Orthofinder. *Escherichia coli* O-antigen genes were used as a reference. The phylogenetic tree was reconstructed using T-Coffee alignment program and IQ-TREE tool.

We selected two *Collimonas* species, one species of genera *Herminiimonas*, *Oxalobacter*, and *Undibacterium*, four *Janthinobacterium*, and two *Collimonas* and *Oxalicibacterium* species. Assemblies of selected species were complete or no more than 10 contigs, however *M. putida* and *M. violaceinigra* genomes contained plasmid sequences. O-antigen genes could be located on plasmids, therefore assemblies with plasmids were considered separately in a further analysis.

We identified clusters of genes encoding the O-antigen for each species. They are dispersed throughout the genome, with no single cluster in any genome. L-rhamnose synthesis (*rml*) genes were present in the genomes of all analyzed species. The most common are genes for GDF-d-mannose synthesis (*manB*, *manC*), genes for protein with nucleotidyltransferase activity (*hddC*), glycosyltransferase (*wbaS*, *wbaX*, *wbdH*), and genes for O-antigen processing (*wzx*, *wzm*, *wzt*). Out of the 20 genomes, 13 contain *wzm/wzt* genes. Surprisingly, two species in *Oxalicibacterium* genus are likely to have different pathways. *O. flavum* possesses *wzt/wzm* genes, *O. faecigallinarum* lacks this gene group. To conclude, we identified conserved and variable parts of the O-antigen clusters. Currently, we are engaged in a more detailed study of the variable parts.

# Diversity and possible metabolic activity of the microbial community in nitrate- and radionuclide-contaminated groundwater

Babich T.L.<sup>1\*</sup>, Kalmykov S.N.<sup>2</sup>, Nazina T.N.<sup>1</sup>

<sup>1</sup> Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Moscow, Russia

<sup>2</sup> Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia

\* *microb101@yandex.ru*

**Key words:** nitrate- and radionuclide-contaminated groundwater, 16S rRNA gene, high-throughput sequencing, denitrification, radionuclides

*Motivation and Aim:* Since 1951, Lake Karachai located at the north of the Chelyabinsk region (Russia) has been used for the disposal of low- and intermediate-level liquid radioactive waste (RW) [1, 2]. Radioactive and other toxic components of radioactive waste were filtered through the permeable lake floor and concentrated in groundwater. Special attention has been paid to hydrochemical and radiochemical monitoring of the groundwaters around Lake Karachai. However, microbiological studies of this object have not been carried out. It is known that microorganisms affect the migration of radioactive and non-radioactive waste components [3–6]. Investigation of radionuclide-contaminated media is important both for determination of the composition and functional activity of microbial communities of these extreme environments, and for the development of biotechnologies for bioremediation of contaminated areas. The aim of the work was to determine the physicochemical and radiochemical conditions, the number and phylogenetic diversity of microorganisms, the isolation of dominant bacteria and the assessment of the possibility of their application for *in situ* bioremediation of groundwater from nitrate and reduction of radionuclides migration.

*Methods and Algorithms:* In this study were used analytical methods for analysis of the chemical and radiochemical composition of groundwater; microbiological methods – for enumeration and isolation of pure cultures of microorganisms; radioisotope methods – for estimation of the rate of microbial processes, sorption and reduction of radionuclides by microorganisms. The prokaryotic community of contaminated groundwater samples was analyzed by sequencing of 16S rRNA genes and Q-PCR of the functional genes of denitrification, nitrite reductase genes (*nirS* and *nirK*).

*Results:* Groundwater samples were collected from observation wells (at the 20–100 m deep) at a distance of 2.2–4.0 km from Lake Karachai. Water samples contained high concentrations of acetate, oxalate, nitrate and sulfate, as well as radionuclides (especially strontium) and heavy metals. The high redox-potential of the medium contributed to the preservation of radionuclides in a mobile oxidized form. Due to oxidative conditions in groundwater, high abundance of aerobic bacteria (up to  $10^5$  cells/ml) was detected. Fermentative (up to  $10^5$  cells/ml), denitrifying (up to  $10^4$  cells/ml) and iron-reducing bacteria (up to  $10^3$  cells/ml) were also present, while the number and activity of sulfate-reducing and methanogenic prokaryotes were low. By high-throughput sequencing of 16S rRNA gene, in contaminated groundwater was revealed a diverse microbial community represented mainly by *Proteobacteria*, uncultivated *Patescibacteria* of the

candidate phyla *Microgenomates* (OP11), *Parcubacteria* (OD1), and *Gracilibacteria* (GN02), and ammonium oxidizing archaea of the phylum *Thaumarchaeota* ('*Candidatus Nitrosopumilus koreensis*', '*Candidatus Nitrosotalea devanaterrea*'). More than 50 pure cultures isolated from groundwater samples belonged to 23 genera and 44 species of aerobic organotrophic and anaerobic denitrifying and iron-reducing bacteria. Bacteria of the genera *Pseudomonas*, *Pusillimonas*, *Rhizobium*, *Shewanella*, *Acinetobacter*, *Roseomonas*, *Pantoea*, and *Dechloromonas* reduced nitrates to nitrites and/or molecular nitrogen. The potential of microbial denitrification *in situ* was confirmed by the detection of *nirS/nirK* genes of denitrifying bacteria in the studied water samples. The genomes of strains of the genera *Pusillimonas* and *Roseomonas* contained genes that determine the reduction of nitrates, resistance to heavy metals and metalloids. These strains were also resistant to ionizing radiation. Bacteria of the genera *Shewanella* and *Pantoea* reduced Fe(III), U(VI) and Np(V) in media with acetate or lactate. Under neutral conditions, the isolated bacteria were able to sorb radionuclides (americium, uranium, plutonium, and neptunium) from the solution. The strains differed from each other by the spectrum of mainly sorbed radionuclides. Ultramicrobacteria and ultramicroforms belonging to the genera *Microbacterium*, *Salinibacterium* and *Chryseobacterium* have been isolated from groundwater.

**Conclusion:** The nitrate- and radionuclide-contaminated groundwater contained diverse population of denitrifying bacteria capable of reducing nitrates to dinitrogen and decreasing the redox-potential of the medium. This contributed to the growth of bacteria that reduced radionuclides and decreased the migration of radionuclides in groundwater. The data obtained are necessary for development of biotechnologies for *in situ* bioremediation of groundwater from nitrate and reduction of radionuclides migration.

### References

1. Rybal'chenko A.I. et al. Deep Injection Disposal of Liquid Radioactive Waste in Russia. Battelle Press: Columbus, OH, USA, 1998;206.
2. Novikov A.P. et al. Radionuclide content in underground waters from observation wells around Karachai Lake. *Radiokhimiya*. 1998;40:484-490.
3. North N. et al. Change in bacterial community structure during *in situ* biostimulation of subsurface sediment cocontaminated with uranium and nitrate. *Appl Environ Microbiol*. 2004;70:4911-4920.
4. Lloyd J.R., Renshaw J.C. Bioremediation of radioactive waste: radionuclide-microbe interactions in laboratory and field-scale studies. *Curr Opin Biotechnol*. 2005;16:254-260.
5. Xu M. et al. Responses of microbial community functional structures to pilot-scale uranium *in situ* bioremediation. *ISME J*. 2010;4:1060-1070.
6. Newsome L., Morris K., Lloyd J.R. The biogeochemistry and bioremediation of uranium and other priority radionuclides. *Chem Geol*. 2014;363:164-184.



## Extremal ecosystems as a basis for the creation of the collection of the ICG SB RAS

Bryanskaya A.V.<sup>1,2\*</sup>, Shipova A.A.<sup>1,2</sup>, Uvarova Y.E.<sup>1,2</sup>, Zadorozhny A.V.<sup>1,2</sup>, Meshcheryakova I.A.<sup>1,2</sup>, Antonov E.V.<sup>1,2</sup>, Korzhuk A.V.<sup>1,2</sup>, Kaplan V.S.<sup>1,2</sup>, Goryachkovskaya T.N.<sup>1,2</sup>, Peltek S.E.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [bal412003@mail.ru](mailto:bal412003@mail.ru)

**Key words:** microbial community, extremal ecosystems, collection of microorganisms

*Motivation and Aim:* In Russia there are many natural objects whose microbial communities are still poorly studied. Though these ecosystems are depository of microbiota with potentially valuable but little or unexplored properties. The microbial communities of salt lakes in the Novosibirsk Oblast and Altai Territory, hydrothermal springs in the Northern Baikal region and Kamchatka, and a number of anthropogenic ecosystems were studied in this work. The purpose of the study was targeted isolation of strains and creation a microorganisms collection of the ICG SB RAS based the data obtained about communities structure in the studied extremal ecosystems as a promising genetic and biotechnological resource.

*Methods and Algorithms:* From 2008 to 2021 we used methods of classical microbiology, light and electron microscopy, methods of molecular biology, and omics technologies.

*Results:* As a result of research carried out on salt lakes in the Novosibirsk Oblast and Altai Territory, thermal springs in the Baikal region and Kamchatka, and a number of anthropogenic ecosystems, a description of microbial communities of more than 50 extreme ecosystems was performed for the first time. We determined dominant and minor components of these ecosystems, as well as their distribution across thermal and chemical gradients. During this period, more than 2500 strains of microorganisms of various taxonomic groups were isolated to the collection of the ICG SB RAS. For more than 1000 strains whole genome sequencing was performed. Phylogenetic, phenotypic and mass-spectrometric characterization was carried out. On the basis of ongoing research, the prospects for the use of strains in biotechnology are assessed and scientific and practical recommendations for their use are developed. The database on the properties of isolated strains is being created.

*Conclusion:* The conducted studies allow us to conclude that microbial communities that have adapted to life in extreme conditions formed in the studied natural ecosystems. The communities are promising objects for both fundamental and applied research. We employed a set of microbiological techniques to create the ICiG SB RAS collection of microorganisms isolated from extremal environments. This collection currently includes over 2.500 strains that are investigated for potential biotechnological applications.

*Acknowledgements:* The study is supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. FWNR-2022-0022. We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education project No. 075-15-2019-1662 [Kurchatov Genomic Center, federal research center ICG SB RAS, Novosibirsk]) at the Federal Research Center ICG SB RAS. The work was conducted at the Multi-Access Center “Collection of Biotechnological Microorganisms as a Source of Novel Promising Objects for Biotechnology and Bioengineering” of the Federal Research Center ICG SB RAS.

## Producers and oxidizers of methane in the bottom sediments of Lake Baikal and their potential role in the formation of anomalously $^{13}\text{C}$ -enriched authigenic carbonates

Bukin S.<sup>1\*</sup>, Lomakina A.<sup>1</sup>, Pogodaeva T.<sup>1</sup>, Khlystov O.<sup>1</sup>, Zemskaya T.<sup>1</sup>, Krylov A.<sup>2</sup>

<sup>1</sup>Limnological institute, SB RAS, Irkutsk, Russia

<sup>2</sup>All-Russia Institute for Geology and Mineral Resources of the World Ocean (VNIIOkeangeologia), St. Petersburg, Russia

\* sergeibukin@lin.irk.ru

**Key words:** Lake Baikal, authigenic carbonates, microbial community, siderite, hydrocarbons

*Motivation and Aim:* In the sediments of Lake Baikal, methanogenesis is the main terminal process in the chain of anaerobic organic matter (OM) degradation. Also at the bottom of the lake there are many zones of discharge of hydrocarbon-containing fluids, in which methane is not only of biogenic, but also of thermogenic origin. Therefore, the study of microorganisms involved in the methane cycle is necessary for understanding the biogeochemical processes occurring in Baikal sediments. One of these processes is the formation of anomalously  $^{13}\text{C}$ -enriched authigenic carbonates ( $\delta^{13}\text{C} \leq +36.3$  ‰ VPDB) in the discharge areas in contrast with almost complete absence of carbonates in the other sediments of the lake. If in marine cold seeps and mud volcanoes, carbonate formation is driven by sulfate-dependent AOM, in Lake Baikal, where  $\text{SO}_4^{2-}$  is 500-fold less abundant than in seawater, active methanogenic destruction of OM and silicate weathering was proposed as such a process [1]. Therefore, the purpose of this study was to reveal the features of the structure of microbial communities and the chemical composition of pore waters in sedimentary layers containing isotopically heavy carbonates, as well as to assess their compliance with the hypothesis of the formation of carbonates during the methanogenic destruction of OM.

*Methods and Algorithms:* We analysed the composition of microbial communities in the carbonate layers and normal bottom sediments of three mud volcanoes MV Malenky, MV Kukuy K-9 and MV Kedr-1. Sediment samples were collected using the benthos and gravity corers. Diversity of microorganisms was assessed via 16S rRNA gene diversity survey using high-throughput sequencing on Illumina MiSeq. The chemical composition of pore water was determined using liquid chromatography, atomic absorption and flame emissivity methods. The isotopic composition ( $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$ ) of carbonates and dissolve organic carbon (DIC) was measured with a Finnigan Delta plus XP mass spectrometer. Scanning electron microscopy and spot analyses of Ca, Mg, Mn and Fe in carbonate samples were carried out using a Quanta 200 microscope with an energy-dispersive X-ray spectroscopy SDD Apollo 10 attachment.

*Results:* In sediments of MV Malenky and MV Kukuy K-9, authigenic siderites ( $\text{FeCO}_3$ ) with moderately positive  $\delta^{13}\text{C}$  values +7.24 and +7.77 ‰ VPDB were deposited in diatom ooze at depths 27–28 and 36–37 cmblf respectively. In MV Kedr-1, siderite nodules were located in clays and had extremely heavy isotopic  $\delta^{13}\text{C}$  values: +32.25 ‰ VPDB (125–127 cmblf), +27.39 ‰ VPDB (127–130 cmblf) and +33.32 ‰ VPDB (150–152 cmblf). At the same time,  $\delta^{13}\text{C}$ -DIC values measured for the bottom sediments of

MV Kedr-1 were on average lower in sedimentary layers with carbonates (from +8 to +10.5‰ VPDB) than in carbonate-free interlayers (from +9.7 to +19.5‰ VPDB).

Analysis of the chemical composition of the pore water showed that the carbonate interlayers fall into the core intervals with elevated ammonium concentrations (0.56–2.88 mg/L). However, statistical analysis did not reveal significant differences between carbonate layers and the closest carbonate-free sediments.

The analysis of the composition and structure of microbial communities associated with siderites revealed that they had a very similar range of dominant taxa, the members of which can destruct organic matter (*Bathyarchaeia*, *Hadarchaeales*, *Cladatribacteriota*, *Chloroflexi*, *Caldisericota*) and produce methane (*Methanomassiliicoccales*, *Methanosaetaceae*, *Methanoregulaceae*). However, the significant proportion of methanotrophic microorganisms (*Methanoperedensaceae*) was also identified in the carbonate interlayers of MV Malenky and MV Kedr-1. We did not detect iron-reducing microorganisms of the genera *Shewanella*, *Geobacter*, *Geothrix*, and *Thermanoerobacter* typically associated with siderites in other ecosystems. Also, the communities of carbonate layers did not have significant differences from the microbial populations of the surrounding carbonate-free sediments.

**Conclusion:** The available data suggest that the activity of microbial communities probably is not the only factor leading to the crystallization of siderites in the sediments of Lake Baikal. The discharge activity and the composition of fluids that provide additional supply of DIC or CO<sub>2</sub> from deeper sediment intervals can also affect the formation of carbonates and their isotopic composition in mud volcanoes. In the sediments of MV Kedr-1, the discharge of secondary microbial gas characterized by the presence of CO<sub>2</sub> enriched in isotope <sup>13</sup>C ( $\delta^{13}\text{C} \leq +30$  ‰ VPDB) was recorded [2], whereas in the sediments of MV Malenky and MV Kukuy K-9, gas have microbial or early mature thermogenic origins with the  $\delta^{13}\text{C}$  values of CO<sub>2</sub>  $\leq +4.3$  and  $\leq +11.0$  ‰ VPD respectively [3; e.g. A. Hachikubo, unpublished]. Under the conditions of additional fluid influx, the deposition of carbonates can occur in the layers with locally elevated DIC concentrations/alkalinity generated during various biogenic processes, including methanogenic destruction of organic matter coupled with silicate weathering, and, possibly methane oxidation, which are insufficient under normal conditions to shift equilibrium towards the deposition of carbonate minerals. In this case, isotopic characteristics of DIC or CO<sub>2</sub> coming with the fluids will largely determine the  $\delta^{13}\text{C}$  values of the formed carbonates that are differ between the study areas.

**Acknowledgements:** The study is supported by the Russian Scientific Foundation grant No. 22-14-00084.

## References

1. Krylov A., Khlystov O., Zemskaya T., Minami H., Hachikubo A., Kida M. et al. First discovery and formation process of authigenic siderite from gas hydrate-bearing mud volcanoes in fresh water: Lake Baikal, Eastern Siberia. *Geophys Res Lett.* 2008;35:L05405.
2. Hachikubo A., Minami H., Yamashita S., Khabuev A., Krylov A., Kalmychikov G. et al. Characteristics of hydrate-bound gas retrieved at the Kedr mud volcano (southern Lake Baikal). *Sci Rep.* 2020;10:14747.
3. Krylov A.A., Khlystov O.M., Hachikubo A., Minami H., Nunokawa Yu., Shoji H. et al. Isotopic composition of dissolved inorganic carbon in the subsurface sediments of gas hydrate bearing mud volcanoes, Lake Baikal: Implications for methane and carbonate origin. *Geo-Mar Lett.* 2010;30:427-437.

## Methanotroph diversity in natural and anthropogenic habitats as an inexhaustible source of novel biotechnologically relevant strains

Dedysh S.N.

Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Moscow, Russia

Dedysh@mail.ru

**Key words:** methanotrophic bacteria, biotechnological potential, methanotroph diversity, strains-producers, genome analysis

*Motivation and Aim:* Aerobic methanotrophs are metabolically unique bacteria that are able to utilize methane as a sole source of energy. A defining characteristic of these organisms is the use of methane monoxygenase enzymes to catalyze the oxidation of methane to methanol. This metabolic capability makes methanotrophs an attractive target for various biotechnological applications for the conversion of methane to a microbial single cell protein or other value-added products. However, despite the long history of studying aerobic methanotrophs, the cultured diversity of these bacteria and, in particular, the number of biotechnologically relevant strains with high growth characteristics remains limited. The project acknowledged in this report was initiated in order to expand the pool of currently available fast-growing methanotrophs by isolating novel strains of these bacteria from various natural and anthropogenic habitats where both methane and oxygen are available.

*Methods:* The samples of activated sludge collected from several wastewater treatment plants, sediments from freshwater ponds and rivers of various location, as well as landfill cover soils were used as sources for isolation of new methanotroph strains. The cultivation approach was adjusted for isolating methanotrophs growing in the temperature range 35–45°C. The obtained isolates were characterized phenotypically and used for the experiments on continuous cultivation in 1.5 l bioreactors. The genome sequences of novel strains were determined by the combination of MiSeq and Nanopore techniques and used for comparative genome analyses.

*Results:* As revealed by the high-throughput 16S rRNA gene sequence analysis, methane-oxidizing enrichment cultures obtained from various samples contained representatives of the genera *Methylococcus*, *Methylomonas*, *Methylocaldum*, *Methylosinus* and *Methylocystis*. Further attempts were focused on isolating novel strains of fast-growing *Methylococcus* species, since the latter have been extensively exploited for large-scale commercial production of microbial proteins. Four novel isolates of these bacteria were isolated from activated sludge, landfill cover soil and freshwater sediments. The determined genome sizes of novel isolates varied between 3.2 and 4.0 Mb. As revealed by the phylogenomic analysis, strains IO1, BH and KN2 affiliated with *Mc. capsulatus*, while strain Mc7 represented a novel species. Highest temperature optima (45–50°C) and highest growth rates in bioreactor cultures (up to 0.3 h<sup>-1</sup>) were recorded for strains obtained from activated sludge. The comparative analysis of all complete genomes of *Methylococcus* species revealed surprisingly high genome synteny and a striking similarity in gene content in all strains of *Mc. capsulatus*,

which also displayed high growth rates on methane. Another research direction was focused on obtaining new strains of fast-growing methanotrophs that produce carotenoids, since the latter increase the value of feed protein. Five novel isolates of yellow- to red-pigmented *Methylomonas* species have been obtained and characterized with regard to their growth characteristics and the composition of produced carotenoids. One of the novel isolates was shown to be especially promising with regard to high growth rates on methane and methanol. This isolate was also assigned to the novel species of the genus *Methylomonas*, '*M. rapida*'.

*Conclusion:* In summary, methanotroph diversity in natural and anthropogenic habitats remains an inexhaustible source of novel biotechnologically relevant strains.

*Acknowledgements:* This 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".

## Third-generation cephalosporin and fluoroquinolone resistance in *Salmonella enterica* associated with antimicrobial resistance genes

Egorova A. \*, Shelenkov A., Saenko S., Chernyshkov A., Kulikova N., Manzeniuk I., Mikhaylova Y., Akimkin V.

Central Research Institute of Epidemiology of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Moscow, Russia

\*bioanna1995@gmail.com

**Key words:** *Salmonella enterica*, WGS, third-generation cephalosporin resistance, *blaCTX-M*

**Motivation and Aim:** *Salmonella enterica* is one of the most important global food-borne pathogens. Since ciprofloxacin (belonging to fluoroquinolones), ampicillin or third-generation cephalosporins (beta-lactams) are the common first-line antimicrobial drugs for the treatment of salmonellosis [1] it is important to study molecular mechanisms of resistance to these drugs and to monitor the dissemination of antimicrobial resistance (AMR) determinants. Thus, the goal of this work was to perform genomic analysis ciprofloxacin- and third-generation cephalosporin-resistant (CIP-3rd) isolates of food origin based on whole-genome sequencing (WGS).

**Methods and Algorithms:** Totally 10 CIP-3<sup>rd</sup> *Salmonella* isolates were obtained in Russia in 2021 from poultry products (n = 6), meat (n = 1) and ready-to-cook products (n = 4). Samples were received from Samara region (n = 1), Tula region (n = 1), Chelyabinsk region (n = 1), Magadan region (n = 1), Vologda region (n = 1), Primorsky Krai (n = 1), Sverdlovsk region (n = 2) and Novosibirsk region (n = 2). The isolates were tested for phenotypic antimicrobial resistance by the boundary concentration method on VITEK MS analyzer. WGS was performed on the Illumina NextSeq platform using Nextera DNA Sample Prep Kit (Illumina, USA). Genome assemblies and comparisons were performed using SPAdes, roary, dnadiff, and custom software.

**Results:** The majority of the analyzed *S. enterica* isolates belonged to ST32 and serovar O7 (n = 6), two were typed as ST548/O21, ST469/O7, while ST11/O9 was revealed in one isolate. All 10 isolates studied were characterized by AMR to ceftazidime (CZD) and ceftriaxone (CRO), which corresponds well to the revealed AMR determinants according to WGS data obtained. So, all *S. enterica* (ST32) carried *blaCTX-M-14* gene which was responsible for 3<sup>rd</sup>-generation cephalosporins resistance [2-4]. ST548, ST469, and ST11 carried  $\beta$ -lactamase genes *blaCMY-2*, *blaCTX-M-1*, and *blaCTX-M-55* respectively. Eight of the isolates presented also possessed resistance to cefepime (FEP) which belongs to fourth-generation cephalosporins. All FEP-resistant samples carried *blaCTX-M* gene family despite the different sequence types. Such dissemination raises concerns and requires further epidemiological surveillance of AMR transmission within *S. enterica* and investigations of the molecular mechanisms of resistance acquisition.

Moreover, all the isolates observed had ciprofloxacin (CIP) resistance. However, the presence of AMR genes to fluoroquinolones (*qnrB19*) and quinolones (*qnrS1*) was revealed only in two and three isolates, correspondingly. We suggest that these

distinctions could be associated with point mutations in a chromosomal virulence gene *gyrA* being the most common mechanism for resistance to quinolones in *Salmonella* [5]. *In silico* search for plasmids showed that all CIP-3rd ST32 samples were characterized by the presence of IncFIB plasmid, while ST548 carried IncC, ST469 had several types of plasmids simultaneously (Col(pHAD28) and IncI1), which was also revealed in ST11 (IncFIB, IncFII, IncI1).

**Conclusion:** Applying WGS technology enabled us to obtain unique genome features of foodborne CIP-3<sup>rd</sup> *S. enterica* isolates. Such features are important both from the standpoint of an increased level of AMR of the pathogens studied and evolutionary changes. The data obtained will contribute to further investigations in epidemiological surveillance of foodborne *S. enterica* isolates.

**Acknowledgements:** The study was supported by the industry research program of Rospotrebnadzor (2021–2025) “Scientific support of epidemiological monitoring and public health measures of the Russian Federation territory” and Russian Federation Government Resolution No. 3116-r.

### References

1. Medalla F. et al. Estimated incidence of antimicrobial drug-resistant nontyphoidal *Salmonella* infections, United States, 2004–2012. *Emerg Infect Dis.* 2017;23(1):29.
2. Kameyama M. et al. Emergence of *Salmonella enterica* serovar Infantis harboring IncI1 plasmid with *blaCTX-M-14* in a broiler farm in Japan. *J Veterinary Med Sci.* 2012;11:88.
3. Tamang M.D. et al. Emergence of extended-spectrum  $\beta$ -lactamase (*CTX-M-15* and *CTX-M-14*)-producing nontyphoid *Salmonella* with reduced susceptibility to ciprofloxacin among food animals and humans in Korea. *J Clinical Microbiol.* 2011;49(7):2671-2675.
4. Ragupathi N.K.D. et al. Draft genome sequence of *blaTEM-1*-mediated cephalosporin-resistant *Salmonella enterica* serovar Typhi from bloodstream infection. *J Glob Antimicrob Resist.* 2016;7:11-12.
5. Dimitrov T. et al. Ciprofloxacin-resistant *Salmonella enterica* serovar Typhi from Kuwait with novel mutations in *gyrA* and *parC* genes. *J Clinical Microbiol.* 2009;47(1):208-211.

## Temperature and carbon source as factors affecting the biooxidation of sulfide concentrate

Elkina Yu., Artykova A., Kolosov A., Nechaeva A., Beletsky A., Gruzdev E., Melamud V., Mardanov A., Bulaev A.\*

Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia

\* bulaev.inmi@yandex.ru

**Key words:** acidophilic microorganisms, biohydrometallurgy, sulfide concentrates

*Motivation and Aim:* Biohydrometallurgical technologies (reactor biooxidation), based on the destruction of sulfide gold-bearing minerals (pyrite, arsenopyrite) by microorganisms, are widely used for processing gold-bearing concentrates. A number of problems limit the efficiency of this treatment of sulfide concentrates, for example, since the oxidation of sulfide minerals resulting in the release of heat, and overheating of the reactor pulp, in turn, leads to inhibition of biological processes, a certain temperature in the biooxidation reactors should be maintained. For this purpose, cooling systems are used, which are not always able to maintain the required temperature due to a number of factors (climatic conditions, equipment wear), and also consume a significant amount of electricity. An important task is to develop approaches that may minimize the negative impact of overheating on the biooxidation of sulfide concentrates. The goal of the present work was to study the biooxidation of sulfide concentrate under different conditions.

*Methods and Algorithms:* The concentrate contained 56 % pyrite and 12 % arsenopyrite, 43 g/t gold. Biooxidation was carried out in laboratory scale reactors in a continuous mode at 40, 45, and 50 °C. The residence time was 6 days, the pulp density was 1 : 10 (S : L). The influence of carbon sources on the biooxidation process was investigated by adding 0.02, 0.04, 0.10 % molasses to the pulp or by supplying CO<sub>2</sub>. Microbial populations were analyzed by metabarcoding of V3-V4 fragments of 16S rRNA genes.

*Results:* At 40 °C, additional carbon sources slightly affected the biooxidation and did not allow increasing gold recovery. At higher temperatures, the use of CO<sub>2</sub> affected biooxidation, and at 50°C allowed increasing gold recovery. At 40 °C, carbon sources practically did not affect the composition of microbial populations, and genera *Leptospirillum*, *Acidithiobacillus* and *Ferroplasma* were predominant. At 45°C, the proportions of the genera *Acidithiobacillus* and *Ferroplasma* were high in the populations of all reactors. The proportion of *Acidiplasma* was high in all reactors, with exception of the reactor with 0.10 % of the molasses. The proportion of the bacteria of the genus *Sulfobacillus* was high in the control reactor as well as in the reactor with molasses. When CO<sub>2</sub> was used, the share of archaea of the genus *Cuniculiplasma* and uncultivated group A-plasma was also high. At 50 °C, representatives of the genera *Acidithiobacillus*, *Sulfobacillus*, and *Acidiplasma*, as well as unknown representatives of the genus *Acinetobacter* prevailed in all reactors. At all temperatures in the reactor where CO<sub>2</sub> was supplied, the total number of microbial cells was 2–3 times higher than in other reactors. *Conclusion:* Thus, the results obtained show that the use of CO<sub>2</sub> as an additional source of carbon can be considered as an approach to prevent the inhibition of the biooxidation of sulfide concentrates with an increase in temperature, and its effect on biooxidation can be explained by the effect on the composition of microbial populations of the reactors.

*Acknowledgements:* The research was supported by the Russian Science Foundation – grant No. 21-64-00019.



# Halophilic bacteria of the genera *Halomonas* and *Marinobacter* from petroleum reservoirs and their possible applications in biotechnology

Ershov A.<sup>1\*</sup>, Sokolova D.<sup>1</sup>, Semenova E.<sup>1</sup>, Tourova T.<sup>1</sup>, Grouzdev D.<sup>2</sup>, Nazina T.<sup>1</sup>

<sup>1</sup> Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Moscow, Russia

<sup>2</sup> SciBear OU, Tallinn, Estonia

\* e.alexey.mail@yandex.ru

**Key words:** petroleum reservoirs, halophiles, alkane oxidation, whole-genome sequencing

**Motivation and Aim:** An exploitation of many oil fields becomes complicated because of high salinity of formation waters. Increased salt concentration makes difficulties in oil production, purification and disposal of oil-contaminated waters. The use of microbial technologies on oil fields with high salinity formation water is limited due to the low solubility of oxygen and the negative impact of salt on microorganisms, including the destruction of the cell walls and denaturation of enzymes [1]. Bacteria of the genera *Halomonas* and *Marinobacter* are often presented in microbial communities of these habitats, including marine reservoirs, deep underground horizons, oil reservoirs, and wastewaters from these sites [2, 3]. Thus, the aim of this study was to determine the composition of the microbial communities of oil reservoirs with highly mineralized formation waters, to isolate halophilic hydrocarbon-oxidizing bacteria, to study their physiological and genomic characteristics and their role in oil biotransformation.

**Methods and Algorithms:** Bacterial strains TAT1 and KAZ22 were isolated from bed 302 of the Romashkinskoe oil field (Republic of Tatarstan, Russian Federation) and from the Uzen oil field (Republic of Kazakhstan) respectively. Strains were identified using the 16S rRNA gene sequencing with an ABI 3730 DNA Analyzer sequencer and the ABI PRISM BigDye™ Terminator v 3.1 reagent kit (Applied Biosystems, USA). The alkane monooxygenases genes (*alkB*, *ladAα*, *ladAβ*, *ladB*, *CYP153A*) were detected using bacterial strains from the collection of the Research Center of Biotechnology RAS and a strain of *Geobacillus icigianus* G1w1<sup>T</sup> provided by the staff of the Institute of Cytology and Genetics SB RAS [4]. The strain genomes were sequenced using the Illumina HiSeq 2500 system (USA). Microbial communities' composition was determined using high-throughput sequencing of the V3–V4 regions of the 16S rRNA genes. Reconstruction of probable metabolic pathways was performed using the NCBI, PATRIC and KEGG online services.

**Results:** The presence of bacteria of the genus *Halomonas* in the Romashkinskoe oil field, characterized by low temperature and high sulfide content in reservoir fluids, and bacteria of the genus *Marinobacter* in the high-temperature Uzen oil reservoir was shown using high-throughput sequencing. Halophilic bacteria *Halomonas titanicae* TAT1 and *Marinobacter lutaoensis* Kaz22 were isolated from oil reservoirs, their physiological characteristics were determined, and their genomes were sequenced. Both strains utilized a wide range of oil hydrocarbons with the formation of biosurfactants in a liquid media with salinity higher than 100 g NaCl/L, and differed in the temperature ranges for growth, which were 4–42°C for the *Halomonas* strain and 22–55°C for the

*Marinobacter* strain. The genomes of the strains contained genes that determine resistance to high salinity and genes for alkane hydroxylases responsible for the degradation of *n*-alkanes with different chain length. In *H. titanicae* TAT1 *alkB* gene was detected and in *M. lutaoensis* KAZ22 *alkB1*, *alkB2*, *alkB3*, and *almA* genes were detected. Several genes responsible for the production of osmoprotectants were presented in both strains. The *betIBA* and *ectABC* gene clusters and separate *ectD1* and *ectD2* genes were found in the *Halomonas* strain, so it was presumably able to produce betaine, ectoine, and hydroxyectoine. The *Marinobacter* strain did not show the ability to synthesize betaine, but possessed the *ectAB* gene cluster with an aspartate kinase gene and separately located *ectC* and *ectD* genes, that is, it was also a potential producer of ectoine and hydroxyectoine. Strain TAT1 grew heterotrophically under anaerobic conditions with organic substrates reducing nitrate to nitrite. This ability makes a possibility of using this strain in biotechnology to suppress the formation of hydrogen sulfide in an oil reservoir. The strain *M. lutaoensis* KAZ22 was incapable of nitrate reduction, and nitrate reductase genes in its genome were absent.

**Conclusion:** Two bacterial strains, *Halomonas titanicae* TAT1 and *Marinobacter lutaoensis* KAZ22, are isolated from the highly mineralized Romashkinskoe (Russian Federation) and Uzen (Republic of Kazakhstan) oil fields respectively, their physiology is studied, and their genomes are sequenced. In the genome of the *Halomonas* strain the *alkB* gene for alkane-1-monooxygenase is found; in the *Marinobacter* strain two genes *alkB1* and *alkB2* and the flavin-dependent monooxygenase *almA* gene are found as well. All these genes belong to the separate phylogenetic clusters. *Halomonas titanicae* TAT1 strain is potentially capable of synthesizing osmoprotectants betaine, ectoine and hydroxyectoine; strain *Marinobacter lutaoensis* KAZ22 is able to produce ectoine and hydroxyectoine. The *Halomonas* strain reduces nitrate with nitrite accumulation, which indicates the possibility of its use to suppress the formation of hydrogen sulfide in the reservoir. Both *Halomonas* and *Marinobacter* strains oxidize oil alkanes under high salinity conditions with the formation of biosurfactants, which indicates the possibility of their application in biotechnologies for cleaning oil pollutions and microbial enhancement of oil recovery.

**Acknowledgements:** The study was supported by the Russian Science Foundation, grant No. 21-64-00019.

## References

1. Fathepure B.Z. Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. *Front Microbiol.* 2014;5:173.
2. An B.A., Shen Y., Voordouw G. Control of sulfide production in high salinity Bakken shale oil reservoirs by halophilic bacteria reducing nitrate to nitrite. *Front Microbiol.* 2017;8:1164.
3. Sokolova, D.S. et al. Sulfidogenic microbial communities of the Uzen high-temperature oil field in Kazakhstan. *Microorganisms.* 2021;9:1818.
4. Bryanskaya A.V. et al. *Geobacillus icigianus* sp. nov., a thermophilic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol.* 2015;65:864-869.

## Bioleaching of sulfide copper-nickel ores

Kolosov A.<sup>1\*</sup>, Latyuk E.<sup>1,2</sup>, Melamud V.<sup>1</sup>, Bulaev A.<sup>1\*</sup>

<sup>1</sup> Federal Research Centre “Fundamentals of Biotechnology”, RAS, Moscow, Russia

<sup>2</sup> Institute of North Industrial Ecology Problems of the Federal Research Centre Kola Science Centre, RAS, Apatity, Russia

\* alexander\_thechemist\_kolosoff@mail.ru

**Key words:** bioleaching, sulfide ores, copper, nickel, chemolithotrophic microorganisms

*Motivation and Aim:* Significant reserves of the most important types of mineral raw materials are concentrated in the Murmansk Oblast, and large number of deposits of sulfide copper-nickel ores have been discovered in this region. At the same time, there are a number of deposits in the region that are not exploited due to the low content of valuable components and of small reserves. Bioleaching is a simple, environmentally friendly and cost-effective approach to process sulfide ores, including low-grade ones. The goal of this work was to perform laboratory trials to determine the possibility of extracting Cu and Ni from copper-nickel ores using the heap bioleaching method.

*Methods and Algorithms:* Ores from the Sopcha and Nud II deposits were objects of the study. The main ore minerals were pentlandite, chalcopyrite, pyrrhotite. The ores were crushed to  $-5+1$  mm. The Sopcha ore sample contained 0.39 % Ni and 0.27 % Cu; the Nud II ore sample contained 0.72 % Ni and 0.20 % Cu. Mixed culture of acidophilic microorganisms was used. The tests were carried out at three temperatures: at 25, 35 and 45 °C, since heap bioleaching processes are characterized by self-heating of the ore pile and it is important to evaluate the possibility of the bioleaching at different temperatures. Solid to liquid ratio in the percolators was 1:1. The duration of experiments with Sopcha ore was 159 days, with Nud II ore – 138 days.

*Results:* The extraction of metals into solution from ore Sopcha ore at different temperatures was: at 25 °C – 19.4 % Ni and 7.54 % Cu, at 35 °C – 15.1 % Ni and 5.6 % Cu, at 45 °C – 13.0 % Ni and 4.5 % Cu. Extraction of metals into solution from ore Nud II at different temperatures was: at 25 °C – 17.3 % Ni and 4.96 % Cu; at 35 °C – 24.53 % Ni and 6.64 % Cu; at 45 °C – 30.58 % Ni and 10.43 % Cu. At the end of the experiment, during the bioleaching of Sopcha ore, uncultivated representatives of Acetobacteraceae bacteria of the genera *Leptospirillum* and *Ferrimicrobium*, archaea of the genus *Cuniculiplasma* and *Ferroplasma*, and uncultivated representatives of *Thermoplasmataceae* predominated in microbial populations at all temperatures. During the bioleaching of Nud II ore, representatives of the genera *Leptospirillum* and *Ferrimicrobium* dominated in microbial populations at 25 and 35 °C, and *Acinetobacter* and *Staphylococcus* dominated at 45 °C.

*Conclusion:* The results of the experiments demonstrated the possibility of long-term bioleaching of the studied ore to extract non-ferrous metals, while the extraction of nickel may be of greater practical interest due to its higher cost and higher leaching rate.

*Acknowledgements:* The study was partially supported by RFBR (project No. 19-35-50073) and by Ministry of Science and Higher Education of the Russian Federation.

# Effect of organic loading rate on the biohythane production in the continuous two-stage anaerobic digestion of confectionery wastewater

Litti Yu.<sup>1\*</sup>, Mikheeva E.<sup>2</sup>, Katraeva I.<sup>3</sup>, Zhuravleva E.<sup>1</sup>, Kovalev A.<sup>4</sup>

<sup>1</sup> Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia

<sup>2</sup> Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

<sup>3</sup> Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

<sup>4</sup> Federal Scientific Agroengineering Center VIM, Moscow, Russia

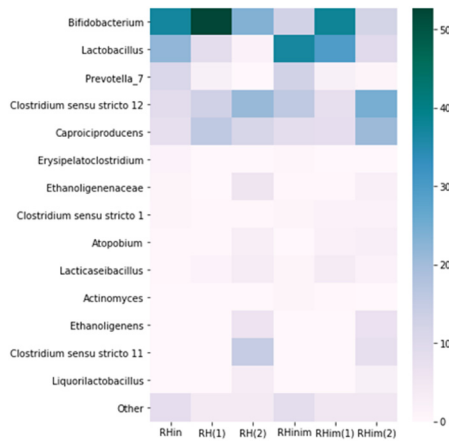
\* [litty-yuriy@mail.ru](mailto:litty-yuriy@mail.ru)

**Key words:** biohythane, confectionery wastewater, two-stage anaerobic digestion, microbial community

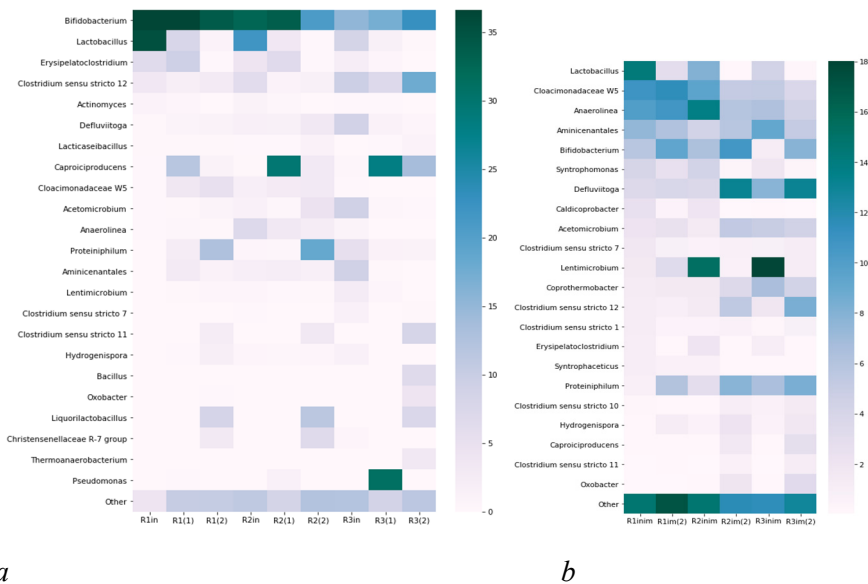
**Motivation and Aim:** A mixture of hydrogen and methane obtained as a result of two-stage anaerobic digestion (TSAD) of various organic wastes is called biohythane. It is attracting particular attention due to its excellent performance in terms of lower greenhouse gas emissions (CO, CO<sub>2</sub> and NO<sub>x</sub>) and high thermal efficiency, so in the future, biohythane could be a potential gaseous alternative energy source [1]. Organic loading rate (OLR) and hydraulic retention time (HRT) are among the main factors affecting biohythane production performance. The purpose of this work was to study the effect of different OLR and HRT on the efficiency of biohythane production, energy recovery and microbial community structure during TSAD of confectionery wastewater.

**Methods and Algorithms:** The TSAD process was carried out in a laboratory biogas plant, consisting of 4 biofilter type reactors, each with polyurethane carriers for biomass immobilization, with a working volume of 900 ml and with an upward liquid flow of 3 m/h. The hydrogen producing reactor (RH) operated in mesophilic mode (37 ± 1) °C with an HRT of 0.42 days, and three methanogenic reactors (R1, R2 and R3) operated in a thermophilic mode (55 ± 1) °C with an HRT of 3, 2 and 1 days respectively. During the 2 phases of the experiment, the OLR was increased, from 12.8, 1.79, 2.69 and 5.38 g COD/(L d) respectively, in RH, R1, R2 and R3, by 2.05 times. Microbial community analysis in planktonic and attached biomass was performed based on the 16S rRNA gene sequencing on the Illumina platform.

**Results:** Butyric acid type fermentation was the main pathway of H<sub>2</sub> production in RH. Both the highest H<sub>2</sub> yield of 87.8 ml/g COD and H<sub>2</sub> production rate of 2352 ml/(L d) were obtained at an OLR of 26.8 g COD/(L d). The highest CH<sub>4</sub> yield of 335 ml/g COD and overall energy yield of 13.36 kJ/g COD were achieved in OLR<sub>RH</sub> of 12.8 g COD/(L d) and OLR<sub>R1</sub> of 1.79 g COD/(L d) configuration of TSAD. CH<sub>4</sub> and overall energy production rate (1543 and 91.56 kJ/(L d) respectively) were highest at OLR<sub>RH</sub> of 26.8 g COD/(L d) and OLR<sub>R2</sub> of 5.5 g COD/(L d). R3, having the lowest HRT of 1 day, failed to produce biomethane. Microbial community in RH was dominated by *Lactobacillus*, *Bifidobacterium*, *Clostridium sensu stricto* 12 and *Caproiciproducens* (Fig. 1); methanogenic archaea were effectively inactivated. Methane was produced through hydrogenotrophic route by representatives of *Methanothermobacter*. The planktonic bacterial communities of methanogenic reactors were dominated by *Bifidobacterium*; *Caproiciproducens* and *Proteiniphilum* were most abundant at lower and higher OLRs respectively. The attached biomass was more diverse and dominated by *Anaerolinea*, *Bifidobacterium*, *Defluviitoga* and *Proteiniphilum* (Fig. 2).



**Fig. 1.** Heatmap of the profile of planktonic and immobilized bacterial communities at the genera level for RH



**Fig. 2.** Heatmap of the profile of planktonic (a) and immobilized (b) bacterial communities at the genera level for R1, R2, and R3

**Conclusion:** Mesophilic-thermophilic mode of TSAD was found to be effective for production of biohythane from confectionery wastewater. The most efficient configurations in terms of OLR and HRT have been found to maximize gas and energy yields and productions rates. Dominant microbial groups were identified to assess their contribution to the performance of TSAD.

**Acknowledgements:** The study was supported by the Russian Science Foundation grant No. 21-79-10153, <https://rscf.ru/project/21-79-10153>.

### References

1. Hans M., Kumar S. Biohythane production in two-stage anaerobic digestion system. *Int J Hydrogen Energy*. 2019;44(32):17363-17380.

## Whole-genome analysis of *Staphylococcus aureus* isolates from ready-to-eat food

Mikhaylova Y.\*, Shelenkov A., Chernyshkov A., Kulikova N., Tyumentseva M., Saenko S., Egorova A., Manzeniuk I., Akimkin V.

Central Research Institute of Epidemiology of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Moscow, Russia

\*mihailova@cmd.su

**Key words:** *Staphylococcus aureus*, whole genome sequencing, antibiotic resistance, virulence determinants, MDR, pathogenic potential, ready-to-eat food

**Motivation and Aim:** The urgency of the food safety issue is increasing every year, since it is one of the main risk factors for public health and the gene pool maintenance [1]. *Staphylococcus aureus* is a versatile opportunistic pathogen, which can survive in diverse environments, grow in many types of foods and can cause food poisoning [2]. Thus, it is important to identify the origin of RTE food-related *S. aureus* strains and to evaluate the pathogenic potential of these isolates. Thirty-five isolates were characterized using the whole genome sequencing (WGS) analysis in terms of population structure, the presence of resistance and virulence determinants, as well as plasmid replicon sequences and CRISPR-Cas systems. To the best of our knowledge, this is the first WGS-based surveillance of Russian RTE food-associated *S. aureus* isolates.

**Methods and Algorithms:** During one year period (2019-2020) RTE food samples were collected from cafes and restaurants of different geographic regions in Russia (from Moscow to Kamchatka), corresponding to eight Russian Federal Centers of Hygiene and Epidemiology. All isolates were identified down to a species level by time-of-flight mass spectrometry (MALDI-TOF MS) using the VITEK MS system (bioMérieux, Marcy-l'Étoile, France). The isolates were tested for phenotypic antimicrobial resistance by the boundary concentration method on VITEK MS analyzer. WGS was performed on the Illumina NextSeq platform using Nextera DNA Sample Prep Kit (Illumina, USA). Genome assemblies and comparisons were performed using SPAdes, roary, dnadiff, and custom software.

**Results:** The analyzed 35 *S. aureus* samples belonged to 15 different MLST-based sequence types and to 20 SpaII types including four new SSR1 sequences of SpaII gene and three groups of AGR system (I-III). Worth noting, 11 *S. aureus* samples under investigation were MRSA and isolated from different types of products (bread, salad, meat products and rolls). We observed a large difference in the carriage of virulence and AMR genes between methicillin-resistant (MRSA) and sensitive (MSSA) isolates studied. We also revealed the isolates with high pathogenic potential including one multi-drug resistant MRSA and separate samples harboring different combinations of enterotoxin genes. By analyzing the isolates belonging to the same/single strain, we were able to identify the differences in their accessory genomes marking their dynamics and plasticity.

The analysis of plasmid sequences showed that six isolates from our set harbored multiple types (more than 2) of plasmid replicons simultaneously, while only four

isolates did not have any rep-sequences at all. It is well known that Staphylococci typically carry one or more plasmids per cell and these plasmids include varied gene content [3]. Bioinformatics analysis of CRISPR in foodborne pathogens is crucial for assessing their potential evolution in order to predict the outbreaks, which is of great importance for food safety [4]. In our RTE food *S. aureus* collection, we revealed three isolates bearing CRISPR-Cas systems of IE type. Additionally, we found the genes encoding potential Cas proteins (Cas3\_0\_I) similar to the CRISPR/Cas type I system in the genomes of five isolates.

At a first glance, the isolates from our collection mostly are non-MDR and do not seem to represent a significant danger to public health. On the other hand, however, given the fact that they were isolated from RTE food, the samples and their sources represent a "means of transfer and dissemination" of important pathogenicity determinants. Moreover, the pathogenic potential of separate isolates studied cannot be ignored, especially for those carrying enterotoxin genes. Our WGS results revealed that 14 samples (40 %) carried at least one enterotoxin gene (*sea*, *sec*, *sed* or *sel*). Additionally, major part of the isolates harboring the *tsst1* gene were MRSA and were assigned to CC22 (seven out of 10). Therefore, the presence of these pathogenic factors in *S. aureus* isolates from RTE food can pose a threat to public health. We can also suggest that such pathogenic potential, including an arsenal of mechanisms providing resistance to antimicrobial drugs, leads to the long-time circulation, transmission and dissemination of *S. aureus* in the community and, in particular, within the RTE food supply chains.

**Conclusion:** Therefore, the development of monitoring criteria for MRSA and MSSA, as well as for other selected foodborne bacteria, may provide a valuable option for controlling the dissemination of antimicrobial resistance and other pathogenic factors. Such criteria could be based on microbiological properties or antimicrobial resistance potential of the isolates studied. We believe that the WGS data obtained will greatly facilitate further studies of foodborne *S. aureus* epidemiology, as well as its genome plasticity, in terms of the acquisition of various genetic elements related to host adaptation, antimicrobial resistance and virulence.

**Acknowledgements:** The study was supported by the industry research program of Rospotrebnadzor (2021–2025) "Scientific support of epidemiological monitoring and public health measures of the Russian Federation territory" and Russian Federation Government Resolution No. 3116-r.

## References

1. Gizaw Z. Public health risks related to food safety issues in the food market: a systematic literature review. *Environ Health Prev Med.* 2019;24(1):68.
2. Sergelidis D., Angelidis A. Methicillin-resistant *Staphylococcus aureus*: a controversial food-borne pathogen. *Lett Appl Microbiol.* 2017;64:409-418.
3. Malachowa N., DeLeo F.R. Mobile genetic elements of *Staphylococcus aureus*. *Cell Mol Life Sci.* 2010;67(18):3057-71.
4. Zhao X., Yu Z., Xu Z. Study the Features of 57 Confirmed CRISPR Loci in 38 strains of *Staphylococcus aureus*. *Front Microbiol.* 2018;26(9):1591.

## Elusive recurrent bacterial contamination in a diatom culture: a case study

Morozov A.\*, Galachyants Y., Marchenkov A., Petrova D., Zakharova Y.

*Limnological Institute, SB RAS, Irkutsk, Russia*

\* morozov@lin.irk.ru

**Key words:** diatoms, axenic cultures, contamination, 16S rRNA amplicon sequencing, *Sphingomonas*, *Fragilaria*

*Motivation and Aim:* Development of axenic cultures remains a crucial step for genomic studies of diatom algae. However, the genome our group has produced from supposedly axenic cultures of *Fragilaria radians* [1] was later found to contain several megabases of *Sphingomonas* sp. genetic material, suggesting contamination with this bacterium [2]. Thus, the goal of this work was to identify the source of this contamination, select non-contaminated libraries (if any) and propose ways to identify similar contamination in the future.

*Methods and Algorithms:* To identify the contaminated libraries, all available genomic and transcriptomic reads were mapped to the assembled genome, estimating the number of reads mapping to *Sphingomonas*-derived scaffolds. To study the contamination of the cultures themselves, the available living cultures was stained with DAPI; total DNA was also extracted from both living cultures and older frozen material, and used for 16S rRNA sequencing on Illumina MiSeq. 16S reads (both from amplicon sequencing runs and genomic libraries) were analyzed in Mothur.

*Results:* The original reads have come from two different *F. radians* strains and a number of sequencing platforms (454, Illumina HiSeq, PacBio). Only a subset of 454 libraries, all produced at LIN SB RAS facilities in 2011–2012, had a non-negligible number of reads mapping to *Sphingomonas* scaffolds, suggesting that contamination with bacterial genomic DNA was limited to this series of experiments. Another 454 library, also from 2011, contained a large amount of bacterial 16S rRNA but very few *Sphingomonas*-derived reads. Analysis of reads mapped to 16S reference alignment shows that they mostly come from a short region within 16S rRNA gene, suggesting that this library was independently contaminated with some metabarcoding amplicon at the sequencing facility.

Analysis of 16S rRNA amplicons from modern cultures and frozen material from 2017 suggests that the former are clean, while the latter is yet again contaminated with *Sphingomonas* sp. For the context, a number of clean libraries was produced from the same strain between 2012 and 2017; phylogenetic analysis also shows that two distinct strains were present in 2011–2012 and 2017, implying two independent contamination events. During all this time, no bacteria were detected using light microscopy with DAPI staining.

*Conclusion:* The produced results suggest at least two independent contaminations with *Sphingomonas* sp., most likely during the high-volume biomass production stage. Since light microscopy has failed to detect contamination in either case, we recommend supplementing it with 16S rRNA amplicon sequencing. Although some bacterial OTUs are detected even in the axenic eukaryotic cultures, they are very numerous and have



very low abundance each (typically in single reads or tens of reads at most). In contrast, contaminated samples contain a single high-abundance OTU or a number of closely related OTUs comprising the absolute majority of non-organellar 16S reads.

*Acknowledgements:* This research was funded by Ministry of Science and Higher Education of Russian Federation project 0279–20201–0009 (121032300191–3) “Analysis of the role of baikalian diatoms’ genes and proteins via bionformatics and physico-chemical biology methods”. Authors are grateful to Emmelien Vancaester and their colleagues at Ghent University for identifying the contamination in published genomic data and sharing their results.

### *References*

1. Galachyants Y.P., Zakharova Y.R., Volokitina N.A., Morozov A.A., Likhoshway Y.V., Grachev M.A. De novo transcriptome assembly and analysis of the freshwater araphid diatom *Fragilaria radians*, Lake Baikal. *Sci Data*. 2019;6(1):183. doi: 10.1038/s41597-019-0191-6.
2. Vancaester E., Depuydt T., Osuna-Cruz C.M., Vandepoele K. Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol Biol Evol*. 2020;37(11):3243-3257. doi: 10.1093/molbev/msaa182.

## Development of Phosphate-accumulating microbial community in a sequencing batch reactor (SBR)

Pelevina A.V.<sup>1\*</sup>, Berestovskaya J.J.<sup>1</sup>, Dorofeev A.G.<sup>1</sup>, Ravin N.V.<sup>2</sup>, Mardanov A.V.<sup>2</sup>, Pimenov N.V.<sup>1</sup>

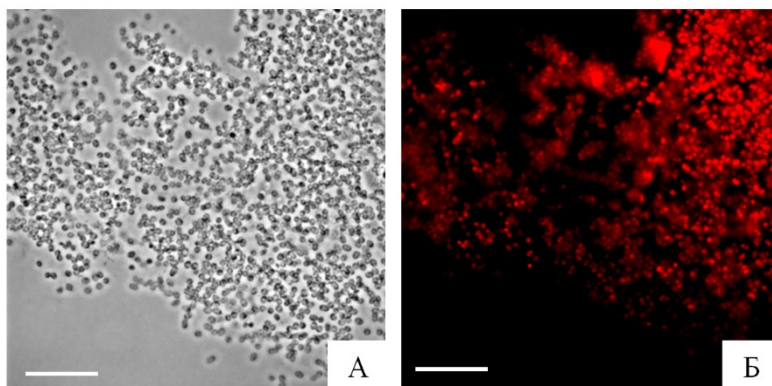
<sup>1</sup> Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Moscow, Russian

<sup>2</sup> Skryabin Institute of Bioengineering, Research Center of Biotechnology, RAS, Moscow, Russian

\*annie.pelevina@yandex.ru

**Key words:** biological phosphorus removal, phosphate accumulating organisms (PAO), wastewater treatment, “*Candidatus Accumulibacter*”

After 150 days of cultivation, a microbial community enriched in phosphate-accumulating organisms (PAO) was obtained in a laboratory bioreactor of sequential-batch action. An analysis of the dynamics of phosphates in the medium and the results of Raman spectroscopy of microbial cells indicated a typical PAO cycle of consumption and release of phosphates during the aerobic and anaerobic periods of cultivation respectively. Analysis of microbial aggregates by FISH revealed a high content of *Ca. Accumulibacter* in it (Fig. 1). *Ca. Accumulibacter* is a typical representative of phosphate-accumulating bacteria.



**Fig. 1.** *Ca. Accumulibacter* as part of the microbial community of aggregates: A – phase contrast, B – cells hybridized with the RAO651 probe (Zeiss 43 light filter). The scale ruler is 20  $\mu$ m

Taxonomic analysis of the studied samples revealed a change in bacterial forms in the microbial community during its development. At the initial stage, the Bacteria domain was represented by four main phyla: *Proteobacteria*, *Bacteroidota*, *Chloroflexota*, and *Patescibacteria*, as well as representatives of archaea, with the dominant phylum *Nanoarchaeota*, which accounted for 5.9 % of the total number of representatives of the microorganism community. During the operation of the bioreactor, the proportion of archaea decreased to zero, and representatives of the phylum *Proteobacteria* began to predominate. Their share by 150 days increased by more than 5 times, reaching 68 %. *Proteobacteria* were mainly represented by the genera *Ca. Accumulibacter*, *Ca.*

Competibacter, *Thiothrix*, *Propionivibrio*, *Dechloromonas*, *Thauera*, *Zoogloea*, which are typical for phosphate removal bioreactors [1–4]. Despite the fact that some representatives of these genera are considered phosphate-accumulating microorganisms (*Ca. Accumulibacter*, *Thiothrix*, *Dechloromonas*, *Thauera*), the dominant representative of this physiological group of bacteria was *Ca. Accumulibacter*, the number of which significantly increased compared to the initial activated sludge microbial community from 0.9 to 13 % of all 16S rRNA gene sequences.

*Conclusion:* Thus, during the operation of the bioreactor, there was a change in the taxonomic composition of activated sludge. By 150 days of cultivation in SBR, a stable functioning phosphate-accumulating microbial community was obtained with a high accumulation of phosphorus in the biomass (16.5 % of the dry weight of the biomass ash-free matter) and accumulation and release of phosphates typical of the PAO phenotype in the aerobic and anaerobic phases of the SBR cycle. The main representative of the PAO in the community was *Ca. accumulibacter*.

*Acknowledgements:* The work was supported by Russian Science Foundation 18-34-00627 and the Ministry of Science and Education of the Russian Federation.

### References

1. Albertsen M., McIlroy S.J., Stokholm-Bjerregaard M., Karst S.M., Nielsen P.H. “Candidatus *Propionivibrio aalborgensis*”: a novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. *Front Microbiol.* 2016;7:1033.
2. McIlroy S.J., Albertsen M., Andresen E.K., Saunders A.M., Kristiansen R., Stokholm-Bjerregaard M., Nielsen P.H. ‘Candidatus *Competibacter*’-lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *ISME J.* 2014;8(3):613-624.
3. McIlroy S.J., Saunders A.M., Albertsen M., Nierychlo M., McIlroy B., Hansen A.A., Nielsen P.H. MiDAS: the field guide to the microbes of activated sludge. *Database.* 2015;2015:1-8.
4. Terashima M., Yama A., Sato M., Yumoto I., Kamagata Y., Kato S. Culture-dependent and - independent identification of polyphosphate-accumulating *Dechloromonas* spp. predominating in a full-scale oxidation ditch wastewater treatment plant. *Microbes Environ.* 2016;31:449-455.

## Microbial anammox communities: structure, physiological and biochemical characteristics, and possibilities of directed modification

Pimenov N.V.<sup>1\*</sup>, Kallistova A.Yu.<sup>1</sup>, Nikolaev Yu.A.<sup>1</sup>, Ravin N.V.<sup>2</sup>, Mardanov A.V.<sup>2</sup>

<sup>1</sup> Winogradsky Institute of Microbiology, FRC of Biotechnology, RAS, Moscow, Russia

<sup>2</sup> Skryabin Institute of Bioengineering, FRC of Biotechnology, RAS, Moscow, Russia

\* npimenov@mail.ru

**Key words:** anammox, microbial community, bioreactor, metagenomics, sewage treatment

*Motivation and Aim:* Anammox is a process of anaerobic oxidation of ammonium with nitrite, resulting in formation of gaseous nitrogen, which is carried out by a multicomponent microbial community, which is based on anammox bacteria belonging to the phylum *Planctomycetes*. According to some estimates, the anammox process is responsible for up to 80 % of the nitrogen turnover in marine ecosystems; it is also the basis of the most efficient biotechnology for nitrogen removal from wastewater. Investigation of these communities by traditional direct microbiological methods is rather difficult, since not only anammox bacteria, but also most other members of the community have not been isolated in pure cultures. The goal of the present work was to study the structure and functioning of the symbiotic anammox community (SAM) for targeted modification of the anammox process and the development of biotechnologies for nitrogen removal from wastewater.

*Results and conclusions:* Two laboratory anammox bioreactors for controlled cultivation of SAM were developed, assembled and launched. In the Anammox-1 sequencing batch bioreactor, the conditions were optimized for growing activated sludge capable of removal of up to 90 % of nitrogen at load of 180 mg N l<sup>-1</sup> day<sup>-1</sup>. The biomass of microorganisms from the Anammox-1 reactor was used to inoculate the Anammox-2 installation, which consisted of three reactors operating in parallel with complete retention of the activated sludge biomass on the non-woven plastic biomass carrier. SAM selection was carried out in the Anammox-2 installation by directed modification of optimal conditions. Successive stages of formation and development of anammox biofilms, from the initial stage of host colonization to the terminal stage of death and lysis, were studied, and the taxonomic composition of SAM was determined. High-throughput sequencing of the 16S rRNA gene fragments revealed predominance of representatives of four phyla in biofilms: *Planctomycetes*, *Chloroflexi*, *Bacteroidetes*, and *Proteobacteria*. Anammox bacteria were represented by species *Ca. "Brocadia fulgida"* and *Ca. "Jettenia asiatica."* Organotrophic members of the phyla *Chloroflexi* and *Bacteroidetes* decompose dead cellular biomass of activated sludge and the exopolymeric compounds formed by anammox bacteria. Nitrifiers of first stage in the community were mainly represented by ammonium-oxidizing bacteria of the genus *Nitrosomonas*; the share of nitrite-oxidizing bacteria was maintained at a low level (< 0.4 %). Analysis of assembled genomes of anammox bacteria revealed the absence of folate biosynthesis pathway, although this compound is important metabolite and it is necessary for carbon dioxide fixation in anammox species. A stimulating effect of folate

on the daily amount of removed nitrogen (dN) was revealed in a community enriched with anammox bacteria of the genus *Ca. "Brocadia"*. Formate was also added to stimulate organotrophic growth and dissimilatory nitrate reduction of anammox bacteria. In the community dominated by anammox bacteria of the genus *Ca. "Jettenia"*, the stimulating effect of formate on dN was revealed. The results obtained support the conclusion that formate and folate can be used to improve the stability and efficiency of the operation of anammox bioreactors for wastewater treatment.

"Synthetic" SAMs were obtained by an enrichment of experimental consortium with stage I nitrifiers and different species of anammox bacteria. Bioaugmentation of the anammox community during reactor start-up increased the efficiency of ammonium removal and led to stable operating mode. However, during subsequent operation of the reactor, exogenous microorganisms were gradually replaced by more competitive autochthonous nitrifiers and anammox bacteria. The introduction of new anammox bacteria at steady state stage of the reactor operation was more effective due to better adaptation of new anammox bacteria in the autochthonous community, and therefore to an increase in the rate of ammonium removal, as well as an increase in the diversity and relative abundance of anammox bacteria in the reactor.

*Acknowledgements:* The study was supported by Russian Science Foundation (project No. 21-64-00019) and Ministry of Science and Higher Education of the Russian Federation.

## Hexadecane-induced stress response of hydrocarbon-oxidizing actinobacterial strain *Tsukamurella tyrosinosolvans* PS2

Romanova V.\*, Grigoryeva T., Laikov A.

*Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

\* *avonamora-94@mail.ru*

**Key words:** stress factors, alkane, quantitative proteomics

*Motivation and Aim:* Bacteria are often encounter to a wide range of stressful conditions in their environment; the most common in nature are nutrient restriction, temperature changes, exposure to various chemical compounds. Under such conditions, microorganisms develop regulatory networks that allow them to overcome adverse conditions. The use of proteomic and transcriptomic methods makes it possible to assess which genes are activated or suppressed in response to exposure to toxic substrates. In the case of biodegradation by alkane-oxidizing bacterial strains, stress may be associated with the chaotropic activity of alkanes. This study was devoted to the effect of hexadecane on the metabolic processes of the *Tsukamurella tyrosinosolvans* PS2 strain using omics approaches.

*Methods and Algorithms:* The *T. tyrosinosolvans* PS2 strain was isolated from petrochemical sludge as a potential degrader of linear alkanes [1]. Targeted quantitative proteomics (LC-MRM/MS) of PS2 strain proteins under different cultivation conditions (sucrose and hexadecane) was performed.

*Results:* A proteomic profile of the PS2 strain with different protein expression was obtained. General, universal stress proteins, as well as cold and heat shock proteins were increased on hexadecane in response to adverse environmental factors. Proteins belonging to the chaperone class had an increased expression on hexadecane, but the HspR protein- a repressor of the synthesis of heat shock proteins, had a reduced expression. Probably, the increased expression of stress proteins in the strain is the result of adaptation and may be an advantage in response to alkanes. In addition, an increased expression on hexadecane of proteins responsible for resistance to antimicrobial agents was found, which increases the tolerance of bacteria to xenobiotic contamination and the ability to biodegradation. The protein systems of resistance to copper, cadmium and arsenic were activated on hexadecane, which makes the strain potentially resistant to heavy metals.

*Conclusion:* Based on the results of quantitative proteomics, hexadecane in the culture medium is a stress factor for the *T. tyrosinosolvans* PS2 strain. Such stress is compensated by changes in the expression of shock proteins and resistance proteins. The observed adaptability of the strain can be useful in the selection of new effective alkane-degrading microorganisms.

*Acknowledgements:* This research was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project No. 0671-2020-0058).

### References

1. Romanova V.A., Boulygina E.A., Siniagina M.N., Markelova M.I., Grigoryeva T.V., Laikov A.V. *Microbiol Resour Announc.* 2019;8:e00218-19.

## Bacteria species diversity on winter grain crops of the field experimental station of RSAU-MTAA named after K.A. Timiryazev

Slovaeva O.<sup>1\*</sup>, Muvingi M.<sup>2</sup>, Yaremko A.<sup>1</sup>, Rubets V.<sup>3</sup>, Igonin V.<sup>3</sup>

<sup>1</sup> All-Russian Plant Quarantine Center, Bykovo, Russia

<sup>2</sup> Peoples' Friendship University of Russia, Moscow, Russia

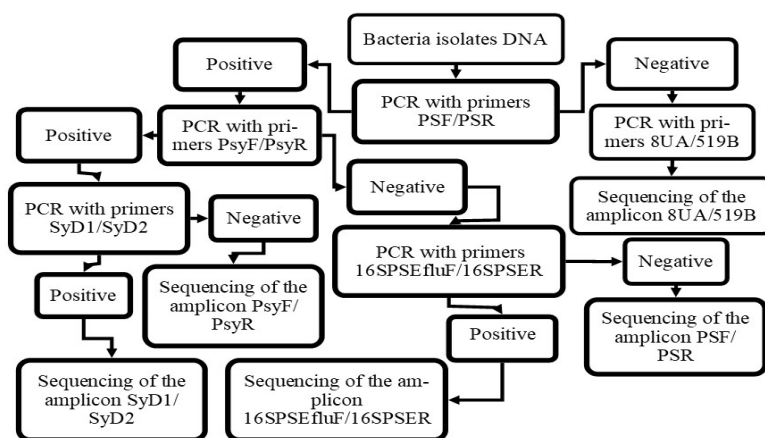
<sup>3</sup> Russian State Agrarian University – Moscow Agricultural Academy named after K.A. Timiryazev, Moscow, Russia

\* slovaeva.olga@gmail.com

**Key words:** diagnostics of phytopathogens, bacterioses of grain crops, PCR, sequencing

**Motivation and Aim:** According to the data of the Federal Customs Service, Russia annually supplies about 17 million tons of grain to countries where grain bacteria pathogens are subject to regulation. The use of species-specific PCR tests is the fastest and most reliable method to diagnose plant pathogens. Diagnostic PCR tests for many grain bacterioses have not been developed or validated. To create a correct algorithm for bioinformatic prediction of a species-specific PCR target, information on the bacterial microbiota species composition of grain crops is required. Both the target bacterial isolates and the accompanying microbiota are needed to validate the designed primers. The purpose of the study is detection and identification of bacteria in cereal crops.

**Methods and Algorithms:** The objects of the study were bacterial isolates from plants of winter wheat, triticale, and rye. The plants were collected in May 2020 on the field experimental station of the Russian State Agrarian University named after K.A. Timiryazev. Bacterial cultures were isolated on CRL medium [4]. Isolates were identified by NCBI alignment of sequences obtained by PCR with primers PSF/PSR [2], PsyF/PsyR, SyD1/SyD2 [1], 16SPSEfluF/16SPSER [3], and 8UA/519B [5]. PCR testing was carried out in accordance with the scheme (Fig. 1).



**Fig. 1.** The scheme for identification of bacterial isolates

**Results:** Bacteria were isolated from 55 different winter grain crop varieties, 47 of which are wheat, 2 ryes, and 6 triticale. The total number of identified isolates is 168. Among them: *Pseudomonas extremorientalis*, *P. trivialis*, *P. poae*, *P. graminis*, *P. synxantha*, *P. cichorii*, *P. chlororaphis*, *P. viridiflava*, *P. fluorescens*, *P. extremorientalis*, *P. frederiksbergensis*, *P. cannabina* pv. *alisalensis*, *P. savastanoi* pv. *savastanoi*, *P. amygdali* pv. *morsprunorum*, *P. syringae* pv. *lapsa*, *P. syr.* pv. *coronafaciens*, *P. syr.* pv. *aptata*, *P. syr.* pv. *atrofaciens*, *P. syr.* pv. *maculicola*, *P. syr.* pv. *cerasicola*, *Erwinia papaya*, *E. billingae*, *Frigoribacterium faeni*, *Clavibacter michiganensis*, *Kineococcus* sp., *Pantoea agglomerans*, *Curtobacterium flaccumfaciens*, *Arthrobacter chlorophenolicus*, *Streptomyces* sp., *Serratia plymuthica*, *Staphylococcus warneri*, *Sanguibacter inulinus*, *S. keddieii*, *Phycococcus* sp., *Cellulomonas* sp., *Oerskovia paurametabola*, *Microbacterium phyllosphaerae*, *M. foliorum*, *Fronidihabitans* sp., *Sphingomonas* sp., *Micrococcus luteus*, *M. yunnanensis*, *Rothia kristinae*, *Bacillus pumilis*, *Rathayibacter festucae*, *Plantibacter* sp., *Actinobacterium* sp., *Rhodococcus cerastii*, *R. yunnanensis*, *R. fascians*, *Agreria* sp., *Lysinimonas* sp., *Subtercola pratensis* and *Pseudoclavibacter helvolus*.

**Conclusion:** The result of detection and identification of bacteria showed that the microbiota diversity of grain crops on the surveyed area is represented by at least 49 species belonging to 29 genera. The vast majority of isolates belong to *Pseudomonas* spp.

### References

1. Guilbaud C. et al. Isolation and identification of *Pseudomonas syringae* facilitated by a PCR targeting the whole *P. syringae* group. *FEMS Microbiol Ecol.* 2016;92:146.
2. Kazempour M.N et al. Isolation and identification of bacterial glum blotch and leaf blight on wheat (*Triticum aestivum* L.) in Iran. *Afr J Biotechnol.* 2009;9(20):2860-2865.
3. Lampugnani C. et al. Quantification of psychrotrophic bacteria and molecular identification of *Pseudomonas fluorescens* in refrigerated raw milk. *Arquivos Instituto Biológico.* 2019;86:e1212018.
4. Slovareva O.Y. Detection and identification of pathogens of bacterial diseases of wheat and barley in the Russian Federation. *MIR J.* 2020;7(1):13-23.
5. Tesic S. et al. Identification of *Pseudomonas cichorii* (Swingle 1925) Stapp 1928 in hydroponic culture of lettuce in Russia. *Vegetable Crops Russia.* 2021;(3):110-115. (In Russ.)



# Microorganisms of low temperature oil field with high-salinity formation water in Tatarstan (Russia) and their potential for MEOR application

Sokolova D.<sup>1\*</sup>, Babich T.<sup>1</sup>, Semenova E.<sup>1</sup>, Bidzhieva S.<sup>1</sup>, Ershov A.<sup>1</sup>, Grouzdev D.<sup>2</sup>, Khisametdinov M.<sup>3</sup>, Nazina T.<sup>1</sup>

<sup>1</sup> Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Moscow, Russia

<sup>2</sup> SciBear OU, Tallinn, Estonia

<sup>3</sup> Tatar Oil Research and Design Institute, Bugulma, Russia

\* sokolovadiyana@gmail.com

**Key words:** microbial community, high-throughput sequencing, 16S rRNA gene, microbial enhanced oil recovery (MEOR)

*Motivation and Aim:* The microbial community of an oil reservoir has a significant impact on the composition of oil and gas and on their displacement from oil-bearing rocks. Investigation of the composition of oil reservoir microbial communities is a necessary step to choosing microbial technologies for enhanced oil recovery (MEOR). The aim of this work was to determine the functional and phylogenetic diversity of microorganisms in the Bavlinskoe oil field in Tatarstan and of the production of oil-displacing metabolites by these microorganisms in order to develop microbial technologies for the enhancement of oil recovery.

*Methods and Algorithms:* Analytical methods were used for analysis of formation water, oil and gas composition; methods of colloid chemistry were used for the detection of surfactants in microbial enrichments; microbiological methods were used for enumeration and cultivation of microorganisms; molecular and bioinformatics methods – for analysis of microbial diversity.

*Results:* Physicochemical conditions, cell numbers and diversity of the microorganisms in the Bobrikovsko-Radaevsky deposits of the Bavlinskoe oil field were determined. Oil reservoirs have a low temperature (16 °C); reservoir water was of the calcium chloride type, characterized by high total salinity (from 109.4 to 254.1 g/l) and slightly acidic conditions (pH 6.0–6.7), the sulfate content in the samples varied from 0.14 to 0.27 g/l, hydrogen sulfide was not detected. By sequencing the V3–V4 region of the 16S rRNA gene in 10 samples of reservoir water, the predominance of bacteria in reservoir water was shown. The dependence of the composition of microbial communities on the geochemical parameters of the habitat (pH, total salinity, content of  $\text{HCO}_3^-$ ,  $\text{SO}_4^{2-}$ ) is shown. Archaeal sequences belonged to methanogens of the genus *Methanohalophilus* of the phylum *Euryarchaeota* (0.2–13.4 % of sequences in libraries), which were most represented in a sample of reservoir water with a salinity of 112.48 g/l. The identified bacteria belonged to the phyla *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Synergistetes*, *Actinobacteria*, and *Thermotogae*. Sequences of uncultivated bacteria of the candidate phylum *Peregrinibacteria* were found in a water sample taken from an observational piezometric well (No. 98). Despite the low content of sulfates in reservoir water, a significant part of the communities in seven samples of reservoir water were deltaproteobacteria of the *Desulfhalobiaceae* family (22.1–83.3 % of sequences in

libraries) containing the halophilic and haloalkalophilic sulfate-reducing bacteria. The second most represented group in these water samples were fermentative bacteria of the genus *Halanaerobium*. Culture methods have shown that the number of fermentative bacteria in production water samples was not higher as  $10^3$  cells/ml. Enrichment cultures of fermentative bacteria obtained from reservoir water were capable of producing significant amounts of volatile fatty acids and alcohols from carbohydrate substrates (starch and sucrose).

*Conclusion:* Formation water of the Bavlinskoye oil field in Tatarstan contain poor, but diverse microbial communities. The activity of the microbial community in the near-bottom zone of injection wells can be activated by the introduction of sugar-containing substrates (starch or molasses) and mineral nitrogen and phosphorus salts in order to form oil-displacing compounds in the oil reservoir. It is necessary to conduct model experiments using isolated microorganisms and cores of oil-bearing rocks to select a suitable microbial technology to increase oil recovery from oil reservoirs. The application of MEOR technologies should be combined with methods of suppressing a large population of sulfidogenic bacteria in the oil reservoir.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant No. 21-64-00019).

# Isolation of new methylophilic species of *Cytobacillus* from deep underground hot spring of Baksan Neutrino Observatory

Tarasov K.<sup>1\*</sup>, Zarubin M.<sup>1</sup>, Yakhnenko A.<sup>1</sup>, Gangapshv A.<sup>2</sup>, Kravchenko E.<sup>1</sup>

<sup>1</sup> *Sektor of Molecular Genetics of the Cell, Dzhelapov Laboratory of Nuclear Problems, JINR, Dubna, Russia*

<sup>2</sup> *Baksan Neutrino Observatory, Institute for Nuclear Research, RAS, Moscow, Russia*

\* *tarasovk49@gmail.com*

**Key words:** methylophilicity, whole-genome sequencing, metabolism

*Motivation and Aim:* Life in hard-reachable corners of the Earth always appeared of great interest due to unique adaptation mechanisms and pathways absent in common “surface” species. Studies of extremophiles already contributed to our understanding the machinery of ribosomes (X-ray structure of ribosome extracted from halophilic bacteria) and provided with such now-ubiquitous enzyme for PCR as Taq-polymerase. Thus while doing another research in Baksan Neutrino Observatory, Kabardino-Balkaria, Russia, we had a chance to take a sample of water from hot highly-mineralized underground spring located at farthestmost tunnel of Observatory. We isolated, sequenced and characterized a strain from that sample. *Methods and Algorithms:* Two metabolic tests API 20 E and API 50 CH by BioMrieux<sup>®</sup> were performed. Genome of strain was sequenced with ONT MinION<sup>®</sup>. Basecalling was performed with Guppy and assembly with Canu. Annotation of genome was performed with two pipelines: NCBI Prokaryotic Genome Annotator and RAST server. Phylogeny inference was performed with GTDB-Tk. During characterization of strain we followed approach proposed by Chistoserdova [1]: that is depicting strain’s metabolism as a sum of modules - each module representing a stage (reaction) for C1-compounds conversion with multiple alternative pathways to perform this stage. Each module depicted by her was thoroughly examined for presence in our strain. All other strain peculiarities were searched for based on subsystem annotation provided by RAST. *Results:* Evidence collected by performing metabolic tests suggested that the strain is unable to utilize the huge part of common multicarbon substances with an exception for glucose, sucrose, maltose, starch. Multiple evidence came from genomic studies (GC-content, distinctive indels, teichoic acid synthetase genes) which suggested that the strain is gram-positive bacteria of genera *Cytobacillus*. Phylogeny inference on GTDB reference tree also mapped this genome into *Cytobacillus* genera with *Cytobacillus oceanisediminis* being species related the most. The strain possesses all necessary genes for methylophilicity with particular modules (formaldehyde oxidation, serine/ribulosemonophosphate assimilation alternatives) being represented with multiple pathways for the same goal. Genes were detected allowing the strain to utilize C1-compounds in dissimilatory way as well as assimilate them for biomass production. Also huge part of genes were located in plasmid of proteins maintaining heavy metal ion balance in strain cells and participating in osmotic regulation. *Conclusion:* Evidence collected from genome sequencing suggests that isolate is Gram-positive facultative methylophilic of genera *Cytobacillus*. Though it cannot utilize methane directly we suggest that other members of microbe community provides nutrition to our strain with C1-compounds. We propose a name for the strain *Cytobacillus oceanisediminis* sp. nov.

## References

1. Chistoserdova L. Modularity of methylophilicity, revisited. *Environ Microbiol.* 2011;13(10):2603-2622.

## Microbial communities in gradient zones of Lake Baikal

Zemskaya T. \*, Bukin S., Chernitsyna S., Ivanov V., Lomakina A., Pavlova O.N., Pogodaeva T., Shubenkova O., Zakharenko A.S.

*Limnological Institute, SB RAS, Irkutsk, Russia*

\* [tzema@lin.irk.ru](mailto:tzema@lin.irk.ru)

**Key words:** microorganisms; metagenomics; shotgun sequencing; genome assembly and binning; metabolic potential

*Motivation and Aim:* Lake Baikal, the oldest lake on the planet (> 25 million years), is of tectonic origin, the bottom sediments of which are about 9 km thick. Distinctive features of Baikal hydrology and hydrochemistry (namely, its great depth, long ice-covered period, low temperatures of surface waters even in the summer, stable temperatures near the bottom, high oxygen concentrations at all depths, and low concentrations of nutrients) make this lake a very peculiar environment for its inhabitants [1], including microbial communities. Microorganisms populating these extreme conditions need to have a flexible metabolism and the enzymatic systems suitable for the specific conditions of Lake Baikal. This study aims to assess the diversity and metabolic potential of prokaryotes inhabiting ecotopes with gradients of oxygen, methane, temperature, light, and sources of carbon and energy.

*Methods and Algorithms:* Water and bottom sediments were sampled based on geological and geochemical data. A comprehensive approach was used for their study, including the analysis of lithological composition of bottom sediments, elemental composition and concentrations of major ions, methane, intensity of fluid flows, DNA analysis of native samples of water column and sediments, functional genes and genomes, culture media with various substrates and energy sources as well as analysis of bacterial and archaeal metabolism.

*Results:* To date, we have obtained a large dataset on the diversity and structure of microbial communities in the water column and bottom sediments of Lake Baikal, taking into account specific geochemical parameters (light, oxygen concentration, methane of different genesis, temperature, concentrations of major ions, elemental composition of pore water, intensity of fluid flows, and sources of carbon and energy), and determined dominant taxa of bacteria and archaea inhabiting different ecotopes. We expect to present the results of the study of microorganisms inhabiting gradients of temperatures, methane and some ions used as electron acceptors. Methane concentration and nitrogen sources determine the diversity and distribution of aerobic methanotrophs, and the rates of methane oxidation are comparable to those observed in other freshwater ecosystems [2]. The results of cultivation and analysis of functional genes and genomes indicated that Baikal methanotrophs differed from known species. Bacteria and archaea involved in anaerobic oxidation of methane (AOM) through metabolic pathways typical of freshwater bodies were present in the sedimentary strata of Lake Baikal. A comparison of the integral rates of aerobic and anaerobic methane oxidation (MO) in the sediments of Lake Baikal revealed that their values were comparable, and in some areas, AOM was much higher than the MO rate in the aerobic zone. Archaea of the family *Methanoperedenacea* and bacteria of the phylum *Methylomirabilota* that carry out AOM

through the denitrifying pathway are diverse among these microorganisms. Analysis of key anaerobic methanotrophy (*mcrA* and *pmoA* genes) indicated a wide diversity of groups of microorganisms involved in these processes in the sediments of Lake Baikal. In particular, we suggested an active role of the members of the class *Bathyarchaeia*, in whose genomes we detected the genes of the methane cycle and which have a wide range of phylotypes in Lake Baikal. Currently, we are conducting laboratory experiments with enrichment cultures of these archaea. Deciphering the genomes and identifying the metabolism characteristics in these microorganisms will significantly elucidate their influence on the composition of discharged gases and their likely participation in the processes of methane oxidation in the bottom sediments of Lake Baikal. In our study, special attention was paid to the assessment of the diversity and spatial distribution of methanogenic archaea in the sediments of Lake Baikal. We detected groups of microorganisms capable of methane generation through four different biochemical pathways. Their ability to produce methane from substrates such as H<sub>2</sub>/CO<sub>2</sub>, acetate and methylated (C<sub>1</sub>) compounds was experimentally shown. We found that hydrogenotrophic methanogens and ‘free-living’ members of the order *Methanomassiliicoccales*, which have the genetic potential to produce methane by reduction of methyl compounds and dihydrogen oxidation, showed the highest relative abundance. Analysis of communities from the aerobic and anaerobic zones of deep sediments and from the coastal zone in gradients of T, O<sub>2</sub> and H<sub>2</sub>S revealed the mass development of colourless sulfur bacteria of two genera, *Thioploca* and *Thiotrix*, represented by a wide range of ecotypes. These bacteria are a basis for the development of a wide range of microorganism species that form a natural biofilter that prevents the influx of hydrogen sulfide, methane and other compounds into near-bottom water from bottom sediments. The area of distribution of bacteria of the genus *Thiotrix* in Lake Baikal is very wide: they form massive fouling in the water of hot springs and in the littoral zone of the lake within the temperature range of +33 to +4 °C and are also part of epibiotic consortiums formed on the surface of the bodies of benthic animals (ostracods and amphipods) living in psychrophilic conditions. Lake Baikal, as a rift lake, is a promising object for studying the deep biosphere. Based on experimental and genomic data, we described for the first time the facultative lithotrophic thermophilic bacteria of the genus *Thermaerobacter* with metabolism atypical for this genus and evolutionarily distant from the type species. The acquired genomic abilities of the investigated strain allow it to function chemolithoautotrophically in deep sediments using inorganic donors and acceptors of electrons with the formation of oxidized compounds of sulfur, carbon and hydrogen necessary for the redox processes carried out by other members of microbial community in the conditions of the low-mineralized bottom sediments of Lake Baikal.

**Conclusion:** The microbiome of Lake Baikal is a rich source of new microbes, enzymes and bioactive compounds, differing from other microbes by the ability to manage their physiology during their survival under psychrophilic conditions with extremely small number of vital elements, sources of carbon (Corg and CH<sub>4</sub>) and energy (SO<sub>4</sub>, NO<sub>3</sub>, Fe<sup>+2</sup>, and Mn<sup>+2</sup>). Studying the characteristics of their metabolism using different approaches can become a key to understanding the role of individual taxa in the ecosystem of Lake Baikal and will provide a more targeted search for strains to use them in the future for biotechnological purposes.

**Acknowledgements:** The study is supported by the Russian Scientific Foundation grant No. 22-14-00084.

## References

1. Kozhov M.M. Biology of Lake Baikal. Moscow: Publ house of USSR Academy of Sciences, 1962.
2. Pimenov N.V., Zakharova E.E., Bryukhanov A.L., Korneeva V.A., Kuznetsov B.B., Tourova T.P., Pogodaeva T.V., Kalmychikov G.V., Zemskaya T.I. Activity and structure of the sulfate-reducing bacterial community in the sediments of the southern part of Lake Baikal. *Microbiology (Moscow)*. 2014;83:180-190.

## 5.3 Section “Industrial biotechnologies: creation of producer strains”



## Constructing robust and efficient PCR-based systems to discriminate *Bacillus cereus* subspecies

Antonets K.S.<sup>1,2\*</sup>, Malovichko Y.V.<sup>1,2</sup>, Merkusheva A.<sup>1,3</sup>, Shikov A.E.<sup>1,2</sup>, Nizhnikov A.A.<sup>1,2</sup>

<sup>1</sup>All-Russia Research Institute for Agricultural Microbiology (ARRIAM), Pushkin, St. Petersburg, Russia

<sup>2</sup>St. Petersburg State University, St. Petersburg, Russia

<sup>3</sup>Saint Petersburg National Research Academic University, St. Petersburg, Russia

\* k.antonets@arriam.ru

**Key words:** *Bacillus cereus*, PCR, systematics

**Motivation and Aim:** The bacterium *Bacillus thuringiensis*, actively used as biocontrol agent against insect pests, is a part of *Bacillus cereus* group. This group now includes a great variety of species from agriculturally important to severely pathogenic among them with unclear systematics. Due to rapid divergence and active horizontal gene transfer it is practically impossible to find “universal” marker to discriminate one species from another. In this work we tried to construct the minimal set of markers which allows to classify strains within *B. cereus* group using PCR.

**Methods and Algorithms:** We constructed the initial dataset using genome assemblies attributed to species and subspecies within *B. cereus* group and available at the NCBI database. The marker set has been constructed by slicing the assemblies into 1 kb fragments and constructing presence/absence table for them.

**Results:** We have developed a workflow that is able to construct a set of markers for classifying bacteria with a certain group based on the representative genomes. The developed workflow has been used to predict such markers for the bacteria in *B. cereus* group. The ability of this set of markers to classify strains within *B. cereus* group has been validated *in silico* and experimentally using qPCR.

**Conclusion:** In spite of the high complexity of genome evolution in *B. cereus* group our approach has been able to find a combination of PCR markers for classifying strains within this group. Using this set of markers for the first time allows high-throughput classification of new strains without expensive whole genome sequencing.

**Acknowledgements:** This study was supported by the Russian Science Foundation (grant No. 20-76-10044).

# Software package for retrosynthesis-based prediction of metabolic pathways

Baboshin M.<sup>1\*</sup>, Shupletsov M.<sup>1,2</sup>, Golubeva L.<sup>1</sup>

<sup>1</sup>Joint Stock Company "Ajinomoto-Genetika Research Institute", Moscow, Russia

<sup>2</sup>Moscow State University, Moscow, Russia

\* Mikhail\_Baboshin@agri.ru

**Key words:** retrosynthesis, bioretrosynthesis, metabolic pathways

*Motivation and Aim:* Finding of metabolic pathways for synthesis of a target compound is a common task in metabolic engineering. Retrosynthesis became a popular approach for seeking pathways in recent years. The method is backward trace from target compound to available metabolites, that models enzyme promiscuity and prediction of hypothetical reactions absent in databases by application of so-called generalized reaction rules [1]. While several retrosynthesis tools were reported [2], the software have shortcomings, such as limited availability, instability, inflexibility, solution loss and proposal of false solutions, lack of appropriate pathway enumeration or thermodynamic analysis. The above shortcomings of current software hinder the practical application of the method. Therefore, more efficient computational tool is required for implementation of retrosynthesis in metabolic engineering. The aims of this work were to develop and test the software package capable to support not only finding of biosynthetic pathway, but their preliminary characterization to simplify the choice of more attractive one, and to create an updated database of generalized reaction rules that takes into account new data on biochemical reactions in living organisms.

*Results:* Given that the database of generalized reaction rules, which is the basis of the retrobiosynthetic approach, depends on completeness of biochemical knowledge, we generated in-house database of reaction rules. Validation of the reaction rules was performed by reproduction of reference reactions (i.e. reactions used for rule generation) and by comparison of global metabolic space with that generated using reaction rules from the database RetroRules [1]. The newly constructed reaction rules reproduce 97 % of their reference reactions. The metabolic spaces generated using the new rules and RetroRules constructed on the same set of reference reactions were close to each other (about 83 % of overlapping), which indicates almost the same quality of reaction rule construction. The entire set of our in-house reaction rules resulted in a larger metabolic space than RetroRules, probably, since the latter were constructed using a less complete data on biochemical reactions.

The software package is organized as a workflow consisting of both in-house components and external software/data and includes following steps: 1) net generation; 2) net processing; 3) pathway enumeration; and 4) pathway ranking.

Net generation means generation of metabolic net containing desired pathways. Restriction of net growth is required to prevent combinatorial explosion. We use constraining by metabolite database as the primary option, but other types of constraints are also possible.

Net processing includes construction of reaction equations, removal of the reactions that do not connect metabolites of the chassis strain with the target compound, and deletion



of inappropriate reactions (stoichiometrically and thermodynamically inconsistent reactions are removed, some other filters may be optionally applied).

Pathway enumeration, i.e. composing of pathways from the list of reactions, is performed by elementary mode analysis. The obtained pathways are added one-by-one to the model of the chassis strain to check possibility of steady-state flux and calculate theoretical yields by flux balance analysis. Pathway ranking is sorting of pathways by fitness criteria to select a pathway for implementation. First, pathways are sorted by number of hypothetical reactions. Then, in each group of pathways with equal number of hypothetical reactions, pathways are sorted by theoretical yield of target product.

To test quality of metabolic net generation, we apply our program to 20 target products (2-amino-1,3-propanediol, 3-methylbutanol, 1,4-butanediol, 2,5-dihydroxybenzoate, benzyl alcohol,  $\beta$ -carotene, muconate, glutarate, lycopene, mesaconate, naringenin, N-methylpyrrolinium, 4-hydroxystyrene, piceatannol, pinocembrin, protopanaxadiol, styrene, terephthalate, vanillin, violacein) with pathways reported in literature that were used earlier for testing of RetroPath2 [2]. We were able to find reference pathway in all 20 cases, compared with 13 out of 20 cases reported for RetroPath2. The result shows that simple constraining of net growth by metabolites may be more efficient than sophisticated constraining methods, such as that implemented in RetroPath2.

All four steps of the workflow were tested on 8 target products. Reference pathway was present in the final list of pathways in all cases, but only in 4 out of 8 cases it matched with the first rank pathway. This, probably, indicates that suboptimal synthetic pathways are often implemented in producer strains.

*Conclusion:* A software package for retrosynthesis-based prediction of metabolic pathways was created; testing on 20 target products confirmed its applicability for finding the pathways with promising properties. This software, together with an updated database of generalized reaction rules, is an efficient tool for *in silico* prediction of metabolic pathways.

### References

1. Duigou T. et al. RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Res.* 2018;47:D1229-D1235.
2. Koch M. et al. Reinforcement learning for bioretrosynthesis. *ACS Synth Biol.* 2020;9:157-168.

## Terahertz radiation proteomic response of the thermophilic bacterium *Geobacillus icigianus*

Bannikova S.<sup>1,2\*</sup>, Khlebodarova T.<sup>1,2</sup>, Meshcheryakova I.<sup>1,2</sup>, Bryanskaya A.<sup>1,2</sup>, Shedko E.<sup>1,2</sup>, Vinokurov N.<sup>3</sup>, Popik V.<sup>3</sup>, Goryachkovskaya T.<sup>1,2</sup>, Peltek S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia*

\* [sbann@bionet.nsc.ru](mailto:sbann@bionet.nsc.ru)

**Key words:** terahertz radiation, *Geobacillus icigianus*, thermophilic microorganism, proteomics

**Motivation and Aim:** Fields of terahertz (THz) radiation utilisation is expanding rapidly, including biomedicine, pharmaceuticals, security systems, and communications, requiring the omics studies of radiation effects on biological systems [1]. Studying proteomic level the responses is one of the most effective methods for external impact results detection. The study on the impact of THz radiation on the temperature-sensitive organisms proteome associated with a number of significant technical difficulties, including the optimal temperature range preservation. With increased resistance of high temperatures of extremophilic species, it is easier to isolate THz radiation effect from thermal one. In this work, we studied the proteomic response to terahertz radiation of the thermophilic bacterium *Geobacillus icigianus*. The bacterium persists under the wide temperature fluctuation conditions with an optimum of 60°C, making its enzyme systems stress resistant.

**Methods and Algorithms:** The experiments were carried out using terahertz free electron laser (FEL) of the Siberian Center for Synchrotron and Terahertz Radiation, designed and employed by the Institute of Nuclear Physics of the Siberian Branch of the Russian Academy of Sciences. Cell culture in LB medium, placed in a specially designed cuvette, were irradiated for 15 min with 0.23 W/cm<sup>2</sup> and 130 µm terahertz radiation. Control sample was incubated using identical cuvette in a thermostat at 60 °C. After, 2D electrophoresis and subsequent mass-spectrometric identification were performed to determine gene expression level alteration and its functional role, based on the results of previously performed whole genome sequencing [2, 3].

**Results:** It was shown, that protein expression level immediate after irradiation and after 10 min incubation following irradiation, varies. THz radiation rapid response involves metabolic systems of electron transport, regulation of transcription and translation, cell growth and chemotaxis, synthesis of peptidoglycan, riboflavin, cofactors NADH, FAD and pyridoxal phosphate, the Krebs cycle, ATP synthesis, chaperone and protease activity, as well as DNA repair, including methylated. When cells are incubated after the irradiation, the systems of cell anti-stress protection, chemotaxis, and cell growth are partially restored, but respiration and energy metabolism, biosynthesis of riboflavin, cofactors, peptidoglycan, as well as components of the translation system are affected, with the amino acid metabolism system being involved.

**Acknowledgements:** This research was funded by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2019-1662 [Laboratory of Molecular Biotechnologies and Kurchatov Genomic Center, federal research center ICG

SB RAS, Novosibirsk]. We also have support by Ministry of Science and Higher Education grant No. FWNR-2022-0022 at the Federal Research Center ICG SB RAS.

### *References*

1. 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.
2. Bryanskaya A.V., Rozanov A.S., Logacheva M.D., Kotenko A.V., Peltek S.E. Draft Genome Sequence of *Geobacillus icigianus* Strain G1w1T Isolated from Hot Springs in the Valley of Geysers, Kamchatka (Russian Federation). *Genome Announc.* 2014;2(5):e01098-14.
3. Kulyashov M., Peltek S.E., Akberdin I.R. A Genome-Scale Metabolic Model of 2,3-Butanediol Production by Thermophilic Bacteria *Geobacillus icigianus*. *Microorganisms.* 2020;8(7):1002. doi: 10.3390/microorganisms8071002.

## Development of optical and ultrasound methods for the multiphase flows investigation and numerical calculations verification in biotechnological process

Kabardin I.K.<sup>1,2\*</sup>, Meledin V.G.<sup>1,2</sup>, Dvoinishnikov S.V.<sup>1,2</sup>, Yavorsky N.I.<sup>1,2</sup>

<sup>1</sup> *Kutateladze Institute of Thermophysics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Optoelectronic Information Technologies, Novosibirsk, Russia*

\* *ivankabardin@gmail.com*

**Key words:** verification of numerical calculations, experiment, laser-doppler anemometry, fiber-optic gas fraction measurements, ultrasound measurements

*Motivation and Aim:* Experimental studies and numerical calculations verification in multiphase flows, including biotechnological processes, is very actual [1]. Both the of optical non disturbance methods development and probe methods development are very important. Institute of optoelectronic information technologies together with the Institute of Thermophysics actively lead the development of such methods for multiphase flows. The aim of this work was to develop methods of laser Doppler anemometry, fiber-optic method for gas fraction diagnosing and ultrasonic methods, as well as probe endoscopic methods for flow investigation in industrial technologies, including biotechnologies.

*Methods and Algorithms:* The first method is the method of laser Doppler anemometry, which makes it possible to measure velocity fields in two-phase flows. The method measure the frequency of the Doppler shift from tracer particles passing through the measurement volume. The method capabilities allows measurement of three components of the velocity vector with an error not exceeding 0.5 %. The second method is a fiber optic method for local gas fraction measurement. The method allows to register the light reflected from the gaseous medium via the optical fiber. The third method is based on the ultrasonic measurement of the Doppler shift from inhomogeneities in a two-phase flow. The error of the latest two methods does not exceed 10 %.

*Results:* The results of the methods application for kinematic flow parameters measurement and gas fraction in various process applications are shown. Both the possibilities of remote application of the methods and the implementation in the probe version are shown. The methods application results made it possible to obtain new fundamental mechanisms and to create new experimental databases for numerical calculations verification.

*Conclusion:* A fundamental scientifically base of new information diagnostic systems for industrial technologies has been created. The results of the work have been implemented at the largest enterprises in the machine-building, nuclear, oil and hydropower industries.

*Acknowledgements:* The work was carried out within the framework of the state contract with the IT SB RAS (121032200034-4 and AAAA-A19-119052190039-8).

### References

1. Meledin V.G. Information optoelectrical diagnostics: Science and industrial technologies. Novosibirsk: Academist, 2015.

## **Individual features of the development of anthrax pathogen strains with different phenotypes on a soil environment model**

Kalinin A.V. \*, Tsygankova O.I., Koteneva E.A., Abramovich A.V.

*Stavropol Antiplague Research Institute of the Rospotrebnadzor, Stavropol, Russia*

\* [jugask@mail.ru](mailto:jugask@mail.ru)

The study of the ability of *Bacillus anthracis* strains with different phenotypic properties to spore germination, reproduction, and sporulation on a medium based on an aqueous soil extract can help assess the role of these processes in the formation and maintenance of soil anthrax foci. We used 12 *B. anthracis* strains, that differ in toxin and capsule formation, proteolytic and lecithinase activity, and also in belonging to the main genetic groups A and B. On a group of strains of anthrax microbes with different plasmid composition and virulence, the possibility of spore germination, reproduction of bacilli, and at least some of them productive spore formation on the soil stead has been revealed. Three scenarios of strain culture development processes were identified: 1 – *B. anthracis* spores remain intact without germinating, 2 – after germination of spores, bacilli are formed, which multiply with different intensity, passing into involutinal forms without spore formation; 3 – the passage of a complete physiological cycle "spore – bacillus – spore". The presence of 2 % blood in the soil environment contributed to the germination of spores and reproduction of the culture, but inhibited the process of sporulation during the observation period of 3 days. No correlation was found between a certain phenotype of the studied strains and *B. anthracis*. The data obtained that only 1–7 % CFU gives rise to the formation of colonies on the soil medium allows us to assume the heterogeneity of the properties of the population of the strains used. Identification of such crops and further detailed study of the complex of their properties will make it possible to identify among them the most significant for the implementation of a complete physiological cycle under conditions simulating soil.

# **Creation of a collection of variants of the *Bacillus anthracis* 1 CO strain with different biological properties to study the adaptation of the anthrax microbe to different living conditions**

Koteneva E.A.\*, Tsygankova O.I., Kalinin A.V.

*Stavropol Antiplague Research Institute of the Rospotrebnadzor, Stavropol, Russia*

\* *postgenom\_stv@mail.ru*

The causative agent of anthrax is characterized by a complex cycle of existence in various morphological and functional forms (spores, vegetative capsular and acapsular), adapted to certain conditions. The short-term phase of existence in the body of sensitive animals is accompanied by active expression of pathogenicity factors and is replaced by spore formation, which ensures the preservation of the microbe in adverse environmental conditions. The study of the adaptation of anthrax microbe strains with different properties to different conditions of existence (macroorganism, soil) will make it possible to identify and characterize, including quantitatively, the key features of strains that affect the manifestation of pathogenic properties. The purpose of the research is the selection of variants of the *B. anthracis* strain, which differ in individual phenotypic properties and genetic characteristics, and their comprehensive study with an assessment of the possibility of using them in experimental work. Materials and methods: we used the original strain *B. anthracis* 1(CO) isolated from an animal with anthrax. The criteria for selecting cultures were their phenotypic properties: cultural and morphological, capsular and toxin formation, enzymatic activity, sensitivity to anthrax bacteriophages, nutritional requirements, *in vitro* virulence. Results: variants were isolated that differ from the original typical strain in capsular formation, which do not show toxin formation on the SOPEK medium, are incapable of spore germination on a minimal nutrient medium, are characterized by weak proteolytic activity, are not able to lyse washed sheep erythrocytes, have lecithinase activity, are resistant to the lytic action of specific bacteriophages, unable to hydrolyze starch and glycogen. Cultures with all variants of the plasmid composition were obtained. Among the studied isogenic variants, four MLVA genotypes and two SNP genotypes were identified. According to the degree of virulence *in vitro*, they belonged to four different groups.

## On the question of activity of oxidative branch of pentose phosphate shunt in *pgl* mutant of *Escherichia coli*

Kovaleva E.S.<sup>1</sup>, Shepelin D.D.<sup>2</sup>, Baboshin M.A.<sup>1</sup>, Golubeva L.I.<sup>1</sup>, Mashko S.V.<sup>1</sup>

<sup>1</sup> *Ajinomoto-Genetika Research Institute, Moscow, Russia*

<sup>2</sup> *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark*

\* *ekaterina\_kovaleva@agri.ru*

**Key words:** <sup>13</sup>C-MFA, oxidative branch of pentose phosphate pathway, 6-phosphogluconolactonase

*Motivation and Aim:* *Escherichia coli* is one of the best investigated organism, which is widely used both in microbiology and for biotechnological production of different biologically active compounds. The special stage in microorganism metabolism investigations started with implementation of <sup>13</sup>C metabolic flux analysis (<sup>13</sup>C-MFA). This approach allows quantitative estimation of actual carbon metabolic fluxes distribution *in vivo*. For example, using <sup>13</sup>C-MFA the role of two transhydrogenases encoded by genes *pntAB* and *udhA* was experimentally determined and characterized [1]. However, there are still some unsolved questions, even about central metabolism of *E. coli*. In particular, there are two opposite points of view about flux distribution in 6-phosphogluconolactonase deficient ( $\Delta pgl$ ) mutant of *E. coli*. For example, applying the flux balance analysis (FBA) to the well known *E. coli* strain BL21, which has *pgl* deficient genotype, investigators assume that oxidative branch of pentose phosphate pathway (oxPPP) is inactive in this strain [2]; this assumption influences on prediction of biotechnological potential of this strain based on genome-scale model. This assumption is based on experimental data of accumulation of product of glucose-6-phosphate dehydrogenase reaction (G6PDH reaction), 6-phosphogluconolactone [3], which is a substrate of 6-phosphogluconolactonase reactions (PGL reaction). In contrast, <sup>13</sup>C-MFA of BL21 [4] and *pgl* deficient MG1655 strain mutant [5], carried out also in our laboratory, has demonstrated only insignificant decrease of flux through oxPPP; the spontaneous hydrolysis of 6-phosphogluconolactone is proposed as the simplest explanation of this phenomenon. In this work we hypothesized that flux through oxPP pathway is actually reduced by near two times, but not to zero, in *pgl* mutant.

*Methods and Algorithms:* For <sup>13</sup>C-MFA of *E. coli* strains grown on minimal medium with 20 %/80 % mixture of [1-<sup>13</sup>C]- and [U-<sup>13</sup>C]-glucose as sole carbon source. Flux calculation and statistical analysis of calculated carbon flux distribution have been provided the OpenFLUX2 software [6] on the basis of labeling data of proteinogenic amino acids and monosaccharides from cellular polymers.

*Results:* Comparative <sup>13</sup>C-MFA of two *E. coli* strains MG1655 and MG1655  $\Delta pgl$  has shown that the flux solution with the best accordance with experimental data (minimal value of sum of squared residuals, SSR) corresponds to carbon flux through the oxPPP of 18 % of input glucose in case of  $\Delta pgl$  mutant in contrast with 25 % in parent strain. The minimal value of SSR is commonly used but conditional criterion. Formally, any flux distribution providing SSR value which passes  $\chi^2$ -criterion can be recognized as acceptable [7]. However flux calculation under assumption that *pgl* inactivation blocks oxPPP led to only statistically unacceptable solution (even minimal SSR value do not pass  $\chi^2$ -criterion). We determined that flux through PGL reaction responded to  $\chi^2$ -criterion on minimal 10 % level. It means that the 2.5-fold decrease in flux through the oxPPP in *pgl* deficient mutant can reflect the real situation.

It is known from literature that inactivation of oxPPP with simultaneous *pntAB* deletion resulted in drastic decrease of growth rate caused by NADPH deficiency. Therefore, we measured growth characteristics for the strains with inactivation of *pntAB* and each of three genes of oxPPP, *zwf*, *pgl*, *gnd*. The  $\Delta zwf \Delta pntAB$  mutant with completely blocked oxPPP, had growth rate which is reduced in 5.9 times relatively to parent strain MG1655. The  $\Delta pgl \Delta pntAB$  and  $\Delta gnd \Delta pntAB$  mutants demonstrated decrease of growth rate in 1.3 and 1.7 times respectively. This additionally confirms that *pgl* inactivation do not abolish activity of oxPPP as in case of *zwf* inactivation. The interesting fact that mutant  $\Delta gnd \Delta pntAB$  grows 1.3 times slower than  $\Delta pgl \Delta pntAB$  mutant despite of activity of Entner-Doudoroff bypass, indicating, probably, that the first strain has more restriction in NADPH supplementation.

Previously it was proposed that the specific bypass through gluconate can function in *pgl* deficient mutant [8]. In this case the flux through *pgl* is zero, but oxPPP is still functioning due to 6-phosphogluconolactone dephosphorylation to gluconolactone followed by its transport out of the cell and hydrolysis to gluconate which is consumed and phosphorylated to 6-phosphogluconate – substrate for 6-phosphogluconate dehydrogenase (GND). Unfortunately, steady-state  $^{13}\text{C}$ -MFA cannot distinguish fluxes through the PGL and this bypass, because both paths possess the same carbon atom rearrangements. Simultaneous inactivation of major gluconate transporters, *gntT*, *gntKU* and *gntP*, reduces growth of  $\Delta pgl$  mutant and growth of  $\Delta pgl \Delta pntAB$  mutant on glucose. These observations indicate that gluconate bypass can participate in keeping of flux through oxPPP in *pgl* deficient strain.

**Conclusions:** Genome-scale metabolic models of bacterial metabolism have a big impact in evaluation of the biotechnological potential of the particular bacterium/strain. Although the whole genome sequence is a necessary basis of these models, a lot of experimental data should be taken into account to make them applicable for reliable prediction. In this work we pay attention to erroneous proposal of inactive oxPPP in *pgl* deficient *E. coli* mutant; this proposal is contradicts to  $^{13}\text{C}$ -MFA of carbon flux distribution in MG1655 $\Delta pgl$  strain and to growth characteristics of double mutants, with inactivated *pntAB* transhydrogenases gene and genes of oxPPP. According to  $^{13}\text{C}$ -MFA the flux through the oxPPP decreases by no more than 2.5-times in the MG1655 $\Delta pgl$  strain.

## References

1. Sauer U. et al. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem.* 2004;279(8):6613-6619.
2. Monk J.M. et al. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Syst.* 2016;3(3):238-251.e12.
3. Meier S. et al. Direct observation of metabolic differences in living *Escherichia coli* strains K-12 and BL21. *Chembiochem.* 2012;13(2):308-310. doi: 10.1002/cbic.201100654.
4. Long C.P. et al. Fast growth phenotype of *E. coli* K-12 from adaptive laboratory evolution does not require intracellular flux rewiring. *Metab Eng.* 2017;44:100-107.
5. Long C.P., Antoniewicz M.R. Metabolic flux responses to deletion of 20 core enzymes reveal flexibility and limits of *E. coli* metabolism. *Metab Eng.* 2019;55:249-257.
6. Shupletsov M.S. et al. OpenFLUX2: ( $^{13}\text{C}$ )-MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb Cell Fact.* 2014;13:152.
7. Zamboni N. et al.  $^{13}\text{C}$ -based metabolic flux analysis. *Nat Protoc.* 2009;4(6):878-892.
8. Kupor S.R., Fraenkel D.G. Glucose metabolism in 6-phosphogluconolactonase mutants of *Escherichia coli*. *J Bio Chem.* 1972;247(6):1904-1910.



## A device reporting cell culture growth in CO<sub>2</sub> incubator via Bluetooth

Litunenکو D.N.\* , Moskalensky A.E.

*Novosibirsk State University, Novosibirsk, Russia*

\* *che14darya@gmail.com*

**Key words:** cell culture, light scattering, portable device

*Motivation and Aim:* Cell cultures are widely used in today's industry and research. Cells are an irreplaceable laboratory material in biomedical research, just as biotechnology cannot be imagined without the use of cell cultures as well.

It is substantial to maintain certain conditions during cultivation, including the control of concentration of cells. For all cells, and especially for bacteria, growth phases can be distinguished. Many existing methods for assessing the state of cell culture require sampling of the material. As cells are cultured under sterile conditions and often inside specialized equipment (for instance, CO<sub>2</sub> incubators), this can be problematic because: 1) the risk of culture contamination increases; 2) the sampling location can influence the result because cells can be unevenly distributed in the flask; 3) sampling leads to a change of (at least) volume, and possibly other parameters of the cell environment 4) this type of measurement cannot be instantaneous. Even if the process is completely automated, there is a period of time between sampling and measurement. Thus, it becomes necessary to monitor the cells without entering the sterile zone.

*Methods and Algorithms:* We present a non-sampling device for monitoring cell culture growth based on light scattering. A horizontal laser beam with a wavelength of 850 nm is directed into the culture flask with cells through the side surface. In this case, the culture flask is placed on a panel with vertical holes located along the laser beam at the same distance from each other. Through these holes, the light scattered by the medium with cells enters the board with photodiodes. The photodiode signals are converted to a digital signal and transmitted via a Bluetooth adapter to an online data store. Cell concentration is estimated from the dependence of the side light scattering signal on the distance along the laser beam.

*Results:* The device was tested on calibration polystyrene beads with size 3.7, 6 and 15 μm. The concentration obtained by the new device coincides with the given concentrations. We plan to demonstrate measurements for cell culture Jurkat soon.

*Conclusion:* This device is easy to reproduce and is quite cheap, while it completely solves the problem of measuring the concentration of cells inside the sterile zone without sampling.

*Acknowledgements:* The study was supported by the grant of the President of the Russian Federation for state support of young Russian scientists MK-4020.2021.1.2, agreement No. 075-15-2021-453.

# The minimal medium irradiated with terahertz radiation induces proteins of homeostasis of transition metal ions and represses proteins of amino acid metabolism when *Escherichia coli* cells are cultivated on it

Meshcheryakova I.<sup>1,2\*</sup>, Goryachkovskaya T.<sup>1,2</sup>, Oschepkov D.<sup>1,2</sup>, Popik V.<sup>3</sup>, Peltek S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia*

\* *irasiris@yandex.ru*

**Key words:** terahertz irradiation, genosensor, *Escherichia coli*

**Motivation and Aim:** It was shown that genosensors cells are activated in response to non-thermal terahertz irradiation, as well as in response to exposure to the irradiated medium [1]. The aim of this work was to study the proteomic response of *E. coli/pKat-gfp*, *E. coli/pCopA-gfp*, and *E.coli/pemrR-gfp* genosensor cells to cultivation in an irradiated medium.

**Methods and Algorithms:** The minimal medium M9 was irradiated with terahertz radiation with a power density of 0.8 W/cm<sup>2</sup> for 15 minutes at a wavelength of 130 μm. The irradiated medium was stored overnight at +4 °C. *E. coli/pKat-gfp*, *E. coli/pCopA-gfp*, and *E. coli/pemrR-gfp* genosensor cells were cultured overnight in LB, then pelleted and resuspended in M9 mixed with an equal volume of irradiated medium, non-irradiated medium, and non-irradiated media with the addition of a genosensor inductor. The non-irradiated medium was also stored overnight at +4 °C. The cells were cultured for 5 hours, after which the protein was isolated and separated by two-dimensional electrophoresis in polyacrylamide gel, the gels were stained with the fluorescent dye Sypro Ruby, and the images of the gels were visualized and analyzed using the PDQuest program. Protein fractions were identified by their peptide mass map by MALDI-TOF mass spectrometry.

**Results:** Comparison of 2D gel electropherogram of sample cultured in irradiated minimal media, non-irradiated media, and media with genosensor inducers revealed differentially expressed protein fractions between samples by at least 1.3 times.

**Conclusion:** Bioinformatic analysis of proteins differentially expressed after cultivation of *E.coli* cells on irradiated minimal medium revealed activation of homeostasis of transition metal ions and inactivation of amino acid biosynthesis in *E. coli* cells.

**Acknowledgements:** This research was funded by the Ministry of Science and Higher Education of the Russian Federation, grant No. FWNR-2022-0022. We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education project No. 075-15-2019-1662 [Kurchatov Genomic Center, federal research center ICG SB RAS, Novosibirsk]) at the Federal Research Center ICG SB RAS. The work was conducted at the Multi-Access Center “Collection of Biotechnological Microorganisms as a Source of Novel Promising Objects for Biotechnology and Bioengineering” of the Federal Research Center ICG SB RAS.

## References

1. Demidova E.V. et al. *Bioelectromagnetics*. 2013;34:15-21.

## The impact of terahertz radiation on living systems

Peltek S.<sup>1,2\*</sup>, Bannikova S.<sup>1,2</sup>, Khlebodarova T.<sup>2</sup>, Shedko E.<sup>1,2</sup>, Meshcheryakova I.<sup>1,2</sup>, Bryanskaya A.<sup>1,2</sup>, Oshchepkov D.<sup>2</sup>, Vasiliev G.<sup>2</sup>, Vasilieva A.<sup>1,2</sup>, Uvarova Y.<sup>1,2</sup>, Kiseleva E.<sup>2</sup>, Popik V.<sup>3</sup>, Goryachkovskaya T.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia

\* peltek@bionet.nsc.ru

**Key words:** terahertz radiation, *E. coli*, *Geobacillus icigianus*, HESc, transcriptomics, proteomics

**Motivation and Aim:** Nowadays terahertz (THz) frequency range is widely used, especially in the fields of security inspection and medical diagnostics. THz radiation is a non-ionizing radiation with a quantum energy of  $\sim 0.01$  eV, corresponding to the vibrational energies of hydrogen bonds that form the supramolecular structures. It was shown that THz radiation transfers biopolymers to gaseous phase as single nanoparticles, retaining their primary structure. Since it is not extensively presented in the natural habitats, the emergence of such an intense source of this radiation requires the study of the possible effects of THz radiation on living organisms.

**Methods and Algorithms:** The Novosibirsk free electron laser generates THz radiation in a range of wavelengths from 50 to 240 microns with the peak of 1 MW powers. Proteomic and transcriptomic analysis were performed on various model organisms including human embryonic stem cells and *E. coli* strains with previously developed genosensors pGlnA-GFP, pMatA-TurboGFP, pSafA-TurboGFP, and pChbB-TurboYFP.

**Results:** Various results on proteomic and transcriptomic responses to THz radiation were obtained. Using MALDI-TOF mass spectrometry it was shown, that radiation near 2.2 THz induce the nondestructive ablation of lysozyme. Bioinformatic analysis indicated that both bacterial and human cells develop a complex response including changes in gene expression and protein production. About 2 % of *E. coli* proteome is affected by THz radiation varying gene expression level of *tdcABCDEFGR*, *matA-F*, *yjjQ*, *dicABCF*, *FtsZ*, *minCDE*, *sfmACDHF*, *yjbEFGH*, and *gfcA* genes, including the induction amyloid biosynthesis through *csg* operon genes. Moreover, THz radiation was shown to cause abnormal morphology of bacterial cells. The study of 2.3 THz irradiation effect on human embryonic stem cells showed no DNA damage, formation of  $\gamma$ H2AX foci, structural chromosomal aberrations, no changes in mitotic index nor morphology, although transcriptional study showed that approximately 1 % of genes expression level was increased, including 15 functional gene classes mostly related to mitochondria.

**Conclusion:** Application of various electromagnetic radiation spectrum range equipment is one of the most critical anthropogenic impact on the environment factors, with also being a source of environmental pollution. The results of the work allow us to state that non-ionizing THz radiation cannot cause mutational events in biological systems, but it affects the metabolic activity and expression of both individual genes and gene cascades. We are currently continuing experiments with thermotolerant bacteria in the hope of understanding the difference in the response of their genome to THz radiation.

*Acknowledgements:* This research was funded by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2019-1662 [Laboratory of Molecular Biotechnologies and Kurchatov Genomic Center, Federal Research Center ICG SB RAS, Novosibirsk]. We also have support by Ministry of Science and Higher Education grant No. FWNR-2022-0022 at the Federal Research Center ICG SB RAS.

### *References*

1. Peltek S.E. et al. Novosibirsk FEL terahertz emission for biological application (plenary). 2nd International THz-Bio Workshop 2011.01.19-20, Seul National University, Seoul, Korea. 2011;37.
2. Bogomazova A., Vassina E., Goryachkovskaya T. et al. No DNA damage response and negligible genome-wide transcriptional changes in human embryonic stem cells exposed to terahertz radiation. *Sci Rep.* 2015;5:7749. doi: 10.1038/srep07749.
3. Serdyukov D.S., Goryachkovskaya T.N., ..., Peltek S.E. 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. Peltek S., Meshcheryakova I., Kiseleva E. et al. *E. coli* aggregation and impaired cell division after terahertz irradiation. *Sci Rep.* 2021;11:20464. doi: 10.1038/s41598-021-99665-3.

## Industrial enzymes production based on genetically engineered microorganisms strains of super-producers

Peltek S.<sup>1,2\*</sup>, Zadorozhny A.<sup>1,2</sup>, Rozanov A.<sup>1,2</sup>, Bogacheva N.<sup>1,2</sup>, Shlyakhtun V.<sup>1,2</sup>, Voskoboev M.<sup>1,2</sup>, Blinov A.<sup>1,2</sup>, Sukhih I.<sup>1,2</sup>, Chesnokov D.<sup>1,2</sup>, Pavlova E.<sup>1,2</sup>, Goryachkovskaya T.<sup>1,2</sup>, Bochkov D.<sup>1,2</sup>, Korzhuk A.<sup>1,2</sup>, Ushakov V.<sup>1,2</sup>, Bannikova S.<sup>1,2</sup>, Meshcheryakova I.<sup>1,2</sup>, Antonov E.<sup>1,2</sup>, Vasilieva A.<sup>1,2</sup>, Shedko E.<sup>1,2</sup>, Shipova A.<sup>1,2</sup>, Bryanskaya A.<sup>1,2</sup>, Uvarova Y.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* peltek@bionet.nsc.ru

**Key words:** enzymes, *Komagataella phaffii*, genetic engineering

**Motivation and Aim:** Enzyme technologies application greatly determines successful development of a diverse national economy sectors. Due to the biotechnological advances many chemical processes in various industry fields, such as textile, leather, food, pulp and paper, fruit and vegetable processing, feed production, and household chemicals, are now being replaced by enzymatic biocatalytic ones. In the Laboratory of molecular biotechnologies of the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, we developed a scientific base of the so called «conveyor system» for industrial microorganism strains of enzyme producers.

**Methods and Algorithms:** The pipeline included creation of a wild type strains collection from previously unexplored unique extreme ecosystems, with further formation of a pure cultures strains sets with certain substrate and bio-catalytic properties. Next, 1860 complete genomes of the collection were fully sequenced and annotated. After, gene expression patterns of the most promising strains were determined, and metabolic activity of industrially significant substrates like lignin or oil sludge was studied. The ability of *Komagataella phaffii* yeast to secrete heterologous proteins in the culture medium was used to create strains producing target enzymatic activities.

**Results:** Using methods of genetic engineering, target enzymes gene transfer into the genome of *K. phaffii* was performed. The effective expression of target proteins in the culture medium was achieved by putting the target gene sequence under the control of the promoter-terminator system of the AOX1 gene, and the leader peptide. Using cross-diffusion filtration and high pressure chromatography preparations highly pure were obtained with subsequent identification by mass spectrometric analysis. To sum up, we developed the technology for preparations of target proteins or enzymes, including membrane methods.

**Conclusion:** Using the methods of modern microbiology, proteomics and bioinformatics, the process of identification, selection and scaled production of target proteins and enzymes for modern industrial purposes, including number of cellulases, alpha-amylases, and lipases, was developed.

**Acknowledgements:** This research was funded by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2019-1662 [Laboratory of

Molecular Biotechnologies and Kurchatov Genomic Center, Federal Research Center ICG SB RAS, Novosibirsk], and as a part of the comprehensive project on high-tech industry “Creation of high-tech production of high-quality plant food proteins” (The agreement on the provision of subsidies from the federal budget for the development of cooperation of a state scientific institution and organization the real sector of the economy in order to implement a comprehensive project for the creation of high-tech industry No. 075-11-2020- 036 from 15.12.2020) in the framework of the Decree of the Government of the Russian Federation of April 9, 2010 No. 218 on the basis of the ICG SB RAS. We also have support by Ministry of Science and Higher Education grant No. FWNR-2022-0022 at the Federal Research Center ICG SB RAS.

### References

1. Beklemishev A.B. et al. Creation of a recombinant *Komagataella phaffii* strain, a producer of proteinase K from *Tritirachium album*. *Vavilov J Genet Breed.* 2021;25(8):882-888. doi: 10.18699/VJ21.102.
2. Rozanov A.S., Korzhuk A.V., Shekhovtsov S.V., Vasiliev G.V., Peltek S.E. Microorganisms of Two Thermal Pools on Kunashir Island, Russia. *Biology (Basel).* 2021;10(9):924. doi: 10.3390/biology10090924.
3. Starostin K.V., Demidov E.A., Ershov N.I., Bryanskaya A.V., Efimov V.M., Shlyakhtun V.N., Peltek S.E. Creation of an Online Platform for Identification of Microorganisms: Peak Picking or Full-Spectrum Analysis. *Front Microbiol.* 2020;11:609033. doi: 10.3389/fmicb.2020.609033.
4. Bryanskaya A.V. et al. Diversity and Metabolism of Microbial Communities in a Hypersaline Lake along a Geochemical Gradient. *Biology (Basel).* 2022;11(4):605. doi: 10.3390/biology11040605.

# Impact of terahertz irradiation on the antimicrobial resistance of *Escherichia coli* JM 103

Shedko E.<sup>1,2</sup>, Uvarova Y.<sup>1,2</sup>, Bannikova S.<sup>1,2</sup>, Goryachkovskaya T.<sup>1,2</sup>, Popik V.<sup>3</sup>, Bryanskaya A.<sup>1,2</sup>, Peltek S.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia*

\* [shedkoed@bionet.nsc.ru](mailto:shedkoed@bionet.nsc.ru)

**Key words:** antimicrobial resistance, terahertz radiation, *Escherichia coli*

**Motivation and Aim:** Nowadays terahertz radiation is widely used in various fields, including application in agriculture and everyday life, especially in medical diagnostics [1]. Also, it is an emerging problem of antimicrobial resistance among both nosocomial and non-nosocomial strains. Nowadays antimicrobial resistance considered the third global cause of death [2]. Previously in our laboratory using RNAseq it was shown, that terahertz radiation has an impact on *E. coli* antimicrobial response genes. Aim of the work was to analyse the impact of terahertz irradiation on the *E. coli* strain, carrying streptomycin resistance gene. **Methods and Algorithms:** Impact of inhibitory and sub-inhibitory concentrations of streptomycin and nalidixic acid were studied. Strain JM 103 with plasmid-mediated streptomycin resistance was grown to the point of log-phase in LB-media. After, cell culture was placed in specially developed cuvette to create 40 µm layer, and irradiated with terahertz laser during 15 minutes. Next, cells were incubated on solid agar at 37 °C. Experiment was carried out with 0, 10, 20, and 30 min incubation in liquid media after the irradiation. On the next morning amount of colonies was counted. **Results:** It was shown that there was an impact of terahertz irradiation on survival of the *E. coli* strain JM 103. With inhibitory concentration of nalidixic acid 400, 800 and 1200 µg/ml no growth was detected, while with concentration of 300 µg/ml bacterial titre corresponded with control sample. For streptomycin with doses of 30, 60 and 90 µg/ml bacterial growth was also shown both in experimental and control samples. **Conclusion:** In our study it was shown that terahertz irradiation has no impact on viability nor plasmid-mediated antimicrobial resistance. Since terahertz radiation is widely used at a present time it is important that it has no impact on antimicrobial resistance levels, especially with the broad application in the field of medical diagnostics.

**Acknowledgements:** This research was funded by the Ministry of Science and Higher Education of the Russian Federation, grant No. FWNR-2022-0022. We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education project No. 075-15-2019-1662 [Laboratory of Molecular Biotechnologies and Kurchatov Genomic Center, federal research center ICG SB RAS, Novosibirsk]) at the Federal Research Center ICG SB RAS. The work was conducted at the Multi-Access Center “Collection of Biotechnological Microorganisms as a Source of Novel Promising Objects for Biotechnology and Bioengineering” of the Federal Research Center ICG SB RAS.

## References

1. KXie J., Ye W., Zhou L., Guo X., Zang X., Chen L., Zhu Y. *Nanomaterials (Basel)*. 2021;11(7):1646.
2. Antimicrobial Resistance Collaborators. *Lancet*. 2022;399(10325):629-655.

## Metabolic engineering of corynebacteria to create a producer of L-valine

Sheremetieva M.E.<sup>1\*</sup>, Khlebodarova T.M.<sup>2,3</sup>, Anufriev K.E.<sup>1</sup>, Ryabchenko L.E.<sup>1</sup>, Leonova T.E.<sup>1</sup>, Kalinina T.I.<sup>1</sup>, Gerasimova T.V.<sup>1</sup>, Rozantseva V.V.<sup>1</sup>, Kolchanov N.A.<sup>2,3</sup>, Yanenko A.S.<sup>1</sup>

<sup>1</sup> National Research Center "Kurchatov Institute", Kurchatov Genomic Center, Moscow, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\*m.e.sheremetieva@gmail.com

**Key words:** *Corynebacterium glutamicum*, L-valine, metabolic engineering, genome editing, whole genome sequencing

L-valine, along with L-leucine and L-isoleucine (valine, leucine, isoleucine), is a proteinogenic branched-chain amino acid (BCAA). As an essential amino acid, valine should be present in the diet of humans and animals: not only is it a building material for proteins, but it also regulates protein and energy metabolism, participates in neurotransmission, and is a donor of an amino group for glutamate biosynthesis.

Valine is used in food industry, medicine and cosmetics, but its main use is as a feed additive. Adding valine to fodder, alone or mixed with other BCAAs, allows for the use of feeds with less crude protein, increases the quality and quantity of meat in pigs and broiler chickens, as well as improves the reproductive functions of farm animals.

The main method of industrial production of valine is biotechnological, the key element of which is strains-producers. Despite the fact that the market for valine is constantly growing, this amino acid is not yet produced in our country. Given the circumstances, the task of organizing local valine production and, consequently, the task of creating industrial producers, are of paramount importance.

One of the most common basic microorganisms for the creation of amino acid producers is the soil bacterium *Corynebacterium glutamicum*. It is well studied from the genetic, biochemical and metabolic points of view, it is safe and easy to grow. Nowadays, producer strains based on *C. glutamicum* are constructed using metabolic engineering, a method based on the analysis of metabolic pathways that allows selection of the most suitable target genes among genes with known functions.

In the present study, the features of corynebacterial BCAA metabolism will be reviewed and strategies for modifying the *C. glutamicum* genome to increase valine production will be proposed. Using several valine-producing strains as examples, we will demonstrate how the level of valine production is affected by modification of the genes responsible for its biosynthesis (*ilvBNCDE*), for the availability of the process precursor (pyruvate) and cofactor (NADPH), for BCAA import and export (*brnQ* and *brnFE* respectively), for the formation of by-products (leucine, isoleucine, D-pantothenate and others).

In addition, the results of the analysis of the complete genomes of valine-producing strains of *C. glutamicum* obtained by nondirectional mutagenesis will be presented. Such an analysis not only makes it possible to identify previously unknown mutations in known genes related to valine production, but also to expand the range of possible target



genes. The results of experimental testing of the effect of new modifications on valine production by *C. glutamicum* strains will also be presented.

The approaches we have tested are of interest not only in terms of creating a corynebacterial valine producer, but may also be applied in the development of producers of other BCAA and related compounds.

*Acknowledgements:* The study is supported by the Ministry of Science and Higher Education of the Russian Federation (075-15-2019-1658).

## Mathematical modeling of gas-liquid flows in bioreactors

Starodumov I.<sup>1\*</sup>, Nizovtseva I.<sup>2</sup>, Svitich V.<sup>3</sup>, Lezhnin S.<sup>4</sup>, Chernushkin D.<sup>3</sup>

<sup>1</sup> Ural Federal University, Ekaterinburg, Russia

<sup>2</sup> Friedrich Schiller University Jena, Jena, Germany

<sup>3</sup> NPO Biosintez, Moscow, Russia

<sup>4</sup> Institute of Thermophysics named after S.S. Kutateladze, SB RAS, Novosibirsk, Russia

\* [ilya.starodumov@urfu.ru](mailto:ilya.starodumov@urfu.ru)

**Key words:** multiphase flow, mass transfer, bioreactor

*Motivation and Aim:* The actual challenge for the researchers is to obtain a relevant model describing liquid and gas turbulent flows [1–4] to develop new bioreactor solutions [5] which tend to maximize gas mass transfer by creating a large gas surface area per unit volume and maintaining this high surface area on the large areas of the total volume of the bioreactor. The methanotrophic microorganisms consume methane as a substrate and breathe dissolved oxygen: ensuring the uniformity of its distribution in volume makes a decisive contribution to the key performance indicators of the system [6]. In order to solve this problem, it is necessary to calculate the predicted values of the gas mass transfer coefficient at various points of a closed contour.

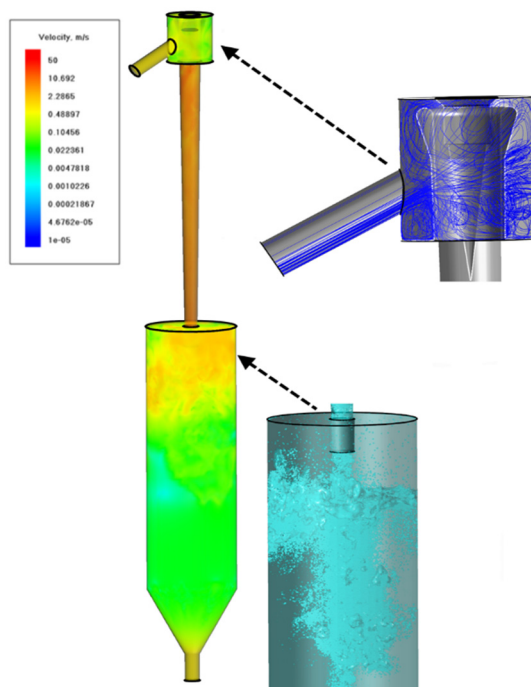
*Methods and Algorithms:* Two models were developed.

Large scale two-phase continuum model, which includes Navier-Stokes equation for liquid and ideal gas, VOF (volume of fluid) model to describe the interfacial interaction and the model of turbulence which is the modification of k- $\epsilon$  model. This model can be used in both low- and high-Reynolds calculations. In first case, the laminar sublayer is resolved by the grid (no wall functions are used), in the second case, the laminar sublayer is not resolved (wall functions are used). The second case, the laminar sublayer is not resolved (wall functions are used). This model satisfactorily predicts the position of the laminar-turbulent transition on the solid surface. The model makes it possible to evaluate the flow intensity modes in different parts of the bioreactor, to identify areas of stagnation, recirculation vortices and zones of high turbulence.

Low scale model of dispersed bubbles transport in liquid continuum, including also Navier-Stokes equation for the continuous liquid phase, convection-diffusion model of dispersed bubble phase transport, convection-diffusion model of mass transfer for the dispersed bubble phase to account for bubble dissolution/growth. Among the main features accounting for gravity, lift and drag forces, bubbles sticking together, their repulsion from the wall should be mentioned. Worth noting that the model has been successfully verified for bubble flows in laminar flows [1].

*Results:* Two models were constructed to describe the process of displacement of the gas and liquid phase in the bioreactor. The large-scale model allows to localize the areas of bubble flows, zones of high turbulence, stagnation/unequal mixing. The low-scale model allows to simulate dispersed bubble flows, the distribution of bubble concentration over the volume, and the contact area of the phases.

Both models allow optimizing the bioreactor design for the target operating modes and evaluating the mass transfer properties.



**Fig. 1.** Liquid velocity (color) and streamlines (blue lines) into bioreactor

**Conclusion:** The highest priority of the simulations was to determine the mass exchange ratio  $K_L = \frac{dc/dt}{a(Cp-C)}$  of the gas fraction, i.e. the intensity of gas dissolution in the liquid. Here  $C(t)$ - dissolved gas concentration,  $Cp$  – equilibrium concentration of dissolved gas,  $a$ -phases contact area. Mass exchange is known to reach its local maximal intense in the areas with bubble flow. Thus, the simulations might be reduced to defining the bubble flow regions (ejector, tank) with model (1) and simulating these regions more detailedly with taking into account model (2) to determine phases contact area  $a$  as well as  $K_L$  prediction.

### References

1. Nizovtseva I.G. et.al. Simulation of two-phase air-liquid flows in a closed bioreactor loop: numerical modeling, experiments and verification. *Math Methods Appl Sci.* 2022.
2. Alekseenko S., Markovich D., Semenov V. Turbulent Structure of a Gas-Liquid Impinging Jet. *Fluid Dynamics.* 2002;37:684-694.
3. Vaidheeswaran A., Hibiki T. Bubble-induced turbulence modeling for vertical bubbly flows. *Int J Heat Mass Transf.* 2017;115:741-752.
4. Timkin L., Gorelik R. Local bubble slip velocity in a downward laminar tube flow. *Thermophys Aeromechanics.* 2020;27(2):259-268.
5. Sauniois M. et al. The growing role of methane in anthropogenic climate change. *Environ Res Lett.* 2016;11:120207.
6. Akberdin I. et al. Methane utilization in *Methylococcus alcaliphilum* 20ZR: a systems approach. *Sci Rep.* 2018;8(1).

## Bioinformatics and experimental approaches in the creation of the drug "Superbact" for parkinsonism's treatment

Stavrovskaya A.V.<sup>1</sup>, Illarioshkin S.N.<sup>1</sup>, Marsova M.V.<sup>2</sup>, Danilenko V.N.<sup>2</sup>

<sup>1</sup> *Research Center of Neurology, Moscow, Russia*

<sup>2</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

\* *alla\_stav@mail.ru*

**Key words:** pharmacobiotics, parkinsonism, lactobacilli, omix technologies, genomic analysis

In recent years, revolutionary changes have taken place in the world in the development and use of pharmacological preparations based on bacteria – inhabitants of the human and animal microbiota and their biologically active ingredients. Pharmabiotics are live biotherapeutic preparations and/or their metabolites and components with established pharmacological ingredients, mechanism of action and aimed at the treatment of specific nosologies. Pharmabiotic preparations are opposed to probiotics, which are used mainly as biologically active supplements and consumed by healthy people.

In this regard, a consensus has formed on the need to study the specific properties of strains and genes that determine the manifestation of the desired effect. In addition to traditional microbiological and biotechnological approaches, a complex of omics technologies – genomic, transcriptomic and proteomic – is used to create pharmacobiotics.

The *Limosilactobacillus fermentum* U-21 strain, due to the high antioxidant activity identified in in vitro and in vivo studies, has shown itself to be a promising candidate for the creation of pharmacobiotic preparations based on it. In particular, the AO activity of its supernatant culture liquid and cell culture was shown. In the model of oxidative stress of the free-living nematode *Caenorhabditis elegans*, the *L. fermentum* U-21 strain increased the median lifespan of the nematode by 25–30 %; in the model of induced parkinsonism in rodents, the introduction of *L. fermentum* U-21 prevented the degradation of brain dopaminergic neurons and pathological changes in internal organs. To identify specific molecular genetic mechanisms that can affect the manifestation of these properties, genomic and comparative genomic analysis was carried out. Bioinformatics search for genes was carried out using the reference catalog of genes of antioxidant proteins found in various types and strains of lactobacilli [<https://github.com/Alexey-Kovtun/Catalog>] created by us on the basis of literary sources and a program for searching for genes/proteins with the desired properties. We also used a comparative analysis of the nucleotide sequences of the studied strain with reference strains and analysis of the nearest environment of the found genes using the Sequence Set Browser. A hypothetical operon was considered to be a set of genes located on the same DNA strand no further than 30 bp. from each other and, in the presence of data on the expression of genes that changed the expression under conditions of oxidative stress in one direction.

The genome of the *L. fermentum* U-21 strain was compared with the *L. fermentum* genomes available in the NCBI database as of February 28, 2022, and with the *L. fermentum* 279 and *L. fermentum* 103 strains that do not exhibit antioxidant properties under experimental conditions. The comparison revealed 138 unique genes, some of

which are extremely rare in this species of lactobacilli and two groups of rare genes, one of which is unique for the strain.

In the genome of the *L. fermentum* U-21 strain, 29 genes were identified that determine proteins with AO properties, namely: genes for thioredoxin complex proteins (*trxA* and *trxC*, *tpx* (*ahpC*), *ahpF*, *nrdH*), which is generally not characteristic of *L. fermentum*; transport protein genes (*cydC*, *cydD*) and glutathione synthesis (*ghsF(AB)*) rare in *L. fermentum*; and also contains a large number of genes for metabolism, transport and binding of heavy metal ions (*copA*, *copB* and others) and metal-dependent transcription factors (including *copR*, *copY/trcY*). The cluster including the MFS transporter and ferrochelatase is found only in 7 genomes of strains of this species deposited in GenBank.

The data obtained are the first step in the identification of genes that determine the properties of the *L. fermentum* U-21 strain. Subsequent studies of the transcriptome and proteome, carried out at the Laboratory of Microbial Genetics of the IOGEN RAS, will provide more information about the genes and their products that determine the unique properties of the *L. fermentum* U-21 strain. Taken together, the obtained data demonstrate the unique AO properties of the strain, allowing it to be positioned as a promising candidate for the development of drugs for the treatment and prevention of various inflammatory diseases, including cardiological, autoimmune diseases, and parkinsonism.

### References

1. Marsova M.V., Abilev S.K., Poluektova E.U., Danilenko V.N. 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.
2. Danilenko V.N., Stavrovskaya A.V., Voronkov D.N., Gushchina A.S., Marsova M.V., Yamshchikova N.G., Ol'shansky A.S., Ivanov M.V., Illarioshkin S.N. The use of a pharmabiotic based on the *Lactobacillus fermentum* U-21 strain to modulate the neurodegenerative process in an experimental model of Parkinson's disease. *Ann Clin Exp Neurol.* 2020;14(1):62-69. doi: 10.25692/ACEN.2020.1.7.
3. Marsova M.V., Poluektova E.U., Odorskaya M.V., Ambaryan A.V., Revishchin A.V., Pavlova G.S., Danilenko V.N. Protective effects of *Lactobacillus fermentum* U-21 against paraquat-induced oxidative stress in *Caenorhabditis elegans* and mouse models. *World J Microbiol Biotechnol.* 2020;36(7):104. doi: 10.1007/s11274-020-02879-2.
4. 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.

## Study of the scaling possibility of the process of obtaining feed protein from natural gas

Tsybmal V.V., Nyunkov P.A., Lalova M.V.\*, Babusenko E.S.

LLC "GIPROBIOSINTEZ", Moscow, Russia

*lalova.m@gibios.ru*

**Key words:** methane-oxidizing bacteria, heterotrophic associates, cultivation/fermentation, feed protein

The carbon substrate is an important aspect in the industrial biotechnology. Thereupon the target product obtaining using relatively cheap non-food raw materials is important. Based on this, microorganisms that are capable of using C1-compounds are of interest, starting from the extremely reduced compound – methane, up to the extremely oxidized compound – carbon dioxide.

The implementation of the technology of obtaining microbial biomass from natural gas as a raw material is a promising, innovative task of biotechnology. The effectiveness of its implementation is determined by the raw material resource, which provides the possibility of creating high-capacity enterprises, as well as the significance of the resulting product – microbial biomass that is characterized by a high protein content (more than 70 %) and high biological value. This will reduce the protein deficiency in animal feed, will ensure an increase in the productivity of fattening farm animals, poultry and fish. The solution to the problem will reduce the import dependence of the Russian Federation on high-protein feed additives (soy, fish, meat and bone meal, etc.).

The efficiency of the technology of obtaining methane-oxidizing bacteria biomass largely depends on the technological potential of the used microorganism culture.

Previously [1, 2] the ability of an associative culture of methane-oxidizing bacteria to develop on natural gas was justified, in the process of cometabolism participate bacteria capable of using as carbon and energy sources the co-oxidation products of methane homologues present in natural gas: ethane, propane and butane, which at a certain concentration inhibit growth of an obligate methane-oxidizing culture.

In the developing of obtaining microbial biomass technology of based on natural gas, in 2016 a new strain of *Methylococcus capsulatus* GBS-15 was selected, which has a complex of industrially valuable properties: the ability to develop at a flow rate up to  $0.36 \text{ h}^{-1}$  on a mineral medium without addition of the growth factors in a wide temperature range of 38–45 °C, pH 5.6–5.9 and in an atmosphere of methane:air (from 1:2.5 to 1:3).

Stable continuous fermentation on natural gas is possible only for an associative methane-oxidizing bacteria culture (with the dominance of the producer of 85–98 %) and heterotrophic satellite bacteria using products of methane homologues incomplete oxidation, and also potentially products of cell autolysis. In fermentation the associative culture the activity of its growth and resistance to stress effects of certain physicochemical parameters noticeably increase.

The isolation of associated bacteria was carried out from the fermentation process carried out in a fermenter  $V = 200 \text{ l}$ , using natural gas with a methane concentration of 96 %, concentration the main homologues of ethane of 2.2 %, propane of 0.8 %, butane of

0.13 %. The fermentation was carried out with a gradually increasing flow rate from 0.1 h<sup>-1</sup> to 0.27 h<sup>-1</sup>. The flow rate was increased after culture growth had stabilized. The studying of the microbial profile using autoselection processes of the established associative culture while the selection of a pure culture of *Methylococcus capsulatus* GBS-15, made it possible to identify 3 heterotrophic satellites. It was found that *Cupriavidus gilardii* predominate among the associates (up to 15 %), and the rest in total no more than 0.5 %. Isolated cultures of methane-oxidizing bacteria and heterotrophic satellites were identified by 16S rRNA and deposited at State Research Institute of Genetics and Selection of Industrial Microorganisms of the National Research Center "Kurchatov Institute". The studies have been carried out in State Research Center for Applied Microbiology and Biotechnology and the report on of non-toxicogenicity, non-pathogenicity and harmlessness of the main producer, heterotrophic associates and the feed product has been obtained. LLC "GIPROBIOSINTEZ" has a technology for obtaining protein feed product, including a full production cycle: starting from sowing an industrial apparatus to obtaining a dry feed product. The cultivation of methane-oxidizing bacteria can be carried out in fermenters of different volumes, with a scaling factor of 10. The fermentation is carried out in reactors with a volume of 5 l, 10 l, 15 l, 200 l and 3 m<sup>3</sup>. Cultivation is carried out in fermenters with a volume of 5 l, 10 l, 15 l, 200 l and 3 m<sup>3</sup>. All fermenters (except 5 and 10 l) operate at positive pressure. The fermentation of an associated culture in a chemostat mode in fermenters with a stirrer with a volume of 5 l and 10 l makes it possible to obtain a productivity of 2.3-2.5 g/l per hour or more. When cultivating in ejection-type fermenters with a volume of 15 l and 200 l, operating at an overpressure of 1.8–2.0 bar, the productivity is about 3 g/l per hour. The increasing the overpressure to 4.0 bar in fermenter with a volume 3 m<sup>3</sup> makes it possible to achieve a productivity of about 4 g/l per hour.

Fermenter volume	Overpressure in fermenter, bar	Mixing device type	Dry weight, g/l	Flow rate D, h <sup>-1</sup>	Productivity, g/l per hour
5 l	–	stirrer	10.0	0.23	2.3
10 l	–	stirrer	8.9	0.28	2.5
15 l	1.8	ejector	11.0	0.26	3.0
200 l	2.0	ejector	10.5–11.2	0.25	2.6–2.8
3 m <sup>3</sup>	2.0–4.0	ejector	14.5–18.0	0.2–0.25	2.8–4.0

The ongoing work in improving the design of the fermenter gives grounds for obtaining higher productivity.

### References

- Galchenko V.F. Methanotrophic Bacteria. Moscow: GEOS, 2001.
- Patent WO 2010/069313 A2/ U-shape and/or nozzle U-loop fermenter and method of fermentation (DK).

## Investigation of promising biotechnological strains from the FRC ICG SB RAS collection using omix technologies and bioinformatics methods

Uvarova Y.<sup>1,2\*</sup>, Bryanskaya A.<sup>1,2</sup>, Shipova A.<sup>1,2</sup>, Vasilieva A.<sup>1,2</sup>,  
Goryachkovskaya T.<sup>1,2</sup>, Shedko E.<sup>1,2</sup>, Peltek S.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* uvarovaye@bionet.nsc.ru

**Key words:** bacteria oil metabolism, KEGG pathway, extremophile microorganism

*Motivation and Aim:* Bacteria are the most common living organisms. The number of studied cultivated bacteria is limited and, particularly, among extremophilic bacteria. These bacteria may utilise various multi-component chemical compounds, but genetic mechanisms of extremophile's metabolism are poorly understood [1]. A unique extremophilic microorganisms collection has been created in the Laboratory of Molecular Biotechnologies of the ICG SB RAS. The aim of the research is an analysis of metabolic activity genetic mechanisms of the extremophilic bacteria from the collection.

*Methods and Results:* Whole genome sequence of 1860 extremophilic bacterial strains from the collection was performed (Illumina MiSeq, USA) and annotated. After, bioinformatic analysis using KEGG PATHWAY was carried out for groups of genes involved in oil sludge metabolism. Next, 263 of all studied strains were selected and grown on the oil sludge. Ten species with the most abundant growth were selected for nanopore sequences and bacterial oxidations chemical oil sludge markers chromatography-mass spectrometry identification. It was shown that bacteria have more than 50 % genes in the following maps: Benzoate degradation [PATH:ko00362]; Aminobenzoate degradation [PATH:ko00627]; Toluene degradation [PATH:ko00623]; Polycyclic aromatic hydrocarbon degradation [PATH:ko00624]; Bisphenol degradation [PATH:ko00363]; Chlorocyclohexane and chlorobenzene degradation [PATH:ko00361]; Naphthalene degradation [PATH:ko00626]; Xylene degradation [PATH:ko00622]; Ethylbenzene degradation [PATH:ko00642]; Caprolactam degradation [PATH:ko00930]; Nitrotoluene degradation [PATH:ko00633]; Fluorobenzoate degradation [PATH:ko00364]; Chloroalkane and chloroalkene degradation [PATH:ko00625].

To reconstruct obtained gene chains, these ten strains were grown on various substrates and analysed using chromatographic analysis. Compounds, related to discreet oil fractions, were selected – vaseline oil, a model mixture of linear n-alkanes, oil sludge, and bitumen. Marker compounds of all found gene chains were found on the chromatograms, with the amount of these compounds varying contingent on the strain. For more complete description of heavy oil fractions degradation genetic mechanisms interactions, it is planned to conduct a transcriptomic analysis for the selected strains.

*Conclusion:* Among 1860 extremophiles from the collection of the ICG SB RAS, strains with the ability of rapid and efficient growth on heavy oil fractions were found. Complete genomes analysis was carried out, showing that studied strains have developed genetic



mechanisms for the various heavy oil fractions utilisation. Chemical markers of the studied genetic chains were identified using chromatographic methods.

*Acknowledgements:* This research was funded by the Ministry of Science and Higher Education of the Russian Federation, grant No. FWNR-2022-0022. We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education project No. 075-15-2019-1662 [Kurchatov Genomic Center, federal research center ICG SB RAS, Novosibirsk]) at the Federal Research Center ICG SB RAS. The work was conducted at the Multi-Access Center “Collection of Biotechnological Microorganisms as a Source of Novel Promising Objects for Biotechnology and Bioengineering” of the Federal Research Center ICG SB RAS.

### *References*

1. Cabral L., Giovanella P., Pellizzer E.P., Teramoto E.H., Kiang C.H., Sette L.D. Microbial communities in petroleum-contaminated sites: Structure and metabolisms. *Chemosphere*. 2022;286:131752.

# **On the search for markers formed in the processes of microbiological metabolism of industrial waste on the example of lignin and oil sludge by gas chromatography-mass spectrometry (GC-MS)**

Vasilieva A.R.<sup>1\*</sup>, Uvarova Y.E.<sup>1,2</sup>, Slynko N.M.<sup>1,2</sup>, Goryachkovskaya T.N.<sup>1,2</sup>, Bryanskaya A.V.<sup>1,2</sup>, Meshcheryakova I.A.<sup>1,2</sup>, Antonov E.V.<sup>1,2</sup>, Tatarova L.E.<sup>1,2</sup>, Lytkin I.N.<sup>1,2</sup>, Shipova A.A.<sup>1,2</sup>, Peltek S.E.<sup>1,2</sup>

<sup>1</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *vasilieva@bionet.nsc.ru*

**Key words:** microbiological metabolism, lignin, oil sludge, gas chromatography-mass spectrometry (GC-MS)

The natural environment is heavily polluted as a result of the extensive use of polyethylene, paints, petroleum products, industrial dyes, toxic chemicals and pesticides, and processed wood products. The indiscriminate use of these chemicals/polymers negatively affects humans and other living systems. One option for reducing the pressure of pollutants on the environment is their metabolic detoxification by microorganisms that use toxic compounds as a source of carbon and energy. In the course of evolution, microorganisms have developed a wide range of enzymes, pathways, and control mechanisms to degrade and recycle xenobiotics. Successful bioremediation requires efficient bacterial strains capable of degrading contaminants to minimal levels in the shortest possible time. Rapid and reliable identification of metabolites produced in the organism-substrate system and an understanding of the mechanisms of metabolic processes are a means of effective microbial selection.

Some of the most common industrial wastes are lignin from wood products and oil contaminants, which are complex hydrocarbon substrates.

In recent years, gas chromatography-mass spectrometry has been widely used for metabolomic studies of living organisms because it is ideal for the analysis of non-polar analytes. This method is highly sensitive and can therefore be used to analyze compounds formed in very small amounts. At the same time, in order to confirm the initiation/progress of oxidative transformation of a wide range of hydrocarbon compounds, it is important to detect substances corresponding to the structural elements of the substrates under study.

The aim of this study was to identify markers formed in the processes of microbiological metabolism of industrial waste on the example of lignin and oil sludge by gas chromatography-mass spectrometry.

When searching for markers of microbiological oxidation of lignin, transformation products of lignin structural elements were considered. Among the numerous compounds in the metabolic mixture, products containing G-fragments of lignin and products of their bacterial metabolism were identified: ferulic acid, guaiacol, vanillin, vanillic acid, succinic acid, malonic acid, oxalic acid; S-fragments of lignin and products

of their bacterial metabolism: sinapic acid, syringal and syringic acid; of the H-fragments of lignin, only p-hydroxybenzaldehyde was identified, indicating that the H-components in this lignin are negligible, or that the metabolic convection rate of these components into non-aromatic compounds is very high. The products identified in the metabolic mixture were found in existing descriptions of the KEGG PATHWAY microbial metabolic pathway maps; the possibility of reactions with these products was confirmed by separate experiments using model compounds that are G-, S-, and H-fragments of lignin as substrates.

Fluorenone and anthraquinone, which are present in the description of the KEGG PATHWAY microbial metabolic pathway maps, were identified and selected as markers of microbial oxidation of oil sludge components. The amounts of these compounds vary depending on the strain used. A possible oxidation product of PAHs is also dibenzfuran, the amount of which increases when microbial cultures are exposed to oil sludge and also varies depending on the strain used.

Thus, the work resulted in the identification and selection of markers that can be used to study the microbiological oxidation of products of industrial waste metabolism, using lignin and oil sludge as examples.

*Acknowledgements:* The study is supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. FWNR-2022-0022.

We thank the Collection of Microorganisms (supported by Ministry of Science and Higher Education project No. 075-15-2019-1662 [Kurchatov Genomic Center, Federal Research Center ICG SB RAS, Novosibirsk]) at the Federal Research Center ICG SB RAS. The work was conducted at the “Multi-Access Center of Proteomics and Metabolomics” of the Federal Research Center ICG SB RAS.

## 5.4 Section

# “Modeling and computer analysis of microbiological systems and processes”



# Reconstruction of a genome-scale metabolic model for chicken whole embryo considering transcriptomics data

Akberdin I.R.<sup>1,2\*</sup>, Kulyashov M.A.<sup>1,2</sup>, Kolmykov S.K.<sup>1</sup>, Pintus S.S.<sup>1</sup>, Yevshin I.S.<sup>1</sup>, Sheng G.<sup>3</sup>, Deviatiiarov R.M.<sup>4</sup>, Stupina A.A.<sup>4</sup>, Gusev O.A.<sup>4,5,6</sup>, Kolpakov F.A.<sup>1,2</sup>

<sup>1</sup> *Sirius University of Science and Technology, Sochi, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Kumamoto University, Kumamoto, Japan*

<sup>4</sup> *Regulatory Genomics Research Center, Kazan Federal University, Kazan, Russia*

<sup>5</sup> *RIKEN Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan*

<sup>6</sup> *Juntendo University, Tokyo, Japan*

\* *akberdinir@biosoft.ru*

**Key words:** genome-scale metabolic model, chicken, skeletal muscle, transcriptomics data

*Motivation and Aim:* One of the rapidly developing modern biological fields, in which advanced technologies are most likely and in-demand developed, is synthetic biology. The main goal of the research field is the creation of new pro- and eukaryotic organs with desired properties based on *de novo* genome synthesis or their deep reorganization using genetic engineering (GE) approaches, bioengineering, systems biology and bioinformatics. Genetic modifications of model lab animals are mainly focused on basic research, while the application of GE technologies in investigations of livestock including chicken plays a key role in the capability to significantly extend existed or create a new niche market. As a perspective, the development of livestock with modified economically useful traits, genetically resistant to hereditary and infectious diseases too is considered. However, a lack of both efficient approaches in the identification of target genes, their regulatory elements determining desired traits and powerful mathematical models enabling quantitative predictions of GE modifications hinders the implementation of this approach in livestock breeding.

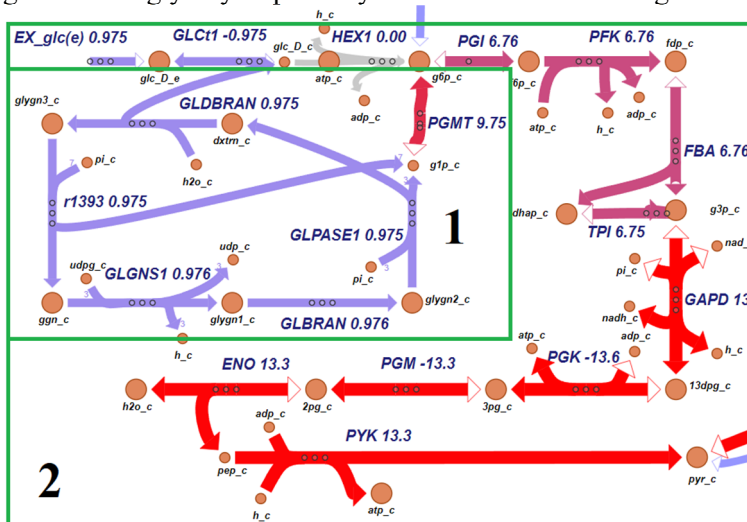
*Methods and Algorithms: Hypometabolic experiments:* As an experimental model, we used three groups: 1) normally developing chicken embryos (38.5C, stages HH1, HH4, HH16 and HH24); 2) embryos in cooling-induced (38.5C→21C) hypometabolic torpor and 3) embryos with re-started normal metabolism (1 hour after 21C→38.5C). Total RNA was extracted from 36 embryos in total (4 embryonic stages x 3-time point x 3 replicates).

*CAGE-seq experimental data and its processing:* CAGE-seq libraries from the total RNA of the embryos were prepared according to the manufacturer's protocol (DNAform, [https://www.dnaform.jp/en/products/library/cage\\_kit/tech/](https://www.dnaform.jp/en/products/library/cage_kit/tech/)). Single-read sequences were analyzed for quality and overrepresented adapter sequences with the theFastQC tool. Quality filtering and adapter trimming were performed with the Trimmomatic tool v0.39. Read mapping on the chicken genome GRCg6a was performed with the STAR local alignment tool. CAGE tag start sites (CTSS) for each sample were imputed using the PromoterPipeline aggregation script from the C1 CAGE protocol. Peak clustering was performed with the Decomposition-based Peak Identification (DPI1) tool. Bidirectional enhancers were identified using the pipeline by Andersson et

al, 2014. Statistical significance of differential expression of TSS and CTSS peaks was calculated using the DeSeq2 tool.

**Genome-scale metabolic model reconstruction:** We utilized the recently published generic genome-scale metabolic model for chicken [1]. Then we fine-tuned the model for chicken whole embryo knowledge and data based on the original transcriptomics data and BioUML platform [2]. The model simulations and analysis was conducted with COBRApy [3] integrated into the platform, whereas Escher web-tool was used to build a metabolic network with fluxes distribution across the map [4].

**Results:** As a result we specified the model for chicken whole embryo metabolic data. We also simulated the metabolic model when glycogen is used as a single carbon substrate for growth, with a growth rate of  $0.34 \text{ h}^{-1}$  on RPMI medium. Flux distributions in gluconeogenesis and glycolysis pathways are demonstrated in Fig. 1.



**Fig. 1.** A genome-scale metabolic network with fluxes distribution across the map predicted by the developed model. The color of arcs in the figure reflects the flux value. The map is constructed in the Escher web-tool [4]. With numbers are indicated two pathways:  
1 – glycolysis pathway, 2 – gluconeogenesis pathway

**Conclusion:** The use of systems biology methods and approaches provides an opportunity to predict the role of various genome modifications on phenotypic traits and their effect on cells. In this regard, the development of genome-scale metabolic models is the most promising approach in the application of GE technologies.

**Acknowledgements:** The study is supported by was supported by Ministry of Science and Higher Education of the Russian Federation grant 075-15-2021-1344.

## References

1. Salehabadi E. et al. 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.
2. Kolpakov F. et al. BioUML - towards a universal research platform. *Nucleic Acids Res*. 2022. In press. doi: 10.1093/nar/gkac286.
3. Ebrahim A. et al. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol*. 2013;7(74). doi: 10.1186/1752-0509-7-74.
4. King Z.A. et al. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comp Biol*. 2015;11(8):e1004321. doi: 10.1371/journal.pcbi.1004321.

# Assembly and annotation pipeline for microorganism genomes from sequence to gene networks

Demenkov P.S.<sup>1,2\*</sup>, Lashin S.A.<sup>1,2</sup>, Ivanisenko V.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *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 promising for biotechnological applications, as well as full sequencing of their genomes. As a result of experimental studies, large amounts of genetic data are generated. Their analysis requires the involvement of modern information and computer technologies, which make it possible, on the basis of genomic information, to reconstruct the gene networks that control the production of target substances.

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 [5]); 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 [2, 4]).

The developed pipeline was used on the data of the collection of microorganisms of the Kurchatov Genomic Center. To date, 1005 genomes have been processed. The average genome length was 6,045,404 bp. The average number of genes in the genome is 5545. KEGG gene networks and metabolic pathways were reconstructed for all analyzed genomes.

*Acknowledgements:* Work 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.

## References

1. Bolger A.M., Lohse M., Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
2. Demenkov P.S. et al. 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.
3. Gurevich A. et al. QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075.
4. Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21:228.
5. Nurk S. et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol.* 2013;20(10):714-737.

# Diversity and distribution of restriction-modification systems in natural microbial multi-species community of Antarctic Deep Lake

Karyagina A.<sup>1,2,3\*</sup>, Rusinov I.<sup>1</sup>, Ershova A.<sup>1,4</sup>, Alexeevski A.<sup>1,5</sup>, Spirin S.<sup>1,5,6</sup>

<sup>1</sup> Belozersky Institute, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia

<sup>3</sup> All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

<sup>4</sup> Moyne Institute of Preventive Medicine, Trinity College Dublin, Dublin, Ireland

<sup>5</sup> Institute of System Studies, Moscow, Russia

<sup>6</sup> National Research University Higher School of Economics, Moscow, Russia

\* akaryagina@gmail.com

**Key words:** restriction-modification system, Antarctic Deep Lake, horizontal transfer of genes

*Motivation and Aim:* The distribution of Restriction-Modification (R-M) systems, which are prokaryotic defense systems, in natural multi-species communities has not been yet studied. Thus, distribution of the R-M system genes in stable and closed microbial community of hypersaline Antarctic Deep Lake [1] is of particular interest.

*Methods and Algorithms:* We performed a bioinformatic analysis of R-M system genes in five metagenomes and in complete genomes of six strains of four archaeal species from Deep Lake. The analysis is based on homology search of metagenomic contigs and genomic sequences over REBASE proteins and on considering R-M system related Pfam domains.

*Results:* We predicted about 5000 R-M system genes, which can be grouped into 1400 clusters of homologous genes (with > 50 % identity of encoded proteins) and 2300 clusters of nearly identical genes (with > 98 % identity). Only 97 clusters of nearly identical genes are represented in the genomes of the four archaeal species. The lists of R-M system genes found in the complete genomes is significantly different from the lists of genes annotated in REBASE. There are more than 60 putative R-M system genes in these genomes not included in REBASE while REBASE contains 18 genes from the four archaeal genomes that do not meet our criteria. For one of the studied species, *Halorubrum lacusprofundi*, we demonstrate high inter-strain heterogeneity of R-M system composition. Most R-M system genes common for different species are located within rather large highly identical regions. We found that some R-M system genes in complete genomes have metagenomic read coverage significantly different in comparison with housekeeping genes of the same species.

*Conclusions:* The results for archaeal complete genomes of most abundant species show significant differences with lists of genes annotated in REBASE for the same genomes, which may be explained by difference in the methods of R-M genes identification. The use of an additional criterion, namely the presence of R-M system-related Pfam domains, can make the search for genes of the R-M systems more reliable compared to only homology search vs. REBASE. Some obvious mistakes in REBASE were identified. The microbial community of Antarctic Deep Lake possesses a lot of diverse R-M systems. Only a small part of their variety is encoded in genomes of the four archaeal



species. A significant difference in R-M system related genes between strains for *H. lacusprofundi* is probably a result of an intense gene exchange. Another sign of an active gain and loss of R-M systems is a huge difference in metagenomic read coverage between R-M system genes within one genome, which can be interpreted as a presence of strains of the same species but with different R-M system repertoires.

*Acknowledgements:* The work was supported by the Russian Science Foundation grant 21-14-00135.

### *References*

1. DeMaere M.Z. et al. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc Natl Acad Sci U S A.* 2013;110(42):16939-16944.

# Finding the ways for potential increasing of L-valine biosynthesis yield in *C. glutamicum* with mathematical modeling

Kazantsev F.<sup>1,2</sup>, Klimenko A.<sup>1,2</sup>, Kropochev A.<sup>3</sup>, Trofimova M.<sup>2</sup>, Lashin S.<sup>1,2</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* kazfdr@bionet.nsc.ru

**Key words:** flux balance analysis, mathematical modeling, L-valine

**Motivation and Aim:** Nowadays the industrial production of amino acids is mostly based on microbiological approaches. The main request of this industry is maximization of the products' yield and minimization of input resources and energy costs. The solution is to develop a new bacterial strain or to optimize metabolic pathways of the existing one. The flux balance analysis allows researchers to assess the potential of various metabolic pathway modifications and suggest the best ones. In this work, we are focused on finding the ways of improving *C. glutamicum* ACTT 14067 strain in L-valine production.

**Methods and Algorithms:** We have used newly published whole-genome model (iCGB21FR) [1] as the basis model. The model was built on the strain *C. glutamicum* ATCC 13032. We have used Python library CobraPy (<https://cobrapy.readthedocs.io/>) as a computational tool and Escher ([escher.readthedocs.io/](https://escher.readthedocs.io/)) for the visualization of results and subsequent analysis. We have fitted the model using the data on Russian National Collection of Industrial Microorganisms strain *C. glutamicum* ACTT 14067 (B-5212) (<https://vkpm.genetika.ru/katalog-mikroorganizmov/show21916/>).

**Results:** We adapted the recently published iCGB21FR full-genome model, which is based on the *C. glutamicum* strain ATCC 13032 - the parental strain for *C. glutamicum* ACTT 14067. A comparative analysis of the genomes of these two strains was performed. The analysis showed a difference in enzyme set including those involved in glycolysis, pentose phosphate pathway, tricarboxylic acid cycle, pantothenate and branched-chain amino acid biosynthesis pathways. The enzymes that had been knocked out were excluded from the original model. A series of computational experiments were performed with different environmental conditions, different carbon sources in order to obtain potential estimates of the productivity of the strain and its difference from the parental strain. To analyze the processes involved, the initial metabolic maps were extended to account for valine biosynthesis and additional carbon sources. A number of model scenarios were calculated and valine yields under aerobic and anaerobic conditions were shown. Varying the target function between biomass growth and valine yield produced upper and lower estimates for valine yield under different conditions. For each scenario, information on strain growth rate, valine yield, and the distribution of carbon across different products, including valine, is presented. The adapted model can be used for assessing the potential valine yield for particular cultivation medium conditions.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

## References

1. Feierabend M. et al. High-Quality Genome-Scale Reconstruction of *Corynebacterium glutamicum* ATCC 13032. *Front Microbiol.* 2021;12:1-18. doi:10.3389/fmicb.2021.750206.

# Finite element modeling of physical processes in multilayer spheroids based on Functional Representations

Kirillov B.\*, Vilinski-Mazur K., Pasko A., Akhatov I.

*Skolkovo Institute of Science and Technology, Moscow, Russia*

\* *Bogdan.Kirillov@skoltech.ru*

**Key words:** mathematical modeling, function representations, tissue spheroids, cellular organoids, 3D bioprinting, viability simulations, implicit functions

*Motivation and Aim:* The current study is motivated by the need to study biological objects that have heterogeneous structures. Such objects emerge in various fields of modern biotechnology, but the current study focuses on objects that come to existence via bioprinting and organoid cultivation. 3D bioprinting – a branch of 3D printing that uses cell-infused hydrogel instead of plastic – is based on tissue spheroids [1], three-dimensional cell cultures, fusing together following the tension forces. Tissue spheroid is similar to an organoid [2] – three-dimensional structure grown from stem cell culture in vitro. Viability of a tissue spheroid and an organoid depends on the same physical process – diffusion of oxygen, nutrients and other important chemicals. Studying the physical processes within a spheroid and an organoid requires non-trivial mathematical apparatus due to the complexity of spheroid/organoid shapes.

*Methods and Algorithms:* We use functional representations (FRep [3]) to provide the geometry of the multilayer spheroid – a three-dimensional cell culture that possesses a layered inner structure. A similar geometry is constructed for an organoid. The geometry is used as a mesh for the Finite Element Method computations (with the use of the Firedrake [4] framework).

*Results:* We propose an approach to model the diffusion processes within a spheroid and an organoid. The method unites numerical modeling with finite elements for solving time-dependent differential equations for diffusion and deformation and geometrical modeling with implicit functions within the Functional Representations (FRep) framework. We perform the simulations for two different kinds of chemicals that could be diffused within an object of interest – oxygen and a fluorescent dye (calcein).

*Conclusion:* The results of our study shows that the proposed method provides for accurate evaluation of the fluorescent calcein concentration field that complies with experimental results [5]. Our method also allows for analysis of cell viability and predictions for emergence of anoxic regions in case of analyzing oxygen diffusion.

## References

1. <https://phys.org/news/2020-07-bioink-cell-bioprinting-d.html?deviceType=mobile>.
2. Takanori T., Wells J.M. Organoids by design. *Science*. 2019;364(6444):956-959.
3. Pasko A. et al. Function representation in geometric modeling: concepts, implementation and applications. *Visual Computer*. 1995;11(8):429-446.
4. Rathgeber F. et al. Firedrake: automating the finite element method by composing abstractions. *ACM Trans Mathematical Soft (TOMS)*. 2016;43(3):1-27.
5. Achilli T.-M. et al. Multilayer spheroids to quantify drug uptake and diffusion in 3D. *Molecular Pharmaceutics*. 2014;11(7):2071-2081.

# Integration of transcriptomic, genomic, and proteomic data in order to assess the impact of translation efficiency features on protein abundance for various prokaryotes

Korenskaia A.E.<sup>1,2\*</sup>, Matushkin Yu.G.<sup>3</sup>, Lashin S.A.<sup>1,2,3</sup>, Klimenko A.I.<sup>1,3</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* korenskaia@bionet.nsc.ru

**Key words:** translation efficiency, protein abundance prediction, codon usage

*Motivation and Aim:* Due to fact that the protein abundance controls many physiological processes, the calculation of protein abundance is of great importance for medical and biological studies. The protein abundance is determined by the regulation of the gene expression on the transcription level, the efficiency of the translation stage, and the post-translation level regulation. It is known that the level of transcript expression impacts protein abundance the most. The correlation coefficient between the expression at the transcriptomic and the proteomic level approaches to 0.5, which has been shown for *Escherichia coli* [3] and *Desulfovibrio vulgaris* [4]. At the same time, translation efficiency tends to reflect gene expression at the basal level, as a result, it could highly correlate with the protein abundance – more than 0.5 for several prokaryotes. Hence, the combination of the transcriptomic data and the translation efficiency features, calculated based on the sequence, are supposed to effectively predict the protein abundance. According to the studies [2], translation efficiency depends on the translation initiation efficiency, which is affected by the mRNA structure near the start codon, and translation elongation efficiency, which impact by codon bias on mRNA, and secondary structures on mRNA. It should be noted that for various prokaryotes these features impact on the protein abundance differently. The goal of this study is to assess the impact of those factors on a protein abundance, considering the gene expression at the protein and transcriptome levels, as well as to build a set of regression models to predict protein abundance in various prokaryotes.

*Methods and Algorithms:* The transcriptomic (RNA-seq) and proteomic (obtained by LC-MS/MS spectroscopy) data were collected from the studies, where these quantifications were carried out under the same conditions. The data gathered for prokaryotes representing various phylogenetic groups. For the provided genomes the translation efficiency features for each protein-coding gene have been calculated. The translation initiation efficiency features are calculated using the ostir [5] tool, the translation elongation efficiency features are obtained by the EloE tool [6] and the BCAWT tool [1]. A number of regression models were built, in which a set of proteins' expression levels acts as a dependent variable, while a set of transcripts' expression levels as well as a translation efficiency parameters act as independent variables for each of the organisms under study. To assess the impact of each parameter, a partial regression coefficient was calculated for each of these features. In addition, Pearson correlation coefficients between a parameter value and expression data were calculated. The

obtained coefficients were compared between the translation efficiency parameters of an organism, as well as between the prokaryotes belonging to various phylogenetic groups. *Results and Conclusion:* It has been demonstrated that both correlation coefficients and partial regression coefficients between translation efficiency features and protein abundance show diversity for various prokaryotes. For example, for *Leptospirillum ferriphilum* it has been shown that a feature, related to the number of secondary structures on mRNA impacts protein abundance the most. Notably, in different conditions correlation between expression at the level of mRNA and at the protein level is diverse (0.55 and 0.63), but adding the discussed parameter increases the correlation coefficient by 0.05. In conclusion, the usage of a combination of data on mRNA abundance and its secondary structure allows one to improve the protein abundance prediction power for certain prokaryotes. However, for some organisms other parameters are more significant, in particular, the characteristics of the codon usage. Thus, for effective protein prediction it is necessary to take into account different factors of translation efficiency for various prokaryotes.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### References

1. Anwar A. BCAWT: Automated tool for codon usage bias analysis for molecular evolution. *J Open Source Softw.* 2019;4(42):1500. doi: 10.21105/joss.01500.
2. Chiaruttini C., Guillier M. On the role of mRNA secondary structure in bacterial translation. *Wiley Interdiscip Rev RNA.* 2019;11(3):1-21. doi: 10.1002/wrna.1579.
3. Guimaraes J.C., Rocha M., Arkin A.P. Transcript level and sequence determinants of protein abundance and noise in *Escherichia coli*. *Nucleic Acids Res.* 2014;42(8):4791-4799. doi: 10.1093/nar/gku126.
4. Nie L., Wu G., Zhang W. Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: A multiple regression to identify sources of variations. *Biochem Biophys Res Commun.* 2006;339(2):603-610. doi: 10.1016/j.bbrc.2005.11.055.
5. Roots C.T., Lukasiewicz A., Barrick J.E. OSTIR: open source translation initiation rate prediction. *J Open Source Softw.* 2021;6(64):3362. doi: 10.21105/joss.03362
6. 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.

# Bioinformatic assessment of factors affecting the correlation between protein abundance and elongation efficiency in prokaryotes

Korenskaia A.E.<sup>1,2\*</sup>, Matushkin Yu.G.<sup>3</sup>, Lashin S.A.<sup>1,2</sup>, Klimenko A.I.<sup>1</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* korenskaia@bionet.nsc.ru

**Key words:** translation elongation efficiency, protein abundance prediction

*Motivation and Aim:* As proteins are the key elements that provide cell function, developing the protein abundance prediction methods based on a gene sequence is the subject of scientific interest. One of the steps of protein abundance prediction is the evaluation of translation elongation efficiency based on mRNA sequence. The efficiency of translation elongation depends on the codon bias, amount and energy of secondary structures in mRNA. EloE (Elongation Efficiency) [1] tool uses these parameters in several combinations classifying a genome under analysis into one of the five elongation efficiency optimization types and to sort protein-coding genes according to the predicted efficiency of the translation elongation stage estimated by the corresponding Elongation Efficiency Index (EEI). The goal of this study is to evaluate the correlation coefficients between experimentally measured protein abundance and predicted elongation efficiency calculated by EloE for prokaryotes belonging to a wide range of taxonomic groups that have not been evaluated before.

*Methods and Algorithms:* The Spearman correlation coefficients between EEI indices and protein abundance, which was obtained by LC-MS/MS spectroscopy, were calculated for 26 prokaryotes. The dependence of this parameter on other factors, such as EEI type, minimal doubling time, mean (M) and standard deviation (R) EEI for ribosomal protein-coding genes, was estimated by the Pearson's correlation coefficient test with bootstrapping.

*Results and Conclusion:* It has been found that the obtained correlation values are not high and vary greatly among the organisms under study. Several factors associated with the correlation level have been revealed. The first factor is the elongation efficiency type – the organisms that belong to the translation elongation efficiency type that is taking into account codon bias only, have significantly higher correlation coefficients. The second factor is taxonomical identity – bacteria that belong to the most studied classes (Bacilli and Gammaproteobacteria) tend to have higher correlation coefficients. The third one is a growth rate – the organisms with higher correlation between EEI and protein abundance show lower minimal doubling time. Finally, we have demonstrated that the initial approach used by EloE that relies on assessing the ranks of ribosomal proteins in the gene list sorted by the base EEI value is adequate to the experimental data of the organisms under study and can be used in further developments of the algorithms. The obtained results are useful for further improvement of protein abundance prediction

methods including those that would take into account not only translation elongation but also other stages that affect the level of gene expression.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### *References*

1. Sokolov V., Zuraev B., Lashin S., Matushkin Y. Web application for automatic prediction of gene translation elongation efficiency. *J Integrative Bioinformatics*. 2015;12(1):256.

# Deciphering a Cu-dependent metabolism of the *Methylovulum alcaliphilum* 20Z<sup>R</sup> methanotroph by integration of transcriptomics data into a genome-scale metabolic model

Kulyashov M.<sup>1\*</sup>, Kolmykov S.<sup>1,2</sup>, Kalyuzhnaya M.<sup>3</sup>, Akberdin I.<sup>1,2</sup>

<sup>1</sup> BIOSOFT.RU, LLC, Novosibirsk, Russia

<sup>2</sup> Sirius University of Science and Technology, Sochi, Russia

<sup>3</sup> San Diego California State University, San Diego, USA

\* m.kulyashov@mail.ru

**Key words:** Cu-dependent metabolism, genome-scale metabolic model, methanotroph

*Motivation and Aim:* Aerobic methane-oxidizing bacteria or methanotrophs have the unique ability to grow on methane as their sole source of carbon and energy. As a consequence, methanotrophs are ubiquitous in the environment and play a major role in the removal of the greenhouse gas methane from the biosphere [1, 2]. Despite the fact that the main metabolic steps of methane utilization by microorganisms have been identified and well studied to date, a detailed understanding of molecular genetic mechanisms that provide an adaptive response at the level of transcription regulation to various growth conditions, high and low pH, temperature, and salinity is still elusive.

*Methods and Algorithms: Transcriptomics data analysis*

To analyze original transcriptomic data on the growth of *M. alcaliphilum* 20Z<sup>R</sup> under various cultivation conditions, obtained as part of a collaboration with Prof. Marina Kalyuzhnaya (San Diego California State University, USA). The modified read sets were aligned to the *M. alcaliphilum* 20Z<sup>R</sup> reference genome available in the NCBI database (ASM96853v1) using the Bowtie2 algorithm [3]. To quantify gene expression based on reads aligned to the reference genome, the featureCounts package was used [4]. At this stage, we used a genomic annotation built within the MicroScope project (MAGE\_000000532.1)[5]. The normalization of the obtained data and the subsequent search for differentially expressed genes were performed using the edgeR package [6].

*Development of the genome-scale metabolic model*

We also integrated published transcriptome data into the previously developed iIA407 metabolic flow model [7]: gene expression data (FPKM values) under 4 culture conditions: 1) growth on methane in the presence of calcium and copper; 2) growth at low methane concentration in the presence of calcium and copper; 3) growth on methanol in the presence of calcium without copper; 4) growth on methanol in the presence of La [8] were integrated into a mathematical model using the E-Flux2 algorithm [9]. To reproduce the cultivation conditions for 20Z<sup>R</sup> growth in the absence of Cu, the mathematical model was modified using the CobraPy software [10]. After integrating the transcriptome data into the model, analysis of the flow balance with biomass maximization was performed using the OptFlux program [11] for each of the cultivation conditions.

*Results:* At the stage of analysis of transcriptomic data sets for 20Z<sup>R</sup>, it was possible to identify biologically significant changes in expression patterns for the ectoine operon



upon varying the osmolarity of the cultivation medium and for mxa-methanol dehydrogenase as a result of the inclusion or absence of copper in the nutrient medium. Integration of transcriptome data into the previously developed model made it possible not only to more accurately predict growth rates under the cultivation conditions used, but also to reveal quantitative differences in the distribution of fluxes depending on the substrate and the presence of key metals for bacterial growth.

## References

1. Chistoserdova L. Methanotrophy: An Evolving Field. In: Methane Biocatalysis: Paving the Way to Sustainability. Springer, Cham. 2018;1-15.
2. Kalyuzhnaya M.G., Collins D., Chistoserdova L. Microbial Cycling of Methane. 2019.
3. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
4. 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.
5. Vallenet D., Calteau A., Dubois M., Amours P., Bazin A., Beuvin M., Burlot L., Bussell X., Fouteau S., Gautreau G., Lajus A. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res*. 2020;48(D1):D579-D589.
6. Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140.
7. Akberdin I.R., Thompson M., Hamilton R., Desai N., Alexander D., Henard C.A., Guarnieri M.T., Kalyuzhnaya M.G. Methane utilization in *Methylobacterium alcaliphilum* 20ZR: a systems approach. *Sci Rep*. 2018a;8(1):1-13.
8. Akberdin I.R., Collins D.A., Hamilton R., Oshchepkov D.Y., Shukla A.K., Nicora C.D., Nakayasu E.S., Adkins J.N., Kalyuzhnaya M.G. Rare earth elements alter redox balance in *Methylobacterium alcaliphilum* 20ZR. *Front Microbiol*. 2018b;9:2735.
9. Kim M.K., Lane A., Kelley J.J., Lun D.S., E-Flux2 and SPOT: validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *PLoS One*. 2016;11(6):e0157101.
10. Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: constraints-based reconstruction and analysis for python. *BMC Systems Biol*. 2013;7(1):74.
11. Rocha I., Maia P., Evangelista P., Vilaça P., Soares S., Pinto J.P., Nielsen J., Patil K.R., Ferreira E.C., Rocha M. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Systems Biol*. 2010;4(1):1-12.

# Genome-scale reconstructions of microbial biodynamics: great expectations, modest current success, and prospects for improvement

Panikov N.S.

Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA  
n.panikov@northeastern.edu

**Keywords:** growth kinetics; conditional expression, cell composition; environmental factors; stress resistance; limitation

*Motivation: The importance of genome-scale models (GEM).* Today, NCBI lists more than 300,000 completed whole-genome projects, they cover almost the entire diversity of cultivable microorganisms. Further progress critically depends on how efficiently sequence data are converted into solid biological knowledge. Regular bioinformatics tools are not sufficient, and much higher expectations are coming from systems biology, a holistic approach aimed to discover emergent properties of biosystems using GEM as a research tool. The ultimate goal of GEM is the accurate prediction of *phenotype* from *genotype* under specified environmental conditions. If successful, it would be a remarkable scientific breakthrough, the next great step of the genomic revolution accelerating progress by substitution of real experiments with verified simulations. An especially high benefit is expected for studies of the hard-to-culture microorganisms, such as biotrophic species (symbionts or parasites poorly growing without the host) or not-yet-cultivable soil and aquatic bacteria with entirely unknown phenotypes.

*Historical roots of contemporary GEMs:* 1) biochemical modeling of metabolic cycles and 2) growth kinetics of microorganisms. The first approach uses equations of enzyme kinetics to analyze the regulatory metabolic circuits. The second approach simulates growth using the mass-balance ODEs for such complex variables as nutrients, cell mass, and products. The still-popular Monod model recognizes nutrients as the only controlling factor. More advanced models (structured, cybernetic, SCM) account for changeable cell constituents and attempt to mimic the regulatory mechanisms. The closest to GEMs is the synthetic chemostat model (SCM) [1], which links growth kinetics (e.g., specific growth rate, SGR) with conditionally expressed macromolecular cell composition (MMCC); it allows successful mechanistic simulation of a wide range of dynamic observation (Figure).

Contrary to enzyme kinetics, microbial growth is to be described in a dualistic way using rate and compositional variables. If SGR and MMCC remain constant in a stable environment, we call it *simple growth* (syn: balanced, steady-state). The real-life environment is not stable causing self-adjusting changes in both SGR and MMCC called the *complex biodynamics*. It prevails over simple patterns (Fig. 1, panels 1–3) but remains reproducible and allows precise kinetic studies.

*Overview of GEM.* Any microbial genome consists of three parts [2]: (i) the M-matrix (Metabolism) encoding enzymes of intermediary metabolism; (ii) the E-matrix (Expression) encoding enzymes and RNAs involved in transcription and translation; and (iii) the O-matrix (Operon), the transcriptional regulatory network (has not yet been incorporated into GEM). The first successful GEM known as FBA (Flux Balance Analyses), deals with the M-matrix. The FBA stems from a biochemical predecessor and

pragmatically ignores time, cell growth, and any kinetic and regulatory considerations. Instead, it uses several simplifying assumptions (steady-state, flux constraints, constant average MMCC) and the optimality principle (linear programming with a single objective function, e.g., the SGR maximization) to uniquely resolve the gigantic M-matrix (up to 30 % of the whole genome). The FBA turned out to be a very popular and efficient computational tool assisting metabolic engineering and displaying a reasonable agreement with key experimental data, such as gene essentiality, the effect of mutations, aerobic-to-anaerobic transition, etc. The next generation of GEM included dynamic FBA (added exchange module simulating nutrients uptake and release of products), ME-models (account of metabolism and expression), and finally the WC-models (the modular whole-cell reconstructions). The ME- and WC models have a larger genome coverage than FBA which became possible due to impressive progress in numeric integration of gigantic models, up to 10,500 non-linear ODE in the WC-model of *E. coli* [3], probably the largest sets of ODE in the history of science.

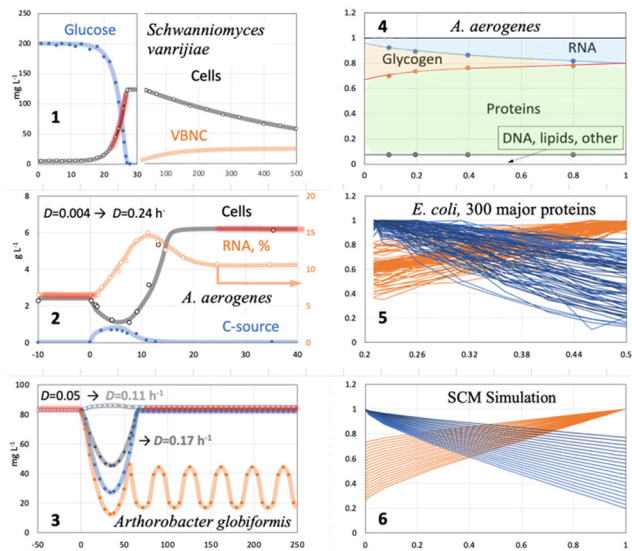
*The limited success of the advanced GEMs.* To reproduce the complex biodynamics, any GEM should satisfy the following three conditions:

1) to account for the effects of environmental factors (EF), as a minimum, the limiting substrate concentration in the M-matrix;

2) to simulate the adaptive MMCC changes including the conditional expression of proteins in response to variable EF;

3) to compute the SGR as a final model's output rather than a fixed input variable. The dFBA models contain the outdated Monod-style exchange module, so they do not meet the conditions 2) and 3) and can reproduce only the simple growth. The published ME-models [5] formally comply with all three requirements but in the wrong way. The most essential drawbacks were i) omitting the nutrient concentration as a

variable, ii) incomplete list of changeable MMCC (see below), and iii) using SGR as an independent variable (widely spread misconception fully discussed in the talk). As a result, the ME models badly failed in an attempt to reproduce the proteomic profile. Ironically, the simple SCM turned out to be a better expression predictor (see Fig. 1, compare 5 and 6). The published WC models [3] assumed the constancy of SGR and MMCC therefore none of the three conditions are met and these models can reproduce



**Fig. 1.** Examples of experimentally observed and simulated simple (red-marked fragments of cells curves) and complex microbial dynamics: 1) batch culture from lag- to death phases including the formation of VBNC (viable-but-not-cultivable cells); 2) shift-up transient in a chemostat with overshoots and undershoots; 3) shift-up with sustained oscillations; 4) MMCC changes in chemostat; 5) proteomic profile changes in a chemostat, orange, and blue curves are the normalized content of the up- and downregulated proteins; 6) proteomic profile simulated by SCM. All experimental data were fitted to the SCM. From [4] with permission

only the simple growth. The talk discusses the benefit of complex data in modeling and explains why the principle *learn to walk before you run*, does not work as expected.

*What are the main deficiencies of existing GEMs and how they can be improved?* Let us face it: *the first pancake is always lumpy*, and the recently constructed dynamic GEMs (dFBA, ME- and WC) have too many drawbacks. They are extremely bulky and not as useful in practice as the static FBA. Moreover, the accuracy of microbial dynamics simulation is not as good as with the best pre-genomic models. One reason is a generation gap: contemporary modelers overlooked the progress in microbial growth kinetics achieved before 2000. The talk specifies how the kinetic framework of existing GEM can be upgraded; the simplest solution is a merging of two models: the regular static FBA and the SCM.

Another discussion point is the strategy of GEM development. The current trend is to increase as much as possible the size of GEM (more genes and metabolites, supplementing the M- and E-matrices with kinetic and regulatory data, etc.), and further improve the computational algorithms. An expectation is that close to 100% coverage of the genome combined with constraints and objective functions like SGR-maximization is enough to provide an adequate simulation. We argue based on the history of mathematical biology that described strategy is counterproductive. Instead, we suggest trying more efforts to test the alternative hypotheses about the systems biology of bacteria. The objective functions should be related to the life strategies of the wild predecessors subjected to natural selection. The SGR-maximization as an objective function can be rejected as a wrong assumption based on the expression profile (see Fig. 1, panel 5). We conclude that the life strategy of *E coli* is to keep a balance between growth and stress resistance, which play the same role as the immune system in mammals.

### References

1. Panikov N.S. Microbial Growth Kinetics. London: Chapman & Hall, 1995.
2. Palsson B.Ø. Systems biology: Constraint-based Reconstruction and Analysis. Cambridge: Cambridge Univ Press, 2015.
3. Macklin D.N., Ahn-Horst T.A., Choi H., Ruggero N.A., Carrera J., Mason J.C. et al. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science*. 2020;369(6502):eaav3751.
4. Panikov N.S. Genome-Scale Reconstruction of Microbial Dynamic Phenotype: Successes and Challenges. *Microorganisms*. 2021;9:2352-2400.
5. O'Brien E.J., Lerman J.A., Chang R.L., Hyduke D.R., Palsson B.O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*. 2013;9:1-13.

# Determination of parameters of a mathematical model of bacterial infection development taking into account antimicrobial resistance

Surnin P.<sup>1\*</sup>, Shishlenin M.<sup>1,2</sup>, Bocharov G.<sup>3</sup>

<sup>1</sup> *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *G.I. Marchuk Institute of Computational Mathematics, RAS, Moscow, Russia*

\* [p.surnin@g.nsu.ru](mailto:p.surnin@g.nsu.ru)

**Key words:** inverse problem, antibacterial resistance, identification of parameters

*Motivation and Aim:* Bacteria that cause certain diseases sooner or later develop resistance to antibiotics used during treatment. This is a natural adaptation process called antimicrobial resistance (AMR). Its development means that the duration of the effectiveness of antibiotics is limited, and their improper and unjustified use contributes to the emergence and spread of antibiotic-resistant bacteria. More and more infectious diseases, such as pneumonia, tuberculosis, gonorrhea and salmonellosis, are becoming more difficult to treat due to the reduced effectiveness of antibiotics. The consequence of antibiotic resistance is longer hospitalizations, increased medical costs and mortality. AMR is today one of the most serious threats to human health, food security and development.

*Results:* We apply mathematical modeling to study the development of AMR [1]. We investigate a mathematical model [2] based on a system of nonlinear ODEs that describes the development of populations of sensitive and resistant bacteria under the action of an antibacterial drug and the immune response of body cells.

However, some parameters of the model cannot be measured experimentally, and therefore an important task is to identify the parameters of the mathematical model.

The inverse problem is to identify the parameters of the mathematical model of AMR, such as the intensity of bacteria that destroy immune cells, the growth rate of resistant bacteria, the constant intensity of conjugation, the frequency of mutations of susceptible bacteria, according to additional information about the population of sensitive and resistant bacteria in the human body.

*Conclusion:* The result of the study is the developed method for solving the inverse problem, which is reduced to minimizing the cost functional by the method of particle swarm optimization and differential evolution.

## References

1. Daşbaşı B., Öztürk İ. On the stability analysis of the general mathematical modeling of bacterial infection. *Int J Eng Appl Sci.* 2018;10(2):93-117.
2. Nechepurenko Y.M., Hristichenko M.Y., Grebennikov D.S., Bocharov G.A. Analysis of bistability of models of viral infections with a delayed argument. Preprints of Keldysh M.V. *IPM.* 2019;017:1-26.

## The secondary structure model of the transcriptional terminator of the *C. glutamicum* *ilvBNC* operon proved by mutational analysis of the operon gene expression and valine production

Titov I.I.<sup>1,2\*</sup>, Ryabchenko L.E.<sup>3,4</sup>, Leonova T.E.<sup>3,4</sup>, Kalinina T.I.<sup>3,4</sup>, Gerasimova T.V.<sup>3,4</sup>, Kolchanov N.A.<sup>1,2</sup>, Yanenko A.S.<sup>3,4\*\*</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup>*NRC Kurchatov Institute-Gosniigenetika, Kurchatov Genomic Center, Moscow, Russia*

<sup>4</sup>*NRC Kurchatov Institute, Kurchatov Genomic Center, Moscow, Russia*

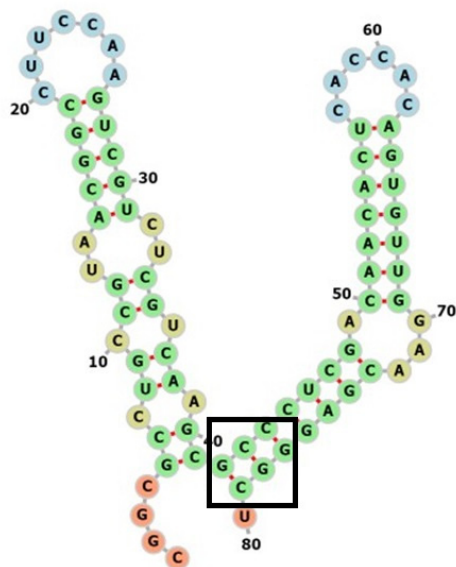
\* [titov@bionet.nsc.ru](mailto:titov@bionet.nsc.ru), \*\* [yanenko@genetika.ru](mailto:yanenko@genetika.ru)

**Key words:** attenuator, transcriptional terminator, RNA secondary structure, bacteria, valine

*Motivation and Aim:* L-Valine is an essential branched-chain amino acid for humans and animals: it is widely added in the products of food, medicine, and feed. L-leucine is predominantly produced through microbial fermentation by using *Corynebacterium glutamicum* as host bacteria. In recent years, continuing efforts have been made in revealing the mechanisms and regulation of L-valine biosynthesis in *C. glutamicum* with the goal for genetic engineering the industrially competitive L-valine-producing *C. glutamicum* strains.

It is known that the *ilvBNC* operon of *C. glutamicum* encodes acetohydroxy acid synthase and isomeroreductase, which are key enzymes of L-valine synthesis. Morbach et al. [1] have shown that the *ilvBNC* transcription is controlled by an attenuation mechanism involving antitermination. In this study, we have built the computer model of *ilvBNC* attenuator secondary structure and examined it by introducing the mutations into the attenuator region and measuring the *ilvB* transcription level and valine production of the mutant strains.

*Results:* The RNA secondary structures of the attenuator region were calculated using the unafold [2] and GArna [3] programs. The attenuator consists of two hairpins that compete for pairing with GCC triplet (Fig. 1). It is important that low-barrier GCC-pairing transition between these hairpins allows fast switching between terminator and antiterminator modes thus providing fluent regulation of *ilvB* transcription. Using this model and mainly focusing on the triplet critical region, we suggested five single and one double mutations and predicted their effect on the level of *ilvB* transcript. Some of these mutations were expected to increase the valine levels while others would retain it. Then we experimentally tested our predictions and found that they are confirmed by experiment. In particular, we observed that two of the mutations drastically change valine production as much as increase it up to 28 times compared to the wild type.



**Fig. 1.** Predicted RNA secondary structure of the wild type of the *ilvBNC* attenuator of *C. glutamicum*. The antiterminator and terminator hairpins are left and right, correspondingly. The rectangular shows the critical region for attenuator regulation

*Conclusion:* In this work, we built the model of the *ilvBNC* attenuator of *C. glutamicum* and supported it with the measurement of *ilvB* transcription level and valine production of the mutant strains. This provides one of the key elements for modelling the regulation of L-valine biosynthesis and for engineering the efficient L-valine-producing *C. glutamicum* strains.

*Acknowledgements:* The work of I.I.T. and N.A.K. was supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

The work of L.E.R., T.E.L., T.I.K., T.V.G. and A.S.Y. was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2019-1658).

### References

1. Morbach S., Junger C., Sahn 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(5):501-507.
2. Markham N.R., Zuker M. UNAFold: Software for Nucleic Acid Folding and Hybridization. In: Keith J. (Ed.). Data, Sequence Analysis, and Evolution, Bioinformatics: Volume 2, Chapter 1. Humana Press Inc., 2008;3-31.
3. Titov I., Vorobiev D., Ivanisenko V., Kolchanov N. A fast genetic algorithm for RNA secondary structure analysis. *Russ Chemical Bulletin.* 2002;51:1135-1144.

# 6 Symposium “Genetics, bioinformatics and systems biology of plants”



- 6.1 Section “Genomics, genetics and systems biology of plants” [592](#)
- 6.2 Section “Developmental biology of plants: computational and experimental approaches” [647](#)



## Large scale analysis of the crop transcriptomic data: analysis of out of the reference transcripts

Afonnikov D.A.<sup>1,2,3\*</sup>, Genaev M.A.<sup>1,2,3</sup>, Shmakov N.A.<sup>1,2,3</sup>, Mustafin Z.S.<sup>1,2,3</sup>,  
Mukhin A.M.<sup>2,3</sup>, Lashin S.A.<sup>1,2,3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* [ada@bionet.nsc.ru](mailto:ada@bionet.nsc.ru)

**Key words:** crop, RNA-seq, large scale analysis, annotation

*Motivation and Aim:* The analysis of crop gene expression based on RNA-seq experiments is one of the most effective ways to search for genes of biological significance. The results are important for geneticists and breeders in the breeding new lines and cultivars of improved stress response, the search for markers of new useful genes. However, most of the results of gene expression analysis published in articles and databases are based only on reference genomic sequences. For agricultural plants, there are more and more data on transcriptomes of varieties and lines, the genotype of which differs from the genotype of the reference organism. Most of these transcriptomes contain sequences that are not detected in the reference genome and can only be obtained by the *de novo* assembly method.

*Methods and Algorithms:* In this work, a large-scale analysis of transcriptomes of 5 crops (maize, rice, tomato, potato and barley) taken from the available SRA archives of NCBI and EBI (over 1200 libraries in total) was carried out. We aimed at identification of “Out Of the Reference Transcripts” (OORT) in RNA-seq libraries and their annotation. For each of the libraries *de novo* transcript sequences were reconstructed and aligned to reference genome. Sequences of two types were identified: (1) transcripts aligned to the unannotated reference genome loci; (2) transcripts unaligned to reference genome.

*Results:* It is shown that the proportion of transcripts that are aligned to the reference unannotated loci varies from 20 to 25 %. Proportion of transcripts unaligned to the reference genome is up to 5 %. For sequences of "new" transcripts not aligned to the genome, the identification of ORFs and amino acid sequences was carried out and their annotation was performed. Some of such transcript were identified as non-coding RNAs, viral and pathogen sequences. We also identified candidate for plant resistance genes among OORT: 181 for unaligned transcripts and more than 1500 for unannotated. Transcripts that are homologous to genes of plant stress response to drought, oxidative stress, high temperatures and genes of plant resistance to pathogens were also identified.

*Acknowledgements:* Development of algorithms 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.

## The low level of variability is depicted in plastomes of early soybean varieties

Aleksandrovich V.\*, Siniuskaya M., Liaudanski A., Mishuk Y., Shatarnova T., Danilenko N., Davydenko O.

*Institute of Genetics and Cytology, National Academy of Sciences of Belarus, Minsk, Belarus*

\* *valeria.alexandrovich@gmail.com*

**Key words:** soybean, genetic diversity, plastid genomes, next-generation sequencing

**Motivation and Aim:** The chloroplast genome of land plants is usually conservative in structure, but variable in sequence. Today the information about the complete plastidial sequences is growing fast. Phylogenetical comparative studies between taxons based on organellar DNA became deeper and more complicated. Intraspecies diversity of organellar DNA became apparent and possibly play an important role in successful breeding and creation of the new varieties in plants. Here we report about the sequencing of complete plastomes of 24 early-maturing soybean (*Glycine max* (L.) Merr.) cultivars. The aim of our study was to evaluate the level of chloroplast genomes diversity in specific group of *Glycine max* varieties.

**Methods and Algorithms:** DNA was obtained from the 7-day-old seedlings by phenol-chlorophorm extraction from the chloroplast fraction, isolated by differential centrifugation. The chloroplast DNA was then sequenced on Illumina MiSeq using Illumina DNA Prep library preparation kit and MiSeq Reagent Kit v3 (600-cycle). The NGS data processing algorithm included reads alignment on reference *G. max* plastid sequence (NC007942.1) using Burrows-Wheeler Alignment Tool followed by converting to .bam format and sorting with Samtools. Variant calling was performed with GATK HaplotypeCaller followed by filtering with Bcftools.

**Results:** Analysis of 24 complete soybean chloroplast genome sequences demonstrated the drastically low level of genetic diversity in studied varieties of *Glycine max*. Only 3 INDELs and 6 SNP were revealed between samples under investigation. 4 SNPs were located in coding sequences of *atpB*, *rps4*, *accD* and *rps3* genes. All of them were synonymic. The studied chloroplast genomes can be subdivided into at least 3 subtypes based on found polymorphisms, possibly linked with origin of maternal form. There can be revealed more distinct subtypes of chloroplast DNA, as we are planning to investigate other cultivars from our collection. Mitochondrial genomes of the same 24 soybean cultivars now are also under investigation.

**Conclusion:** Here we report about low level of variability in chloroplast DNA of *Glycine max*. Possibly it is the result of the founder effect – evidently a narrow range of ancestors was used when this group of varieties was created. Mitochondrial genomes of the same 24 soybean cultivars now are under investigation.

**Acknowledgements:** The study was carried out as a part of the State Scientific Research Program «Biotechnologies» 2021–2023, Subprogram 2 «Structural and functional genomics», assignment 2.1.3.

## Genomic regions associated with fiber-related traits in flax

Bankin M.<sup>1\*</sup>, Kanapin A.<sup>2</sup>, Samsonova A.<sup>2</sup>, Rozhmina T.<sup>3</sup>, Samsonova M.<sup>1</sup>

<sup>1</sup> *Mathematical Biology & Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>2</sup> *Centre for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>3</sup> *Laboratory of Breeding Technologies, Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

\* *mikle.p.bankin@gmail.com*

**Key words:** GWAS, flax, fiber-related traits

*Motivation and Aim:* For thousands of years mankind have cultivated flax for its fiber [1]. Flax fiber was used as a textile raw material for production of cords, weaving yarn and more later fashionable garment and high-quality fabric upholstery. Since 1930's a unique mechanical properties of flax fibers paved the way to their application as reinforcement in composite materials [2]. The current information on genomic regions associated with fiber is not limited but is also quite inaccurate as the reduced representation sequencing methods were employed instead of whole genome sequencing and in most cases the biparental crosses were used. Several studies identified QTLs associated with fiber yield and fiber content, fiber technical length, straw weight and cell wall [3–5] analyzing RIL populations from crosses between linseed and fiber flax parents. Here we expand the knowledge on genetic control of fiber traits by mining 306 flax accessions maintained at Federal Research Centre of the Bast Fiber Crops. An in-depth characterization of genetic regions controlling fiber-related traits could provide foundation for further genomic selection and breeding efforts aimed on development of cultivars with superior fiber quality.

*Methods and Algorithms:* 306 flax specimens containing 182 fiber flax accessions, 120 linseed flax accessions and 4 accessions of unknown morphotype were selected from the collection of the Federal Research Center for Bast Fiber Crops. The dataset included landraces, elite cultivars and breeding lines. All plants have been grown in the experimental field of Federal Research Center for Bast Fiber Crops in Torzhok, Tver region, Russia. A detailed description of plant growing and phenotyping was published in [6]. DNA was extracted from collected leaves with DNeasy Plant Mini Kit (Qiagen). DNA sequencing was performed at the BGI using Illumina protocol generating paired-end reads 150bp in length. 1143.850625 Gbytes of raw data comprising 7.626 billion reads with an average of 9.3x coverage or 3.7 Gbp per sample were generated. The genome-wide association analyses were performed with GAPIT3 R package (models: GLM, MLM, CMLM, FarmCPU, SUPER, Blink) [7] and GEMMA model [8].

*Results:* Univariate GWAS analysis identified 206 significantly associated QTNs, of which 21 were associated with several traits. Most QTNs were associated with technical length (69 QTNs), elementary fibre length (49 QTNs) and fiber content (33 QTNs). Analysis of genes around QTNs revealed 7 potential candidate genes. GWAS results using multivariate model implemented in GEMMA package identified 354 QTNs. Analysis of genes around QTNs, taking into account the LD value for each chromosome, revealed 13 potential candidate genes.

**Conclusion:** The predicted candidate genes were subdivided into five functional groups: (1) genes related to the biosynthesis and subsequent modification of polysaccharides that make up the secondary and tertiary cell walls of flax fiber; (2) genes associated with vesicular transport and maintenance of intracellular trafficking of the cell wall components; (3) genes associated with the processes accounted for by acid-growth hypothesis (4) genes associated with hormonal regulation, and (5) genes associated with resistance to plant diseases. Our results provide novel insights into the genetic architecture of flax fiber-related traits and pinpoint potential candidate genes for further in-depth studies.

**Acknowledgements:** This work is supported by Peter the Great St. Petersburg Polytechnic University in the framework of Russian Federation's Priority Strategic Academic Leadership Programme (Agreement 75-15-1038-2021-1333).

### References

1. Allaby R.G. et al. Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the sad2 locus. *Theor Appl Genet.* 2005;112:58-65.
2. Fogorasi M.S., Barbu I. The potential of natural fibers for the automotive sector – review. *IOP Conf Ser: Mater Sci Eng.* 2017;252:12-44.
3. Zhang J. et al. Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 2018;18:160.
4. Kumar S. et al. QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). *Theor Appl Genet.* 2015;128:965-84.
5. Xie D. et al. Genomic variations and association study of agronomic traits in flax. *BMC Genomics.* 2018;19:512.
6. Rozhmina T. et al. A comprehensive dataset of flax (*Linum usitatissimum* L.) phenotypes. *Data Brief.* 2021;37:107224.
7. Wang J., Zhang Z. GAPIT Version 3: Boosting power and accuracy for the genomic association and prediction. *Genom. Proteom. Bioinform.* 2021;19(4):629-640.
8. Zhou X., Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44:821-824.

## Inheritance of the type growth and maturity in *Aegilops squarrosa*, a donor of the D hexaploid wheat genome

Chepurnov G.<sup>1,2\*</sup>, Blinov A.<sup>1</sup>, Goncharov N.P.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* ChepurnovGY@bionet.nsc.ru

**Key words:** *Ae. tauschii*, Vrn-D1, Vrn-D2, gene, sequencing, vernalization

Wheat is one of the main food crops of mankind. In conditions of global and impending change, the impact on a favorable environment to changing environmental conditions has an impact on the genes that control precocity. According to some estimates, the genes of the *Vrn* system (response to vernalization), which determine not only the type, but also the rate of development of wheat plants, make the main contribution to the characterization. All diploid species of *Triticum* L., with the exception of *T. sinskajae*, appear mainly as winter forms. The growth of most spring forms of these species is late-ripening. The last remark also applies to diploid species of the genus *Aegilops* L. The only exception is the species *Ae. squarrosa* L. (syn. *Ae. tauschii*), which are the most early maturing forms not only for the genera *Triticum* and *Aegilops*, but also for the wheat tribe as a whole. Vegetation period of accessions K-1954 and K-992 *Ae. squarrosa* in the conditions of Novosibirsk, on average for three years, the study did not exceed 60 days. The type of development (spring vs. wintering) and precocity in hexaploid ( $2n = 6x = 42$ ) common wheat *Triticum aestivum* L. determine for *VRN-1*, *VRN-2*, *VRN-3* loci. One of the donors of the main genomes of common wheat is *Ae. tauschii* ( $2n = 2x = 14$ ) is the object present study. The paper presents the results of the determination of allelic variants of the *Vrn-1D* gene and the established nucleotide sequences of the *ZCCT-1D* and *ZCCT-2D* genes of the *Vrn-2D* locus. Total DNA was isolated from 28 spring and winter accessions of *Ae. tauschii*, PCR amplification of the region of the first intron of the *Vrn-1D* gene was carried out. In this development, which, as demonstrated in previous studies [4, 5], is manifested by the presence of the trait "loss of the vernalization requirement". In 6 samples, there was no deletion in the intron, which is found to be the presence of an intact gene in them. Thus, the sign of "loss of the vernalization requirement" in this accessions of *Ae. tauschii* does not disrupt the *Vrn-1D* gene. In this samples, nucleotide sequences were established for the sections of the *ZCCT-1D* and *ZCCT-2D* genes, structural changes in which (deletions of 24 and 1 bp respectively), as described in the description [1–3, 6], disrupt the functionality of genes. According to the results of the study, all 6 samples did not reveal published deletions, which reveals the presence of other alleles in them, which determine the sign of "loss of the vernalization requirement". They probably have other, previously undetectable, disorders within the *ZCCT-1D* and *ZCCT-2D* genes or changes in traces upstream of the *Vrn-3D* gene promoter, already described *Vrn-3B* for cv. Hope. Also, one should not exclude the possibility of disruption of the association of sites with the TaGRP2 protein in the first region of the *Vrn-1D* intron without availability a large deletion. The presence of such an allele was going to help clarify the specific binding sites and complement the already observed literature data, however, the study of changes in the sequences of the *Vrn-1D* and *Vrn-3D* genes will be developed in the future.

*Acknowledgements:* This work was supported by the Russian Science Foundation (grant No. 22-16-20026).

### *References*

1. Distelfeld A., Tranquilli G., Li C., Yan L., Dubcovsky J. Genetic and molecular characterization of the *VRN2* loci in tetraploid wheat. *Plant Physiol.* 2009;149:245-257.
2. Dubcovsky J., Loukoianov A., Fu D., Valarik M., Sanchez A., Yan L. Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. *Plant Mol Biol.* 2006;60:469-480.
3. Kippes N., Chen A., Zhang X. et al. Development and characterization of a spring hexaploid wheat line with no functional *VRN2* genes. *Theor Appl Genet.* 2016;129:1417-1428.
4. Shigeo T., Kayoko K., Kensuke F., Fuminori K. Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. *Genes Genet Syst.* 2011;86(3):183-195.
5. Xiao J. et al. O-GlcNAc-mediated interaction between *VER2* and *TaGRP2* elicits *TaVRN1* mRNA accumulation during vernalization in winter wheat. *Nat Commun.* 2014;5:4572.
6. Yan L. et al. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science.* 2004;303:1640-1644.

# Integration of ChIP-seq and RNA-seq data in structure-function analysis of *cis*-regulatory elements

Dolgikh V.<sup>1,2\*</sup>, Wiebe D.<sup>1,2</sup>, Levitsky V.<sup>1,2</sup>, Zemlyanskaya E.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *volleydollmc1@gmail.com*

**Key words:** regulation of transcription, *cis*-elements, *cis*-regulatory code, transcription factors

**Motivation and Aim:** The precision of gene regulation by transcription factors (TFs) is controlled at multiple levels, including the physical properties of TF–DNA interactions and TF oligomerization, among others. This is reflected in a complex structure of *cis*-regulatory elements. They may represent *cis*-regulatory modules consisting of several TF binding sites; besides, divergent binding sites may correspond to one TF. Due to a high combinatorics, the systemic experimental studies of *cis*-regulatory complexity are tricky. Bioinformatics integration of ChIP-seq data (which, along with the target TF binding site, contain the footprints of its partners) with RNA-seq data (which reflect transcriptional response of potential TF targets) can facilitate the systemic studies of *cis*-regulatory elements. However, to our knowledge, no software for such integrated analysis is currently available. Therefore, we developed a pipeline for structure-function analysis of *cis*-regulatory elements based on integration of ChIP-seq and RNA-seq data.

**Methods and Algorithms:** The pipeline is implemented in R programming language. It consists of three modules. Module (I) performs a repetitive *de novo* motif search in ChIP-seq peaks with HOMER, followed by motif clustering, cluster analysis and annotation of the inferred motifs using universalMotif and igrph R packages. Module (II) analyzes pairwise co-occurrence of the inferred motifs in ChIP-seq peaks with MCOT. Module (III) estimates an enrichment of the putative regulatory elements defined at the previous stages in certain promoter regions of differentially expressed genes using Fisher’s exact test.

**Results:** The pipeline uses ChIP-seq peaks for TF of interest, the set of differentially expressed genes, which reflect transcriptional activity of potential TF targets, and the background set of genes as input data. Noteworthy, the algorithm beneath the pipeline is capable of automatically (with no expert involvement) reducing technical artifacts of *de novo* motif search, and segregating related motifs, which could potentially represent divergent binding sites for the same TF. The pipeline output contains the list of motifs and composite elements (the simplest *cis*-regulatory modules consisting of two TF binding sites), which potentially regulate the target TF-driven transcription, as well as the functional characteristics of each regulatory element, including association with transcriptional response. The output is supplemented with clear visualizations. The pipeline application for analysis of *cis*-elements involved in regulation of hormone responses in *Arabidopsis thaliana* identified different modes of transcriptional regulation and allowed classifying hormone-related TFs.

**Conclusion:** Our results demonstrate that the developed pipeline can foster the screening studies of structure-function organization of *cis*-regulatory elements

**Acknowledgements:** This work was supported by the RSF grant No. 20-14-00140.

## Structural variation in flax *Linum usitatissimum* L. genome

Duk M.A.<sup>1,2\*</sup>, Kanapin A.A.<sup>1</sup>, Samsonova A.A.<sup>1</sup>, Rozhmina T.A.<sup>3</sup>, Samsonova M.G.<sup>1</sup>

<sup>1</sup> Centre for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

<sup>2</sup> The Ioffe Institute, St. Petersburg, Russia

<sup>3</sup> Federal Research Center for Bast Fiber Crops, Thorzhok, Russia

\* duk@mail.ioffe.ru

**Key words:** flax, genomics, sequencing, structural variation

**Motivation and Aim:** Flax is one of the significant agricultural crops for Russia cultivated for fiber and oil. The characterization of the genetic diversity is of great importance for the sustainability and diversification of the production of flax, as well as for successful breeding programs. The collection of the Federal Research Center for Bast Fiber Crops covers the most genetic diversity of this crop. It includes flax varieties from Eurasia with a large proportion of inherited Russian local forms, which distinguishes it from the collections used in previous genetic studies. Here we analyzed two types of structural variation, namely, presence-absence variations (PAV) and copy-number variations (CNV) in individual genomes from this collection.

**Methods and Algorithms:** Two datasets were analyzed, one of 100 accessions sequenced with a mean coverage of 20.6x, and the other of 306 accessions with mean coverage 9.3x. All plants have been grown in the experimental field of Federal Research Center for Bast Fiber Crops in Torzhok, Russia. For new insertion analysis we considered reads which were more than 1000 bp in length, do not align to the reference genome and have less than 25 % identity with it. CNVnator was used to analyze deletions and copy number increase [1]. Functional content of regions with structural variation was analyzed using flax genome annotation, Pfam database [2] and Gene Ontology (GO).

**Results:** Insertions of new regions in individual genomes (i.e. regions absent in the reference genome) were enriched for genes encoding proteins with antimicrobial functions (*MatE*, *Antimicrobial10*), phytopathogen cleavage (*TAXi\_C*, *TAXi\_N*), viral activity and controlling DNA-processes, energy exchange and cell membrane formation. Deletions in individual genomes happen in regions with genes involved in response to oxidative and salt stress, immune response (*TPR*, *PPR*, *ARM*), response to drought and temperature fluctuations (*WD40*, *TPR*, *LRR*), and energy exchange. Copy number increase happens in regions enriched for genes controlling response to abiotic stress (*RING/U-box*, *PPR*, *TTR*, *LRR*), seed formation and response to light. Both insertions and deletions contain plenty of domains related to retrotransposon activity, reflecting transposone relocation in individual genomes.

**Conclusion:** We found large structural variation in individual genomes of flax accessions that predominately happens in regions controlling stress and immune responses, membrane remodelling and energy metabolism. We conclude that individual genomes adapt to regional growing conditions through structural variation. Our results provide a valuable resource for future pangenomic studies.



*Acknowledgements:* The research is funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program "Priority 2030" (Agreement 075-15-2021-1333 dated 30.09.2021).

### *References*

1. Abyzov A., Urban A.E., Snyder M., Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21(6):974-84. doi: 10.1101/gr.114876.110.
2. Mistry J., Chuguransky S., Williams L. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. doi: 10.1093/nar/gkaa913.

## CENH3 gene duplication leads to heterogeneous organization of centromeric chromatin in rye

Evtushenko E.<sup>1\*</sup>, Elisafenko E.<sup>1,2</sup>, Gatskaya S.<sup>1</sup>, Vershinin A.<sup>1</sup>

<sup>1</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* [evi@mcb.nsc.ru](mailto:evi@mcb.nsc.ru)

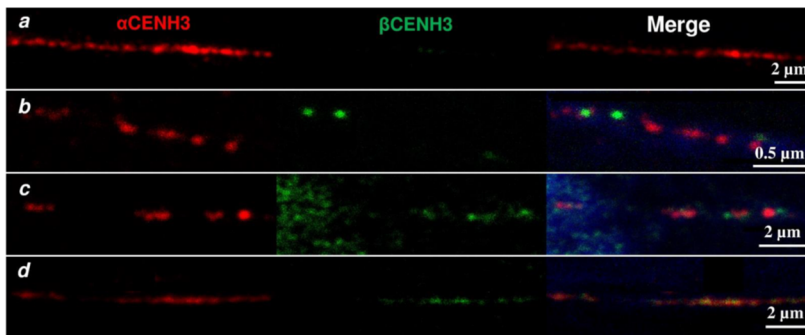
**Key words:** CENH3 variants, gene duplication, *Secale cereale*, functional DNA motifs, *CENH3* expression, subdomain organization of nucleosomes, CENH3 loading

**Motivation and Aim:** Gene duplication and the preservation of both copies during evolution is an intriguing evolutionary phenomenon. Although an important role is attributed to gene duplications in generating evolutionary novelty and adaptation [1], this role is often not so evident. Their preservation depends on the importance of the function they perform. The central component of centromere specification and function is the centromere-specific histone H3 (CENH3). Some cereal species (maize, rice) have one copy of the gene encoding this protein, while some (wheat, barley, rye) have two, and so they represent a good model for a comparative study of the functional activity of the duplicated *CENH3* genes and their protein products. We investigated the organization of the *CENH3* locus in the rye (*Secale cereale* L.) genome to identify functionally important sites in the vicinity of the  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes, the expression levels of these genes at different stages of plant development, the subsequent loading of their products, the  $\alpha$ CENH3 and  $\beta$ CENH3 proteins, into nucleosomes and patterns of their organization in linear centromeric chromatin using elongated DNA fibers.

**Methods and Algorithms:** The rye genome assembly was downloaded from GenBank ([https://www.ncbi.nlm.nih.gov/assembly/GCA\\_902687465.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_902687465.1)). The search for *CENH3* sequences in the genome was performed using the *TBLASTN* program from the AB-BLAST package [2]. Multiple alignment of regulatory regions in the vicinity of the  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes was performed using *MUSCLE*. Identification of functional sites was performed using TSSPlant and NSITE-PL [3, 4] on <http://www.softberry.com>. Plants were grown in a greenhouse. The relative expression levels of both functional CENH3 variants were examined in different rye tissue types (meiotic anther, carpel, leaf, root, stem) using quantitative real-time PCR. To find out whether proteins are synthesized from the  $\alpha$ *CENH3* and  $\beta$ *CENH3* variants, whether these proteins are loaded into nucleosomes of extended chromatin fibers from young etiolated leaf nuclei and how they are organized within the centromeric chromatin, we employed CENH3 variant-specific antibodies in combination with super-resolution spatial structured illumination microscopy (3D-SIM) and confocal microscopy.

**Results:** The alignment of the primary structure of -300 and +100 regions relative to the TSS (transcription start site), where regulatory elements responsible for specific transcription regulation normally reside, showed that they are highly conserved both the  $\alpha$ *CENH3* and  $\beta$ *CENH3* genes of different cereal species. Thus, the amount of pressure exerted by negative selection on these regions during the evolution of the *CENH3* locus was rather low. This statement is supported by a comparison of the sets of functional

sites between  $\alpha$ CENH3 and  $\beta$ CENH3. A common feature of both genes is the complex architecture of their promoters; along with TATA boxes identifiable by TSSPlant. These sets contain both common regulatory motifs, TATA boxes and the DPE and INR motives characteristic of TATA-less promoters, and those specific for each gene. These extensive sets of functional sites make the regulatory regions of the  $\alpha$ CENH3 and  $\beta$ CENH3 genes universal. A comparison of the expression levels of both genes between different developmental stages of two genotypically different rye varieties, ‘Imperial’ and ‘K69’, revealed a clear-cut correlation between expression levels and cell division intensity. Assuming that separate clusters of signals on DNA fibers correspond to particular nucleosomal subdomains, we identified main patterns of the co-organization of the CENH3 variants, which correspond to the main polynucleosomal subdomains in centromeric chromatin: 1) clusters consisting only of  $\alpha$ CENH3, the maximum size of which is about 40 kb (Fig. 1a), or  $\beta$ CENH3 signals, the maximum size of which is about 12 kb (Fig. 1b); 2) clusters consisting of alternating signals of both proteins, their size is from about kb to 120 kb (Fig. 1c, d). Clusters with CENH3 signals are interrupted by gaps with nucleosomes containing the canonical histone H3 (Fig. 1, b, c). We propose that these main patterns of polynucleosomal CENH3 subdomains form part of centromeres.



**Fig. 1.** Different types of nucleosome clusters visible on extended chromatin fibers

*Conclusion:* We determined the organization of the *CENH3* locus in rye and identified the functional motifs in the vicinity of the *CENH3* genes. Both CENH3 variants are transcribed and the corresponding proteins are centromere-incorporated in a tissue specific manner. Using extended chromatin fibers, we revealed main patterns of loading CENH3 proteins into polynucleosomal domains in centromeric chromatin where alternating nucleosomal clusters contain both CENH3 variants and nucleosomes containing the canonical histone H3.

*Acknowledgements:* The study is supported by grant of Russian Foundation for Basic Research (project No. 20–04-00699).

### References

1. Flagel L.E., Wendel J.F. Gene duplication and evolutionary novelty in plants. *New Phytol.* 2009;183:357-364.
2. Gish W. 1996-2019. <https://blast.advbiocomp.com>.
3. Solovyev V.V., Shahmuradov I.A., Salamov A.A. Identification of Promoter Regions and Regulatory Sites. In: Ladunga I. (Ed.). *Computational Biology of Transcription Factor Binding*. Methods in Molecular Biology, vol 674. Humana Press, Totowa, NJ. 2010.
4. Shahmuradov I.A., Solovyev V.V. Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinformatics.* 2015;31:3544-3545.

## Changes in the phenology of perennial plants in Western Siberia against the background of global warming

Fomin E.<sup>1\*</sup>, Fomina T.<sup>2</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Central Siberian Botanical Garden, SB RAS, Novosibirsk, Russia*

\**fomin@bionet.nsc.ru*

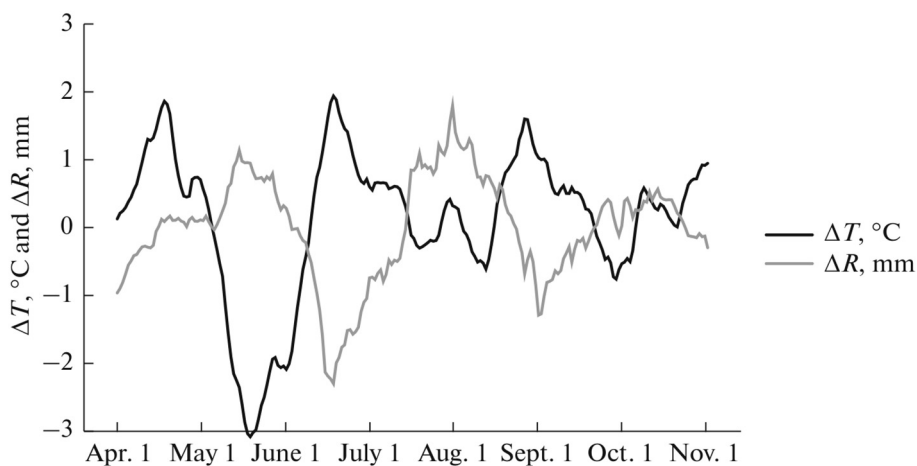
**Key words:** climate change, meteorological trends, phenological trends, species of natural flora

*Motivation and Aim:* Global warming in recent decades is considered by most researchers to be the dominant factor affecting biodiversity indices and a shift in time of the occurrence of phenological events [1]. Changes in plant phenology affect productivity, interspecific competition, seasonal activity of pollinators, pollen release [2] and also has consequences for the spread of plant pests and diseases [3]. It is expected that the effect of warming on the growth and development of plants, especially in combination with other abiotic and biological stressors--drought and increased pressure of pathogens--will increase the problems for both wild and cultivated species [4]. The aim of the study was to identify deviations of local meteorological trends from the global one and to estimate the variability of the dates of phenological development of perennial plants due to a local trend of climate change in Novosibirsk.

*Methods and Algorithms:* This study was carried out in the Central Siberian Botanical Garden, Siberian Branch, Russian Academy of Sciences (55°2'29.4" N, 82°56'4.6" E). Seventy-eight species of perennial plants from the collection of decorative species of natural flora, for which phenological observations were performed at the time interval from 1996 to 2015, served as test subjects. Data on the Ogurtsovo weather station (no. 29638) were used to characterize the climatic indices for the period of study. Data on phenological and meteorological observations were processed with standard statistical methods using R and MS Excel packages, as well as our own programs written in C++. During the study period, we calculated intraseasonal trends of air temperature T and rainfall R (from April 1 to October 31) by the construction many individual (for each date of the year) linear trends with the subsequent calculation of a simple moving average in a window with a width of 30 days to smooth out short-term fluctuations and to detect the main changes. To obtain phenological trends and assess their confidence, a bootstrapping method was used [5] suitable for any probability distributions and sample sizes.

*Results:* During the study period, an average change in air temperature in the warm season of the year in Novosibirsk was +0.16 °C; total precipitation from April to October did not change (-0.01 mm). The construction of linear temperature and precipitation trends revealed intraseasonal periods with considerable changes in the main meteorological indices (Fig. 1). Although the beginning of the season became noticeably warmer ( $\Delta T_{av} = +0.83$ ), the late spring period became colder and wetter ( $\Delta T_{av} = -1.87$ ,  $\Delta R_{av} = +0.47$ ). At the same time, the probability of hot and dry weather increased both in the first half of summer ( $\Delta T_{av} = +0.89$ ,  $\Delta R_{av} = -1.12$ ) and in early autumn ( $\Delta T_{av} = +0.80$ ,  $\Delta R_{av} = -0.37$ ). As a result, the duration of the prefloral and floral periods characterizing the intensity of seasonal development decrease significantly in all species.

In the species flowering in the first half of the season, the decrease was 5–8 days; in summer ones, it was 10 days, while late-summer ones begin to bloom up to 24 days earlier. The effect of increased summer temperatures and dryness accompanying the climatic trend is obvious. The flowering periods are characterized by a very high variability for all groups with the trends of different direction and different value.



**Fig. 1.** Change in meteorological indices in Novosibirsk for 1996–2015. *X* axis, calendar dates; *Y* axis, difference between values of moving averages for the trends of average daily air temperature  $\Delta T = T_{2015} - T_{1996}$  and precipitations  $\Delta R = R_{2015} - R_{1996}$

**Conclusion:** Meteorological local trends of the warm season in Novosibirsk for 1996–2015 noticeably higher than the global trend: spring has become longer, colder and wetter, while early summer and early autumn have become warmer and drier. As a result, the prefloral period became much shorter, which indicates an acceleration in the development of perennials. At the same time, the trends in the duration of flowering vary greatly in magnitude and are multidirectional. Phenological trends are consistent with trends in climatic indicators and describe up to 16 % of the variability in the dates of phenophases and the duration of phenological periods. Otherwise, their variability is due to inter-annual and intra-seasonal fluctuations in weather factors, as well as the high inertia of seasonal development processes in perennial plants.

**Acknowledgements:** The study is supported by Central Siberian Botanical Garden, SB RAS (AAAA-A21-121011290025-2) and the Institute of Cytology and Genetics, SB RAS (0259-2021-0009).

### References

1. Zorina A.A. et al. Climatic reasons for the shift in the timing of flowering plants in the Central Forest Reserve. *Povolzh Ekol Zh.* 2020;1:52-65.
2. Badeck F.-W. et al. Responses of spring phenology to climate change. *New Phytol.* 2004;162:295-309.
3. Forrest J. et al. Toward a synthetic understanding of the role of phenology in ecology and evolution. *Phil Tran. R Soc B.* 2010;365:3101-3112.
4. Lippmann R. et al. Development of wild and cultivated plants under global warming conditions. *Curr Biol.* 2019;29:R1326-R1338.
5. Efron B. Bootstrap methods: Another look at the jack-knife. *Ann Stat.* 1979;7(1):1-26.

## ***StCDF1* gene editing in wild potato protoplasts**

Fomin I.N.<sup>1\*</sup>, Egorova A.A.<sup>1,2</sup>, Koloshina K.A.<sup>1</sup>, Gerasimova S.V.<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* [ifomin@bionet.nsc.ru](mailto:ifomin@bionet.nsc.ru)

One of the major domestication traits in potato is tuberization at long-day conditions [1]. A key regulatory gene for tuberization time control in potato is *StCDF1* encoding CYCLING DOF FACTOR family protein. Late tuberization at short day is associated with intact allele of this gene and early tuberization at long day is associated with mutations leading to either truncated gene product or modification at C-terminus [2]. Genome comparison between wild and domestic potato showed that most studied long-day potato varieties contain truncated version of the *CDF1* gene [3]. The aim of the present study is the reconstruction of frameshift mutation leading to premature stop-codon at C-terminus of the *CDF1* gene product in wild potato. Two wild potato species namely *Solanum chacoense* and *S. polyadenium* were selected for modification. The structure of the *CDF1* gene was examined in selected genotypes. In order to integrate the precise mutation and create the truncated version of the CDF1 factor two gRNA were designed and cloned in plasmid vectors containing also Cas9 endonuclease gene. Both vectors were used for protoplast transfection with further DNA extraction and target region sequencing. Mutations were found in both wild potato species. The experiment showed the possibility to edit *CDF1* gene in potato cells. Developed gene editing strategy can further be applied for whole plants for modification of tuberization time trait in wild potato.

*Acknowledgements:* the study is funded by RFBR, project number 20-016-00217.

### *References*

1. Egorova A.A., Chalaya N.A., Fomin I.N., Barchuk A.I., Gerasimova S.V. De Novo domestication concept for potato germplasm enhancement. *Agron.* 2022;12:462. doi: 10.3390/AGRONOMY12020462.
2. Kloosterman B., Abelenda J.A. et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature.* 2013;495:246-250. doi: 10.1038/nature11912.
3. Hardigan, M.A., Laimbeer, F.P.E. et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *PNAS.* 2017;114(46):E9999-E10008. doi: 10.1073/pnas.1714380114.

## Whole-genome-based breeding in legumes

Gentzbittel L.

*Project Center for Agro Technologies – Skolkovo Institute of Science and Technology, Moscow, Russia*

*l.gentzbittel@skoltech.ru*

**Key words:** GWAS, Genomic Prediction, quantitative genetics, Machine Learning, disease resistance, yield

Legumes (*Fabaceae*) are a nutritious staple of diets around the world. Grain legumes (e.g. soybean, chickpea, bean, pea, etc...) are mostly used as source of edible proteins and oil. Forage legumes (alfalfa, *Vicia* spp.) are among the most effective sources of protein for animal husbandry.

The *Fabaceae* comprise annual and perennial species, also including some tree species (e.g. *Acacia*, *Prosopis*). The different species also evolved a large diversity of reproductive strategies, from self-pollination (e.g. soybean, pea, peanut) to full outcrossing (e.g. alfalfa) and some intermediate cases (e.g. faba bean). The complexities of the genomes are vast, from small sized diploid genomes (e.g. *Medicago truncatula*,  $2n = 2x = 16$ , genome size of 500Mb) to diploidized allopolyploids (eg. Soybean,  $2n = 2x = 40$ , genome size = 1.15Gb), to allotetraploids (cultivated peanut,  $2n = 4x = 40$ , genome size = 2.7Gb) and to very large genomes (faba bean,  $2n = 2x = 12$ , genome size = approx. 13Gb).

The reproductive strategies of the different legume crops imposed the creation and the use of very diverse breeding schemes to improve quantity, quality and resilience to biotic and abiotic stresses. The complexities of the genomes also imposed various molecular and bioinformatics strategies to be developed. We will present how Genome Wide Association Studies (GWAS), Marker Assisted selection (MAS) and Genomic Selection (GS) are implemented in some legume crops, reviewing strategies for molecular characterization, high-throughput, high quality phenotyping and computational methods.

## CRISPR-like elements and MGEs in *Arabidopsis thaliana* nuclear genome

Gorbenko I.V.<sup>1\*</sup>, Petrushin I.S.<sup>1,2</sup>, Konstantinov Yu.M.<sup>1,2</sup>

<sup>1</sup> Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia

<sup>2</sup> Irkutsk State University, Irkutsk, Russia

\* gorbenko@sifibr.irk.ru

**Key words:** *Arabidopsis*, CRISPR, transposon

Previously we reported on the discovery of CRISPR-like elements in *Arabidopsis thaliana* genome. In the ecotype Col-0 we found 1372 arrays containing direct repeats and possessing 1 to 8 spacers with a length ranging from 61 to 743 bp. Most arrays possess 1–2 spacers and thus resemble mini-CRISPR from archaeal viruses. The length distribution of the detected arrays is discrete, which speaks in favor of the canonical CRISPR structure of the detected elements. The fact that fairly aligning arrays are often found across different ecotypes suggests their ancient origin. According to present data, CRISPR can distribute across genomes residing in special transposons – casposons. We suggest that transposons in general could be the mechanism of CRISPR emerging in plant genomes. More than 30 % of CRISPR-like elements found in *A. thaliana* Col-0 were located within mobile genetic elements (MGEs), mainly belonged to the DNA/MuDR, LTR/Gypsy, and RC/Helitron superfamilies. It should be noted that CRISPR were found inside VANDAL-type MGEs, that according to recent data can remain active due to their ability to demethylate DNA. CRISPR-containing MGEs are basically located in the pericentromeric regions, probably because the TEs there retain their structural integrity more. Analyzing the diversity of CRISPR residing in TEs, we found that arrays with a high degree of homology are usually located within MGEs that are of the same superfamily, which speaks in favor of their possible emergence as a component of MGEs. It is possible that CRISPR arrays spread throughout the plant genome as a part of transposons, but then 70 % of MGE-hosts were destroyed. Given this origin, plant CRISPR-like systems may be trans-acting and utilize the molecular machinery of transposons as Cas proteins. Considering the abundance of CRISPR-like elements, we suggest that they could play a role as a component of defense mechanism against pathogenic DNA and foreign MGEs.



# Reconstruction and analysis of potato *Solanum tuberosum* pangenome of Siberian cultivars

Karetnikov D.I.<sup>1,2,3\*</sup>, Genaev M.A.<sup>1,2</sup>, Nesterov M.A.<sup>1,2</sup>, Ibragimova S.M.<sup>1,2</sup>, Vasiliev G.V.<sup>1,2</sup>, Toschakov S.V.<sup>4</sup>, Gavrilenko T.A.<sup>5</sup>, Afonnikov D.A.<sup>1,2</sup>, Salina E.A.<sup>1,2</sup>, Patrushev M.V.<sup>4</sup>, Kochetov A.V.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup> *Kurchatov Institute, Moscow, Russia*

<sup>5</sup> *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*

\* *karetnikovmit@bionet.nsc.ru*

**Key words:** potato, *Solanum tuberosum*, pangenome, genome structural variation, DNA sequencing, genome annotation

**Motivation and Aim:** Genome sequence of the crop opens the way for studying the structure of its chromosomes, distribution of repetitive and coding sequences, identification and annotation of genes. It makes possible to study gene functions and develop markers to find associations with phenotypic traits. However, there may be significant structural variations in the genomes of different organisms of the same species that cannot be represented based on the reference sequence alone and need to be described within the concept of pangenome, the complete set of genes of organisms of the same species. Sets of both common and variable genes can be distinguished in the pangenome. The variable part is particularly important for the study because it is formed by genes associated with plant adaptation to environmental conditions, resistance to diseases and abiotic and biotic stresses. *Solanum tuberosum* L. (potato) is the most important crop produced almost all over the world. Total production is over 370 million tons per year and the number of varieties is over 4000. A feature of potato genetics is the autoteraploid genome and high heterozygosity due to the predominantly vegetative method of propagation. In this paper, a study of pangenome of potato varieties of Siberian breeding was performed.

**Methods and Algorithms:** DNA from plants of 7 varieties was sequenced by paired-end short reads using the Illumina platform, allowing to obtain DNA assembly at the exome level. This assembly was further improved by using the DM1-3516 R44 reference genome. Structural variation was assessed for genotypes and were compared between Siberian cultivars and additionally four foreign varieties sequenced earlier [1].

**Results:** Based on the obtained assemblies, protein-coding genes were identified and the pangenome was analyzed. We determined its variable and conservative parts. It was shown that the pangenome structure of Siberian varieties is open, i.e., the pangenome size grows indefinitely as the number of individual genomes increases.

**Conclusion:** The results obtained provide an opportunity to further study the features of the structure and functions of protein-coding genes for potato varieties of Siberian cultivars.

**Acknowledgements:** 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.

## References

1. Kyriakidou M., Anglin N.L., Ellis D. et al. *Scientific Data*. 2020;7(1):1-6.

## Whole genome sequencing and development of KASP markers for genotyping the West Siberian wheat stem rust population

Kelbin V.N.\*, Muterko A.F., Skolotneva E.S., Sergeeva E.M., Salina E.A.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

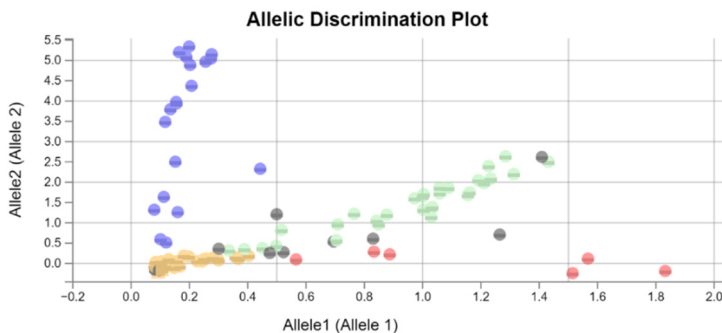
\* kelbin@bionet.nsc.ru

**Key words:** wheat stem rust, *Puccinia graminis*, genomic sequencing, KASP-markers

*Motivation and Aim:* *Puccinia graminis* f. sp. *tritici* is a biotrophic pathogenic fungus that causes the stem rust disease of bread wheat and leads to significant yield loss. During the last decade the multiple episodes of stem rust infection have been noted in the West Siberia region. The ability of the fungal pathogen to infect wheat plants is determined by the combination of virulence genes (race), at the same time either resistance or susceptibility of the host plant is provided by corresponding resistance genes (*Sr*). During the *P. graminis* life cycle its genome undergoes the recombination events, leading to the different combinations of virulence genes and together with multiple allelism make the pathogen population highly heterogeneous. The estimating of the pathogen race composition and the migration routes of virulent races in West Siberia is of great importance for the breeding of resistant bread wheat varieties. The traditional phytopathology methods are laborious and time-consuming. Therefore, there is a need for diagnostic set of genotyping PCR markers for pathogen population. The aim of the study is the development of allele-specific KASP (Kompetitive Allele Specific PCR) genotyping markers for *P. graminis* f. sp. *tritici* in the West Siberia region. For this reason, the whole genome sequencing and comparative transcriptomic analysis of the West Siberia stem rust lines was carried out for the first time.

*Methods and Algorithms:* Four lines of wheat stem rust from the Novosibirsk and Omsk regions with different combinations of virulence genes were selected for the study: Wh\_18\_6 and Wh\_18\_7 (differing by ability to infect the plants with *Sr6*, *Sr11*, *SrTmp* genes), and Duet\_sp\_1, Duet\_sp\_2 (*Sr6*, *Sr30*). The genome sequencing for each line was carried out as 2x150 bp reads on NextSeq550 (Illumina) at ICG SB RAS Genomics Core Facility Novosibirsk, Russia. The representative transcriptome, used for development of KASP markers, included predicted for *P. graminis* coding sequences [1], which are completely and unambiguously mapped with BLAT [2] to the reference genome of this species [3]. The paired reads were trimmed by quality, error corrected with Musket [4] and merged. Obtained fragments were mapped to the representative transcriptome using custom mapper, based on Minimap [5] and SSW library [6], and alignments were filtered. SNP searching was carried out during analysis of the alignments of all samples with the use of SAMtools mpileup [7]. The designed KASP-markers (LGC, Biosearch Technologies, United Kingdom) were tested on stem rust samples from Western Siberia and adjacent regions of the Russian Federation. KASP amplifications and allelic discriminations were performed using QuantStudio 5 (Applied Biosystems) Real-Time PCR machine and the data was analyzed using the QuantStudio™ Design and Analysis Software V2.6.0 (Applied Biosystems). Genotyping reactions were performed following the guideline of “Guide to running KASP genotyping on the ABI 7500 instrument”.

**Results:** The Illumina reads obtained in this work (for Wh\_18\_6, Wh\_18\_7, Duet\_sp\_1, Duet\_sp\_2 lines) were assembled in the the 135 contigs with a total length of 62.3 Mbp (*P. graminis* reference genomic sequence is of 85 Mbp). From the 379923 polymorphic sites, total of 16610 SNPs were selected for the development of 1061 KASP markers, whereof the 20 markers approximately evenly distributed across the genome were selected. Since the aim of this work was to produce KASP markers useful to monitor the phytosanitary situation and determine the races of the pathogen fungi, the KASP tests were performed to genotype 120 accessions from Western Siberia and other Russian regions. All studied KASP assays yielded consistent results, confirming the ability of markers to clearly distinguish between the Wh\_18\_6, Wh\_18\_7 and Duet\_sp\_1, Duet\_sp\_2 lines, indicating to correctly selected SNPs. Many of the KASP assays were diagnostic and polymorphic between samples from different regions of Russia (Fig. 1).



**Fig. 1.** Allelic Discrimination plot for selected KASP assays showing clustering of genotypes stem rust from regions of the Russian Federation on the Y- and X-axes. Genotypes colored red have a HEX-type allele; genotypes colored blue have a FAM-type allele; green-colored have both types of alleles (heterozygotes); black and orange dots represent non-template control

**Conclusion:** The genome sequencing of the stem rust lines from the West Siberia, having contrasting combinations of virulence genes, was carried out for the first time. By comparison of genomic data obtained from our samples with the *P. graminis* reference transcriptome the 20 KASP-markers were selected. The results obtained showed good resolution of developed KASP markers to characterize the pathogen population in the Western Siberia and other regions of Russia. The entire panel of markers is successful for diagnostic tasks and can be used in the phytosanitary monitoring and to determine the races of the pathogen fungi.

**Acknowledgements:** The study is supported by the Budget Project FWNR-2022-0007.

### References

1. Upadhyaya N.M. et al. Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. *Front Plant Sci.* 2015;5:759.
2. Kent W.J. BLAT--the BLAST-like alignment tool. *Genome Res.* 2002;12(4):656-664.
3. Duplessis S. et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *PNAS.* 2011;108(22):9166-9171.
4. Liu Y., Schröder J., Schmidt B. Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics.* 2013;29(3):308-315.
5. Li H. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics.* 2016;32(14):2103-2110.
6. Zhao M., Lee W.-P., Garrison E.P., Marth G.T. SSW Library: An SIMD Smith-Waterman C/C++ Library for Use in Genomic Applications. *PLoS ONE.* 2013;8(12):e82138.
7. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987-2993.

## The new method of DNA barcoding

Kiryanova O.<sup>1\*</sup>, Gubaydullin I.<sup>1,2</sup>, Chemeris A.<sup>3</sup>

<sup>1</sup> Ufa State Petroleum Technological University, Ufa, Russia

<sup>2</sup> Institute of Petrochemistry and Catalysis – Subdivision of the Ufa Federal Research Center, RAS, Ufa, Russia

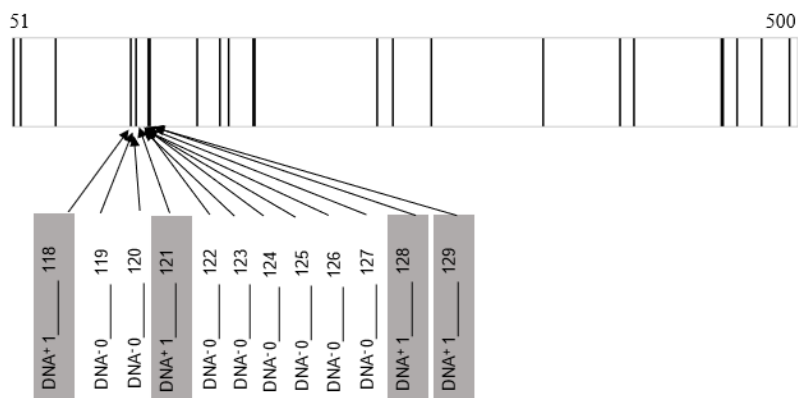
<sup>3</sup> Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Center, RAS, Ufa, Russia

\* [olga.kiryanova27@gmail.com](mailto:olga.kiryanova27@gmail.com)

**Key words:** DNA barcoding, DNA identification, *in silico* genome analysis

**Motivation and Aim:** The classification and systematization of living organisms is currently an actual task. A popular method for genomes classification is DNA barcoding. This method requires the presence of a reference DNA fragment, which allows to explicitly identify living organisms [1]. The cytochrome oxidase gene was assumed as a reference region for the living organisms. However, this gene does not lead to the satisfactory results for the plants systematization. Here we propose a new method for DNA certification of living organisms on the example of plants.

**Methods and Algorithms:** We have utilized *in silico* multiplex PCR analysis. Namely, we have found the assumed primer annealing sites in a genome. Search of the annealing positions was performed using the Boyer-Moore algorithm, where the primers are samples and genome is a text. The amplicons corresponding to the annealing sites were converted into a binary format. It should be noted, that amplicons were selected in the range from 51 to 500 nucleotides. Such amplicons have been unambiguously determined experimentally to within a nucleotide. The binary sequence is easy enough to translate into a barcode format. An example of the resulting barcode is presented on Fig. 1.



**Fig. 1.** The principle of constructing a barcode

**Results:** We have conducted a comparative analysis of the following plant genomes: wheat (6 genomes), hevea (3 genomes), and arabidopsis (9 genomes). The proposed approach was utilized to obtain barcodes for each group. Each set of barcodes indicated the differences and similarities for certain plants. It should be pointed out, that different primers provide different barcodes for the same genome. In this case, computer analysis

makes it possible to identify successful primers for the experimental PCR. As a result, the optimal number of primers could be determined for DNA certification. However, primers and amplicons should be treated as a whole to ensure the uniqueness of any DNA barcode and provide a successful distinction between living organisms.

*Conclusion:* The presented DNA certification can reveal similarities and differences at the level of large order taxa (species, family) however it more reliable for the certification of cultivars. Hence, it allows to test several primers as well as to get an idea of the success experiment. Thus, the proposed method provides a reliable and convenient way to store genetic data in a digital format, and reduce the time consumption in PCR tests planning.

*Acknowledgements:* The study was funded by Russian Foundation for Basic Research (17-44-020120).

### *References*

1. Shneyer V.S. DNA barcoding of animal and plant species as an approach for their molecular identification and describing of diversity. *J General Biol.* 2009;70(4):296-315.
2. Gasfield D. Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. St. Petersburg: Nevskii dialect, 2003.

## Dissection of candidate genes for grain texture in common wheat

Kiseleva A.A.<sup>1\*</sup>, Leonova I.N.<sup>2</sup>, Salina E.A.<sup>1</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* [antkiseleva@bionet.nsc.ru](mailto:antkiseleva@bionet.nsc.ru)

**Key words:** *Triticum aestivum*, grain texture, grain hardness, *puroindolines*, GWAS

**Motivation and Aim:** Grain texture (grain hardness) is a key trait that influences the milling and baking quality of wheat. Previously, it was shown that the Ha locus on 5DS, containing the *puroindoline a*, *puroindoline b* and *GSP-1* genes, has a crucial role in endosperm texture. Although *puroindoline* genes are considered to be major determinants of common wheat grain texture, a number of studies have demonstrated the complex nature of the trait. However, no specific genes were proposed to be involved in grain hardness organization.

**Methods and Algorithms:** There were 92 varieties adapted to the environments of the Siberia region used in this study. We estimated two parameters for grain texture in two years: flour particle diameter (FPD,  $\mu\text{m}$ ) and flour particle specific surface (FPSS,  $\text{cm}^2/\text{g}$ ). SNP genotyping was performed using the Illumina Infinium 15 k Wheat platform by TraitGenetics GmbH. The population structure was analysed using the R-based STRUCTURE-like inference algorithm LEA. For association analysis, a mixed model approach implemented in the R package “GENESIS” was used. The resulting mixed linear model considered fixed effects of SNPs and population structure and random effects for kinship. To estimate the expression of genes in the detected loci, we used data of the common wheat cultivar Azhurnaya developmental time course.

**Results:** Analysis of the flour particle size and flour particle specific surface of wheat varieties cultivated in different environments demonstrated high heritability (0.79–0.8) and strong repeatability of the traits. The genotyping with gene-specific primers demonstrated that twenty-three varieties had allele *Pina-D1k*, which lacked both *Pina* and *Pinb* genes (double null allele). Association analysis with the use of SNP genotyping (Illumina 15 K Wheat) and 2-year phenotyping confirmed the key role played by *puroindoline* genes (5D) in determining the grain texture of Russian wheat varieties. The analysis also detected significant SNPs on the 1B, 3A, 5B, 6A, 6D, 7B, 7D chromosomes. The most promising were SNPs on chromosomes 5B and 7B. These MTAs demonstrated the highest significance after FDR correction. We dissected a number of candidate genes and proposed a possible mechanism for their contribution to endosperm texture determination. These genes are involved in the metabolism of galactolipids (DGDG) and carbohydrates (1,3- $\beta$ -glucan).

**Conclusion:** Wheat grain texture is a complex trait. The crucial role of *puroindoline* genes in grain texture determination was confirmed. New loci, highly associated with the trait, were identified and dissected.

**Acknowledgements:** The study was supported by Kurchatov Genomic Center of ICG (075-15-2019-1662).

# Analysis of the size and color characteristics of wheat grains and their relationship to storage time and germination

Komyshv E.<sup>1,2\*</sup>, Efimov V.<sup>1</sup>, Genaev M.<sup>1,2,3</sup>, Koval V.<sup>1,3</sup>, Gierke P.<sup>4</sup>, Börner A.<sup>5</sup>, Afonnikov D.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>4</sup> *Landesanstalt für Landwirtschaft und Gartenbau, Dezernat 42, Prüf- und Anerkennungsstelle für Saat- und Pflanzgut, Halle (Saale), Germany*

<sup>5</sup> *Leibniz Institute of Plant Genetics and Crop Plant Research, OT Gatersleben, Seeland, Germany*

\* *komyshv@bionet.nsc.ru*

**Key words:** cereal grains, phenotyping, color characteristics, germination

*Motivation and Aim:* Grain coat is the main barrier between the grain and the external environment. A number of important biological functions are associated with its characteristics: moisture absorption, grain viability, resistance to preharvest germination. Thus, grain coat properties are important for maintaining its physiological state and longevity. In this regard, there is an ongoing interest in plant breeding in studying the grain morphological and color traits. One of the promising and relevant approaches in this regard is the use of digital cameras, which are inexpensive, but at the same time of sufficient quality for accurate color and size/shape reproduction of the research object [1].

*Results:* A computer method was developed for phenotyping the color and size/shape characteristics of cereal grains, based on the analysis of two-dimensional digital images. The developed method was used to obtain and analyze grain characteristics in 44 recombinant inbred lines of bread wheat, in which genbank storage time and germination ability were evaluated. No significant relationship with the storage time was detected for the grain shape/size traits, while 90 % of the color traits demonstrated such a correlation. The most significant negative correlations were found between the harvesting year and the traits of grain redness: the greater the storage time, the more intensive is red color component for the grains. At the same time, it was shown that grains of longer storage time have lighter coat. Analysis of linear correlations between germination of wheat seeds of different genotypes and harvesting years and their seed traits revealed a negative linear relationship between the red component of coat color and germination: the redder the grains, the lower their germination rate.

*Conclusion:* The results obtained demonstrate manifestations of metabolic changes in the coat of grains associated with storage time and their relationship with a decrease of seed viability. Our data showed that the intensity of the red component of the color of the grain coat has a negative relationship with the germination of seeds: the redder the grains, the lower their germination.

*Acknowledgements:* The study was funded by the Kurchatov Genomic Center of the ICG SB RAS, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

## References

1. Komyshv E.G., Genaev M.A., Afonnikov D.A. *Vavilov J Genet Breed.* 2020;24(4):340.
2. Afonnikov D.A., Komyshv E.G., Efimov V.M., Genaev M.A., Koval V.S., Gierke P.U., Börner A. *Plants.* 2022;11(1):35.

## Differential gene expression analysis in barley *Nud* gene mutants

Korotkova A.M.<sup>1\*</sup>, Skripileva A.I., Vihorev A.V.<sup>1</sup>, Kolosovskaya E.V.<sup>1</sup>,  
Gerasimova S.V.<sup>1,2</sup>, Khlestkina E.K.<sup>3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*

\* [korotkova@bionet.nsc.ru](mailto:korotkova@bionet.nsc.ru)

**Key words:** CRISPR, transcriptome, *Nud*, barley

**Motivation and Aim:** Barley (*Hordeum vulgare* L.) is an important food crop in the world and in Russia in particular. Naked barley provides some benefits in terms of human nutrition and health. The *Nud* gene encodes a transcription factor conceivably involved in the formation of a lipidic layer at the grain surface of hulled barley. If the *Nud* gene function is lost, lemma and palea do no longer adhere to the grain's pericarp, which entails the non-adherent hull (i.e. naked) phenotype. In a recent study, we obtained *nud* loss-of-function lines by site-directed mutagenesis using customized RNA-guided Cas9 endonuclease in the spring barley cultivar 'Golden Promise'.

**Methods and Algorithms:** One of the obtained *Nud* mutant barley line showed the best results in analysis of agronomically relevant properties has been chosen for transcriptome analysis. The purpose of the experiment is to obtain a set of genes involved in the formation of the nakedness property in barley. Isogenic lines with contrasting phenotypes are a perfect experimental model system for the elucidation of gene function. For the experiment, two isogenic lines (differing only in one *Nud* gene) were taken: control Golden promise line and *nud* mutant line. Grain's pericarp were collected from the grain at two stages of spike development:

I). After the end of flowering (early milk development),

II). After filling the grain (dough development – the scales begin to adhere to the grain).

RNA was isolated from the samples and transferred for transcriptome analysis.

**Results and Conclusion:** After analysis we've got a large table with transcription activity levels of more than 20,000 genes. We selected a group of genes among them with upregulation (51) in the mutant and a group of genes with downregulation (185). 19 genes were chosen for further analysis. Primers were picked for each transcribed gene sequence, and upregulation or downregulation of the expression of these genes was confirmed using quantitative PCR analysis.

**Acknowledgements:** This work is carried out with the support of the Russian Science Foundation No. 21-66-00012.

### References

1. Korotkova A.M., Gerasimova S.V., Khlestkina E.K.. *Vavilov J Genet Breed.* 2019;23(1):29-37.
2. Gerasimova S.V., Hertig C., Korotkova A.M., Kolosovskaya E.V., Otto I., Hiekel S., Kochetov A.V., Khlestkina E.K., Kumlehn J. *BMC Plant Biol.* 2020;20:255.
3. Gerasimova S.V., Khlestkina E.K., Kochetov A.V., Shumny V.K. *Russ J Plant Physiol.* 2017;64(2):141-155.



# Wheat yield estimation based on analysis of UAV images at low altitude

Kozhekin M.<sup>1,2\*</sup>, Genaev M.<sup>1,2</sup>, Koval V.<sup>1,2</sup>, Slobodchikov A.<sup>3</sup>, Afonnikov D.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* m.kozhekin@g.nsu.ru

**Key words:** wheat detection system, computer vision, UAV

*Motivation and Aim:* Protocols for manually counting the density of ears in crops have been the only way to estimate yields for a long time. However, this method is labour-intensive and time-consuming. An alternative is the development of automated systems operating in the field [1]. Most of such systems allow to obtain 2D images of crops and use computer vision methods for their automatic processing, in particular, for counting ears in the image. Modern methods of image analysis based on neural network algorithms and deep learning allow ears identification on the image of crops and counting their number with high accuracy [2–4]. The use of these technologies is justified due to the lower cost and acceptable accuracy compared to the labour costs of manual human observation.

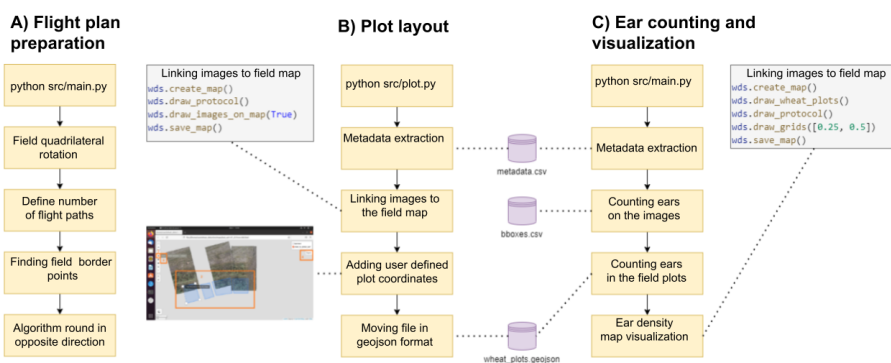
*Methods and Algorithms:* At this research faster rcnn [5] and efficient-det [6] object detection models were trained and evaluated on 3 datasets: GWHD 2020, 2021 [7, 8] and wheat images collected in a field of SibNIIRS on July 29, 2021 at stage of ear formation phase. Manual counting of ears was performed after harvesting from the area of 0.25 m<sup>2</sup>. The number of productive stems was counted in four replications.

The ear recognition accuracy estimation was used for neural network prediction results on an additional sample of images that were added to the 2021 GWHD dataset (not used in the network training). The average precision (AP) and mean average precision (mAP) for bounding boxes identified by neural networks with IoU over 50 % were used as described in [9]. After image processing and ear counting in the WDS system, we marked the coordinates of the plots. The system counted the number of ears per plot. Knowing that the area of each plot is equal to 24.75 m<sup>2</sup>, we obtain the ear density. Pearson and Spearman correlation coefficients were used to compare the values obtained by manual counting and our system.

*Results:* The results of testing the models of recognition and spike counting in the image are shown in table below.

Model testing was performed on the 27 sets of images added to GWHD in 2021. Each set is provided by one of 10 institutions. The variability of performance metrics and shooting conditions varies greatly. The best accuracy (73.54 on the mAP metric) is provided by the efficient-det model. The arithmetic mean mAP for these samples is 41.40 for faster rcnn and 37.51 for efficient-det. A comparison of the ear density estimates made using our approach and those made manually showed that the Spearman and Pearson coefficient values between them are 0.6176,  $p$ -value = 0.0013 and 0.5405,  $p$ -value = 0.0064 respectively. The structure of developed WDS package is shown below: (A) construction of the flight plan on the quadrangular field by 4 points, (B) marking of plot boundaries, (C) counting the number of ears on each plot.

	frcnn	effdet		frcnn	effdet		frcnn	effdet
ARC_1	39.81	34.63	CIMMYT_3	46.72	33.29	Terraref_1	19.29	14.60
Arvalis_10	50.53	69.74	KSU_1	22.03	38.33	Terraref_2	6.89	4.97
Arvalis_11	29.33	35.06	KSU_2	37.49	39.99	Ukyoto_1	34.25	36.62
Arvalis_12	51.37	40.97	KSU_3	30.52	37.33	ULiège-GxABT_1	25.14	33.68
Arvalis_7	62.79	43.13	KSU_4	13.61	26.00	UQ_10	49.70	39.84
Arvalis_8	57.31	47.16	NAU_2	72.45	69.98	UQ_11	30.18	24.95
Arvalis_9	44.60	36.09	NAU_3	73.26	73.54	UQ_7	50.64	46.49
CIMMYT_1	37.84	20.91	NMBU_1	55.72	37.84	UQ_8	41.23	32.99
CIMMYT_2	63.66	42.13	NMBU_2	33.13	22.17	UQ_9	38.28	30.33



Software package, instructions and scripts for installation are available at [https://github.com/SI07h/wheat\\_detection](https://github.com/SI07h/wheat_detection).

**Conclusion:** We have developed a software package to estimate wheat yield based on counting the number of ears in UAV images of wheat crops, which does not require image stitching. The software package allows to form a flight plan for low altitude flying over the crops (~3 m), to count the number of ears on each image by a deep learning neural network, to link the obtained images to the crop map, and to visualize the density of ears for the studied crops.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Centre of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

## References

1. Jin X. et al. *IEEE Geosci.* 2020;9(1):200-231.
2. Khaki S. et al. Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. arXiv preprint arXiv:2103.09408. 2021.
3. Md Mehedi et al. *Plant Methods.* 2018;14(100).
4. Zhao J. et al. *Remote Sensing.* 2021;13(16):3095.
5. Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS'15, 7-12 Dec 2015, Montreal, Canada. 2015.
6. Tan M., Pang R., Le Q.V. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, CVPR, 14-19 Jun 2020, remote. 2020.
7. David E. et al. *Plant Phenomics.* 2020;2020:3521852.
8. David E. et al. *Plant Phenomics.* 2021;2021:9846158.
9. Yu J., Zhang W. *Sensors.* 2021;21(9):3263.

## Ability to obtain of ITS1 and ITS2 DNA barcode for metabarcoding of historical herbarium specimens

Krinitina A.<sup>1\*</sup>, Samoilov A.<sup>2</sup>, Seregin A.<sup>1</sup>, Zhukova S.<sup>1</sup>, Speranskaya A.<sup>2</sup>

<sup>1</sup> *Lomonosov Moscow State University, Biological Department, Moscow, Russia*

<sup>2</sup> *Research Institute of Systems Biology and Medicine, Moscow, Russia*

\* *krinitina@msu-botany.ru*

**Key words:** historical herbarium, DNA metabarcoding

*Motivation and Aim:* In the archival premises of the botanical gardens and academical herbariums, a large number of specimens are stored. These samples were collected many years ago. Some of them are valuable as a historical object, because they were collected by famous scientists of the past or they were collected during geographical expeditions of a one-two hundred years ago. However, often the specimens from these collections cannot be used for botanical investigations. Because they were not identified during the expedition and the transport storage conditions of the biomaterial in those ancient times did not allow it to be preserved in good enough quality for taxonomy identification of specimens.

Hugh Cuming's collection of flowering plants kept in the Moscow University Herbarium (MW) contains 420 type specimens, 403 non-type specimens and 106 specimens of unknown status. More than 120 specimens from the Philippines and Malaysia have not even been identified.

High throughput sequencing is increasingly being used by biological laboratories associated with field of plants in routine practice. DNA metabarcoding has been actively developing recently, allowing qualitative (to the level of species or genus for some taxa) analyze of complex biological mixes. It employs high-throughput sequencing (HTS) and comparative analysis of specific DNA sequences called "DNA barcodes" to discriminate species present in the mix. DNA barcoding has been widely used in various areas of botanical research, for example authentication and identification of medicinal or archeological plants [1, 2].

*Methods and Algorithms:* DNA from 122 herbarium samples (Cuming's collection from the Philippines and Malaysia) was extracted using modified CTAB-methods [3]. After quality (including DNA fragment length analysis) and quantity evaluation DNA samples were purified using AmPure magnetic beads (Beckman Counter, UK). A simplified two-step PCR using primers for DNA barcodes (ITS1 and ITS2) fused with Illumina adaptor sequences was performed for DNA library preparation as described elsewhere [4, 5]. Products of the first PCR for each barcode were mixed equimolar (or by volume if product concentration was below detection level) for each sample, indexed in the second PCR, and sequenced on the MiSeq platform with the MiSeq Reagent Kit v3 for 600 cycles (2 × 300 nt paired-end) (Illumina, USA).

Raw sequencing reads were trimmed using the trimmomatic software v.0.38 [6]. *De novo* assembled contigs (SPAdes 3.14) using for determinate of taxonomic classification by BLAST-based bioinformatic pipeline.

*Results:* The average concentration of the extracted DNA was 10.02 ng/μL (min 0,1 ng/μL, max 43,8 ng/μL). The average length of DNA fragments was in the 15–150 bp range and 38 % of samples contained DNA fragments about 2,500 bp long.

We were able to prepare the amplification product of the ITS1 and ITS2, the yield of which allowed us to use it for the preparation of libraries, for 59 samples. The amplification of ITS1 was more successful. The product was obtained for 59 samples, while ITS2 amplification was worse: only 29 samples were able to obtain the product. For these samples, the product corresponding to both (ITS1 and ITS2) marker was obtained as a result. There were no samples for which only the ITS2 was obtained. After treatment the sequencing data, assembled contigs length of 44 samples corresponded to the length of the product obtained. However, for 15 samples, it was not possible to assemble fragments using the applied pipelines. It is possible that the use of other sample assembled methods will make it possible to obtain DNA fragments suitable for further analysis.

*Conclusion:* Thus, we can say that ITS1 and ITS2 DNA metabarcoding can be used even for samples that have been stored for more than 185 years, even though DNA becomes highly fragmented during storage.

*Acknowledgements:* Acknowledgement to D. Omelchenko for help with sample sequencing. This research was funded by Russian Foundation for Basic Research, project No. 20-29-01039.

### *References*

1. Techen N. et al. DNA Barcoding of Medicinal Plant Material for Identification. *Curr Opin Biotechnol.* 2014;25:103-110.
2. Gismondi A. et al. Grapevine Carpological Remains Revealed the Existence of a Neolithic Domesticated *Vitis Vinifera* L. Specimen Containing Ancient DNA Partially Preserved in Modern Ecotypes. *J Archaeo. Sci.* 2016;69:7584.
3. Doyle J.J., Doyle J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin.* 1987;19:11-15.
4. Speranskaya A.S. et al. Comparative Analysis of Illumina and Ion Torrent High-Throughput Sequencing Platforms for Identification of Plant Components in Herbal Teas. *Food Control.* 2018;93:315-324.
5. Omelchenko D.O. et al. Improved Protocols of ITS1-Based Metabarcoding and Their Application in the Analysis of Plant-Containing Products. *Genes.* 2019;10:122.
6. Bolger A.M. et al. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics.* 2014;30:2114-2120.

## Development of CRISPR/Cas9-based genome editing constructs for Nevsky and Udacha potato cultivars

Larichev K.<sup>1,2\*</sup>, Sergeeva E.<sup>1,2</sup>, Karetnikov D.<sup>1,2</sup>, Salina E.<sup>1,2</sup>, Kochetov A.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *k.larichev@g.nsu.ru*

**Key words:** CRISPR/Cas9, guide RNA design, potato, U6 promoter

*Motivation and Aim:* Potato (*Solanum tuberosum*) is a staple food crop that has a significant impact on the world's food supply, warranting the continuous development of potato's breeding techniques. Apart from being autotetraploid and heterozygous, there is a large degree of heterogeneity between various cultivars of potato. Because of this, utilization of conventional breeding techniques is very challenging, which complicates development of potato cultivars with desirable properties. CRISPR-based methods of genome editing are able to address these issues by allowing breeders to introduce new traits without disrupting the existing genetic background. As a result, currently CRISPR/Cas9 editing is the most promising approach to developing new commercially viable potato varieties. However, high heterogeneity and heterozygosity of potato still complicate the use of CRISPR based methods of genome editing, as they rely on having the exact sequence of target region, and, in case of potato, publicly available reference sequences may differ significantly from the sequence of a particular cultivar.

In this work, we have focused on designing guide RNAs (gRNAs) to induce knock-out mutations in granule-bound starch synthase (GBSS) gene of Russian potato cultivars Nevsky and Udacha. We have also aimed at constructing plasmid vectors for efficient gRNA expression in potatoes by using cultivar-specific U6 snRNA promoters. The GBSS gene was chosen as it is responsible for amylose content in starch, and its knock-out is known to result in amylose-free starch phenotype [1–3]. This amylose-free starch is better suited for certain industrial applications than the wild-type starch.

*Methods and Algorithms:* Genomes of cultivars Nevsky and Udacha were sequenced on the Illumina MiSeq platform. Sequences of GBSS gene from Nevsky and Udacha were derived from genome assemblies using BLASTn algorithm and a GenBank GBSS sequence X58453 as a reference. In order to find gRNA target regions in GBSS (13 exons, 2948 bp length), which would be identical for all the alleles in both cultivars, we aligned Nevsky and Udacha GBSS sequences with reference GenBank sequences A23741 and X58453 using UGene (alignment algorithm ClustalΩ). Regions that were identical in all sequences were chosen as query sequences for gRNA design. gRNA design was performed in three programs: CRISPOR, CRISPRdirect and CRISPR-P 2.0. We only selected gRNAs which had the highest specificity according to all three programs. The secondary structure of selected gRNAs was analyzed in RNAfold. Only gRNAs with optimal secondary structure (mostly free 5' end, three characteristic stem loops) were chosen [4]. As a reference, we also used gRNA to first exon designed by Johansen et al [3]. The lack of heterogeneity in sites, targeted by designed gRNAs was further confirmed using data from Illumina reads of PCR-amplified GBSS fragments from cultivars Nevsky and Udacha. gRNAs were either synthesized with GeneArt gRNA

Precision Synthesis Kit (Invitrogen), or by *in vitro* transcription from pUC57-sgRNA plasmid [5]. gRNA cleavage efficiency was tested *in vitro* and *in vivo*. For *in vitro* assay we used PCR-amplified fragments of GBSS gene containing the target site. Fragments were incubated with Cas9 protein and one of gRNAs. Cleavage efficiency was determined by gel-electrophoresis in agarose gel. For *in vivo* assay, protoplasts were transformed with either ribonucleoprotein complexes (RNP) of Cas9 and gRNA, or with pHDE-35S-Cas9-mCherry plasmid, encoding Cas9 protein and one of gRNAs. To determine efficiency of transformation, DNA of transformed protoplasts was extracted and used as a template for PCR-amplification of target regions. These PCR products were sequenced with either Sanger sequencing, or on the Illumina MiSeq platform. To drive more efficient gRNA expression from plasmid vectors in potatoes we used potato U6 snRNA promoters. U6 promoter sequences of cultivars Nevsky and Udacha were acquired from genome assemblies of these cultivars using BLASTn algorithm and GenBank potato U6 promoter reference sequences Z17290, Z17292, Z17293, Z17301. Identified promoter variants were generated by PCR-amplification and inserted into pAtU6-sgRNA plasmid instead of *Arabidopsis thaliana* U6 promoter by NEBuilder HiFi DNA Assembly (New England Biolabs), effectively making it into a pStU6-sgRNA plasmid. The entire gRNA expression cassettes (U6 promoter, gRNA, U6 terminator) were then inserted in pHDE-35S-Cas9-mCherry binary vector [6], suitable for both biolistic and *Agrobacterium*-mediated transformations.

**Results:** We designed five gRNAs with high predicted specificity: three for the first exon, one for the second and one for the tenth. *In vitro* efficiency was the highest for gRNAs 3, 4 and 5. None of our gRNAs performed intended cleavage *in vivo* in PEG-mediated RNP transformation of potato protoplasts. Investigation of plasmid-based transformation efficiency is still ongoing. We identified five distinct U6 promoter variants in both cultivars. Promoter sequences varied between cultivars. Each promoter was used to create new vectors – pStU6-sgRNA, suitable for gRNA expression in plants. Efficiency of identified promoters is currently being tested.

**Conclusion:** Using assembled genomes of cultivars Nevsky and Udacha we were able to design five gRNAs targeting GBSS gene with high predicted specificity, and to identify five distinct variants of U6 promoter in each cultivar. The latter were used to construct new expression vector – pStU6-sgRNA, which can be used to efficiently express gRNAs with a specific target sequence in potato plants. It is worth noting that currently there are no plasmid vectors containing potato U6 promoter available for purchase, which makes our vector especially valuable.

**Acknowledgments:** The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

## References

1. Veillet F. et al. The *Solanum tuberosum* GBSSI gene: a target for assessing gene and base editing in tetraploid potato. *Plant Cell Rep.* 2019;38(9):1065-1080.
2. Andersson M. et al. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* 2017;36(1):117-128.
3. Johansen I.E. et al. High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. *Sci Rep.* 2019;9(1):1-7.
4. Liang G. et al. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci Rep.* 2016;6(1):1-8.
5. Shen B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods.* 2014;11(4):399-402.
6. Gao X. et al. *Plant Physiol.* 2016;171(3):1794-1800.

# Pangenome construction and analyses for a population of flax parasitic fungus *Fusarium oxysporum* f. sp. *lini*

Logachev A.<sup>1\*</sup>, Kanapin A.<sup>2</sup>, Samsonova A.<sup>2</sup>, Bankin M.<sup>1</sup>, Stanin V.<sup>1</sup>, Rozhmina T.<sup>3</sup>, Samsonova M.<sup>1</sup>

<sup>1</sup> *Mathematical Biology & Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>2</sup> *Center for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>3</sup> *Laboratory of Breeding Technologies, Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

\* *loga4ov@gmail.com*

**Key words:** population genomics, pangenome, *Fusarium*, fusariosis, flax, virulence

Pangenome usually defined as a full gene repertoire of a species, infraspecific taxa, or population [1]. Upon this conception pangenome divides on core genes, shared between all individuals, and accessory genes, which are absent in one or more individuals [2]. Application of this term to microorganisms is very useful since some of them have significant variation in gene content between individuals within a species. Sordariomycete, *Fusarium oxysporum* Shlecht., is widely distributed soil-born fungus that may appear as saprophyte, or have parasitic lifestyle, infecting different plants, including economically important crops. Distinction between saprotrophic and parasitic forms of this species lies in different gene content possessing by this forms. Phytopathogenic special forms (*forma specialis* – *f.sp.*) of *F. oxysporum* have very strict specificity to their host-plants, that is also determined by differences in genomes, and more than 200 of special forms are described by today. Earlier we reported about sequencing and analysis of 5 *Foli* genomes with chromosome-level assemblies and described that karyotype of *Foli* divides at least on two parts: conserved chromosomes composed on sequences with shared synteny to other special forms, and variable chromosomes that lack of such synteny [3, 4].

Here we report the pangenome construction and characteristics for *F. oxysporum* f.sp. *lini* (*Foli*), that infects flax and cause fusariose wilt of plant, based on full genome sequences of 13 strains. We conducted a population genomic analysis of different pangenome gene categories. We determined a repertoire of *SIX* (*Secreted In Xylem*) genes, and virulence of these strains. We also determined MAT-idiomorphs and vegetative compatibility (VC) groups. To discover the relations between observed *Foli* strains and to other special forms we inferred phylogenies on selected single genes, SNPs, and on protein-coding genes.

**Methods:** We sequenced 3 genomes of low-virulent *Foli* isolates with Illumina technology, and one of them was also sequenced with PacBio technology. Together with 5 genomes we published earlier [3] and 5 other available *Foli* genomes [5], these 3 new genomes were included in population genomic analysis. All investigated 13 isolates were provided by Federal Research Center for Bast Fiber Crops. Pangenome construction was carried out based on genes identified in complete genomes of 13 *Foli* isolates. We performed resampling of all possible combinations of 1–13 genomes and modelled the core genome and pangenome size curves [2]. We also conducted population genomic

analyses of different gene categories according to [6, 7]. Chromosomal synteny between the genomes was analysed by blastn pairwise alignment and visualised with R.

*SIX* genes (*SIX1* – *SIX14*), coding small effector proteins, were determined by blastn searching for homologous sequences on 13 *Foli* genomes. The results of the search were then confirmed by PCR with respective primers for each of 14 *SIX* genes on genomic DNA of 11 strains. To investigate the expression level of *SIX* genes during colonisation of plant we used RNAseq data on mycelium culture, 3rd and 5th day of flax infection, obtained earlier from infection experiments with flax seedlings. Phylogenetic trees were constructed for each *SIX* gene found in our 13 strains with IQ-TREE.

MAT-idiomorphs were determined by blastn search through all 13 genomes with known sequences of MAT-locuses for *F. oxysporum* (GeneBank: AB011379 for MAT1-1 idiomorph, and AB011378 for MAT1-2 idiomorph) [8]. Phylogeny was inferred on *MAT1-2-1* gene sequences with JC69 model, and maximum likelihood tree was constructed with IQ-TREE. Expression level of *MAT1-2-1* gene in culture and during infection was determined by RNAseq data. Phylogenomic analyses of 13 *Foli* strains completed with 2 other *Fusarium* genomes used as outgroups were performed with protein sequences found only in one copy in each genome based on a cluster analysis. The identified single-copy orthologs then were aligned and ML phylogeny were inferred using RaxML.

**Results:** Pangenome size estimated on 13 genomes comprises of 14283 genes. We found that the core genome stabilised at approximately 967 genes. Number of accessory genes did not stabilise and increased almost linearly, and was equal to 10337 on 13 genomes. Number of singletons in each genomes varied from 138 to 453. CAZymes were mostly belong to core genes, genes coding secretory proteins relate to core and accessory genes while effectors belong to accessory genes and even singletons. Repertoire of *SIX* genes in *Foli* pangenome comprises *SIX1*, *SIX7*, *SIX10*, *SIX12* and *SIX13*. Two analysed isolates had no *SIX* genes at all, four isolates did not possess *SIX7* and *SIX12*, one isolate lack of only *SIX7* gene. Phylogenetic analysis on found *SIX* genes shown clear clusterization of strains by their special forms.

We found that 12 of 13 *Foli* isolates shared MAT1-2 locus with only one intact *MAT1-2-1* gene with 2 introns exactly as appear in other heterothallic sordariomycetes. One isolate had MAT1-1 idiomorph and possessed 3 genes inside the locus: *MAT1-1-1*, *MAT1-1-2* and *MAT1-1-3*. Phylogeny that was inferred for 12 MAT1-2 strains on *MAT1-2-1* gene depicts real origin of investigated strains and demonstrates 2 large distinct clades that include 11 *Foli*, and 1 small clade with only one *Foli* strain. Moreover, we demonstrate that *MAT1-2-1* gene do expressed in mycelium of MI39 isolate despite *Foli* does not have reported sexual reproduction.

**Conclusions:** We found that in *Foli* pangenome accessory genes occupied approximately 72 % of all genes, while core genes comprises only 6 %, and remaining 22 % were singletons. This fact reveals that *Foli* has highly plastic genome with prevalence of genes not shared with all individuals in the population. Isolates with high virulence have set of 5 different *SIX* genes while low-virulence strains lack at least *SIX7* gene. Phylogenomic analyses shown that investigated *Foli* population divides on 2 big distinct clades and 2 small clades each consisted each only of one *Foli* isolate. MAT-locus analyses reveal that all individuals possessed MAT1-2 idiomorph and the population is quite homogenous on mating type. Our first attempt to construct pangenome for *F. oxysporum f.sp. lini* provides new insights into genomic studies on populations and give rise to further investigations on phytopathogenic fungi.



*Acknowledgements:* The study is supported by the Russian Science Foundation, grant No. 19-16-00030.

### *References*

1. Hurgobin B., Edwards D. SNP discovery using a pangenome: has the single reference approach become obsolete? *Biology*. 2017;6(1):21.
2. Plissonneau C., Hartmann F.E. Croll D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology*. 2018;16(1):1-16.
3. Kanapin A., Samsonova A., Rozhmina T., Bankin M., Logache A., Samsonova M. The genome sequence of five highly pathogenic isolates of *Fusarium oxysporum* f. sp. lini. *Mol Plant-Microbe Interactions*. 2020;33(9):1112-1115.
4. Samsonova A., Kanapin A., Bankin M., Logachev A., Gretsova M., Rozhmina T., Samsonova M. A genomic blueprint of flax fungal parasite *Fusarium oxysporum* f. sp. lini. *Int J Mol Sci*. 2021;22(5):2665.
5. Dvorianinova E.M., Pushkova E.N., Novakovskiy R.O., Povkhova L.V., Bolsheva N.L., Kudryavtseva L.P., Rozhmina T.A., Melnikova N.V., Dmitriev A.A. Nanopore and Illumina genome sequencing of *Fusarium oxysporum* f. sp. lini strains of different virulence. *Front Genet*. 2021;12:662928.
6. Badet T., Oggenfuss U., Abraham L., McDonald B.A., Croll D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology*. 2020;18(1):1-18.
7. Alouane T., Rimbert H., Bormann J. et al. Comparative genomics of eight fusarium graminearum strains with contrasting aggressiveness reveals an expanded open pangenome and extended effector content signatures. *Int J Mol Sci*, 2021;22(12):6257.
8. Yun S.H., Arie T., Kaneko I., Yoder O.C., Turgeon B.G. Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. *Fungal Genet Biol*. 2000;31(1):7-20.

# Amyloid formation is a widespread phenomenon in land plant species

Nizhnikov A.A.<sup>1,2\*</sup>, Antonets K.S.<sup>1,2</sup>

<sup>1</sup> All-Russia Research Institute of Agricultural Microbiology, St. Petersburg, Russia

<sup>2</sup> St. Petersburg State University, St. Petersburg, Russia

\* a.nizhnikov@arriam.ru

**Key words:** amyloid, plant, protein structure, protein fibril, protein aggregate

**Motivation and Aim:** Amyloids are highly ordered protein aggregates best known for their association with dozens of incurable diseases in humans and animals. However, numerous recent studies have demonstrated the functional significance of amyloid formation in different species of archaea, bacteria, and eukarya. Plants represent a poorly studied group in the field of amyloid biology, in which amyloids had not been identified before our studies.

**Methods and Algorithms:** Different bioinformatic and experimental methods were used as described in [1] and [2].

**Results:** Bioinformatic prediction of the distribution of potentially amyloidogenic regions in proteomes of 76 land plant species have revealed an unusual enrichment of seed storage and defense proteins with such regions [1]. Later, first plant protein that forms amyloids under native conditions *in vivo* has been discovered [2]. Proteomic analysis of seeds from different plant species has revealed a set of proteins forming amyloid-like aggregates *in vivo* (unpublished data).

**Conclusion:** The data obtained suggest that seed storage proteins of different plant species are enriched with potentially amyloidogenic regions. Experimental verification confirmed amyloid properties of several plant proteins *in vivo* [2 and unpublished data]. Overall, these data demonstrate the widespread occurrence and functional significance of the phenomenon of amyloidogenesis in land plants.

**Acknowledgements:** This work was made with the support of the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2020-920, date November 16, 2020 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”.

## References

1. Antonets K.S., Nizhnikov A.A. Predicting amyloidogenic proteins in the proteomes of plants. *Int J Mol Sci.* 2017;18:e2155.
2. Antonets K.S., Belousov M.V., Sulatskaya A.I., Belousova M.E., Kosolapova A.O., Sulatsky M.I., Andreeva E.A., Zykin P.A., Malovichko Y.V., Shtark O.Y., Lykholay A.N., Volkov K.V., Kuznetsova I.M., Turoverov K.K., Kochetkova E.Y., Bobilev A.G., Usachev K.S., Demidov O.N., Tikhonovich I.A., Nizhnikov A.A. Accumulation of storage proteins in plant seeds is mediated by amyloid formation. *PLOS Biology.* 2020;18(7):e3000564.

## Adaptation to polyploidy in Siberian *Arabidopsis lyrata*

Novikova P.Y.

Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

\*pnovikova@mpipz.mpg.de

**Key words:** polyploidy, whole-genome sequencing, adaptation, meiosis, self-compatibility

**Motivation and Aim:** Polyploid organisms have more than two sets of chromosomes due to either whole-genome duplication (WGD) in the same species, giving rise to autopolyploids or whole-genome hybridization in allopolyploids formed from the combination of more than two sets of chromosomes from different species. Polyploidy is a recurrent evolutionary process: most eukaryotes had at least one round of WGD in their ancestry [1] and about 30 % of flowering plants [2] are contemporary polyploids. However, new polyploids are rarely successful in the long run [3, 4], with the majority becoming quickly extinct [5] due to immediate reproductive disadvantages. Using widespread *Arabidopsis lyrata* species with varying ploidy levels as a model, we investigate what makes a polyploid successful and determine how they overcome challenges ranging from cellular to population levels.

**Methods and Algorithms:** We obtained whole-genome sequencing data from herbarium and live samples of *A. lyrata* covering the entire distribution of the species in Eurasia. Using bioinformatic inferences (allele frequencies and S-locus genotyping) and cytological measurements (flow cytometry and karyotyping) we identified ploidy and mating-type of each sample. Using phylogenomic and population genetic approaches we describe evolutionary history of polyploid lineages in *A. lyrata* and *A. kamchatica*. And finally, we performed selection scans measuring divergence between diploid and tetraploid lineages and diversity within the tetraploid lineage to identify potentially adaptive changes linked to increased ploidy level.

**Results:** Our results suggests that *A. lyrata* is the most polyploidy-rich species in *Arabidopsis* genus, and includes multiple autopolyploid lineages in Europe and Siberia. We show that *A. lyrata* lineage that contributed to the origin of *A. kamchatica* was self-compatible, which promoted the establishment of the new allopolyploid. We describe the mechanism of self-incompatibility loss in both *A. lyrata* (Siberian selfing lineage) and *A. kamchatica*. We also show that adaptation to autopolyploidy in Siberian *A. lyrata* took place and involved meiotic pathways, similarly to tetraploid *A. arenosa* in Europe. **Conclusion:** *A. lyrata* is a new exiting model to study how polyploids originate, survive and succeed.

### References

1. Edger P.P., Pires J.C. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* 2009;17:699-717.
2. Van de Peer Y., Mizrahi E., Marchal K. *Nat Rev Genet.* 2017;18:411-424.
3. Mayrose I., Zhan S.H., Rothfels C.J., Magnuson-Ford K., Barker M.S., Rieseberg L.H. et al. Recently formed polyploid plants diversify at lower rates. *Science.* 2011;333:1257.
4. Van de Peer Y., Ashman T.-L., Soltis P.S., Soltis D.E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell.* 2021;33:11-26.
5. Levin D.A. Minority Cytotype Exclusion in Local Plant Populations. *Taxon.* 1975;24:35-43.

## Natural variation of *E1-E4*, *E9*, *GmFT5a* genes in 180 soybean accessions adapted to Novosibirsk regions

Perfl'ev R.<sup>1\*</sup>, Shcherban A.<sup>1</sup>, Potapov D.<sup>2</sup>, Maksimenko K.<sup>2</sup>, Polyudina R.<sup>2</sup>, Salina E.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Siberian Research Institute of Forages, SFSCA RAS, Novosibirsk, Russia

\* Perfl'evRN@bionet.nsc.ru

**Key words:** maturity time, soybean, *E* genes

**Motivation and Aim:** *E1-E4* genes have the major effect on maturity time and provide soybean adaptation to wide range of latitudes. Also, some variations in flowering time are associated with the *E9* and *GmFT5a* genes. The natural variation of the *E1-E4* genes has been widely studied in different germplasm collections of China, Europe, America, India, but not in Russia. The Novosibirsk district is a unique region for soybean production, because in the past there were no adapted varieties for cultivation in this region. At present, in Siberian Research Institute of Forages has bred and selected accessions that are able to fully maturity in such a northern cold region. Which alleles of the *E1-E4*, *E9* and *GmFT5a* genes provide adaptation to these conditions is unknown. Thus, the aim of our study is to reveal which alleles of this genes provide soybean adaptation and variation of maturity time in the Novosibirsk conditions.

**Methods and Algorithms:** Soybean accessions for DNA extraction were provided by Siberian Research Institute of Forages (172) and by EFKO (8) company. The 180 accessions were grown in the experimental field of the Siberian Research Institute of Forages (55°) under natural daylength conditions. The date of the emergence (VE), the first flower appearance (R1) and full maturity (R8) was recorded as the average between of three replicates. We used previously developed molecular markers to determine the alleles of the *E1-E4*, *E9* and *GmFT5a* genes. The box-plot were plotted in R 4.1.

**Results:** We found the high frequency of recessive alleles of *E1-E4* genes (Table 1). All accessions carry the dominant allele *E9*, which reduce the flowering time. For the *GmFT5a* gene, only three cultivars Kassidy, Pripyat and Aktai carry the *GmFT5a-H4* haplotype which associated with early flowering.

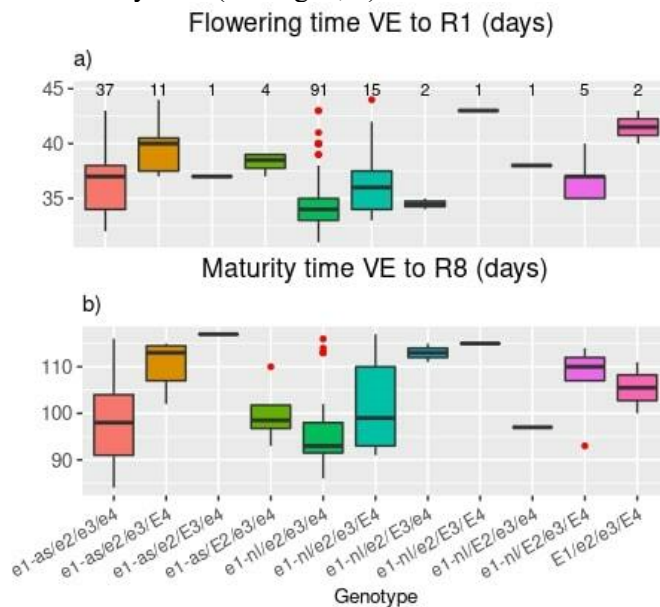
**Table 1.** Defined alleles of genes *E1-E4*

<i>E1</i>	<i>n</i> *	%	<i>E2</i>	<i>n</i>	%	<i>E3</i>	<i>n</i>	%	<i>E4</i>	<i>n</i>	%
<i>e1-nl</i>	112	62,2	<i>e2</i>	168	93,3	<i>e3-fs</i>	122	67,7	<i>e4-SORE-1</i>	93	51,6
<i>e1-as</i>	58	32,2	<i>E2</i>	12	6,6	<i>e3-tr</i>	38	21,1	<i>e4-kes</i>	46	25,5
<i>E1</i>	2	1,1				<i>e3-ns</i>	10	5,5	<i>E4</i>	41	22,7
<i>e1-fs</i>	1	0,55				<i>E3-Ha</i>	8	4,4			
						<i>e3**</i>	2	1,1			

\* – number of accessions, \*\* – mix of haplotypes *e3-fs* and *e3-tr*

In field conditions, 10 varieties did not reach full maturity and were dropped out from further analysis. Almost all of these cultivars carry dominant alleles *E3* or *E4*. Recessive

alleles of the *E1-E4* genes, which reduce the maturity time, have a very high frequency in studied soybean population (see Table 1). Accessions with *e1-nl/e2/e3/e4* and *e1-as/e2/e3/e4* genotype have the earliest flowering time (Fig. 1a) and maturity (see Fig. 1, b). Accessions carrying dominant alleles *E3* or *E4* genes have the latest flowering time (see Fig. 1, a) and maturity time (see Fig. 1, b).



**Fig. 1.** a) Flowering time (VE to R1) of 170 cultivars of different *E* genotype, the numbers under boxes are the number of cultivars in each *E* genotype. b) Maturity time (VE to R8) of 170 cultivars of different *E* genotype

**Results:** Recessive alleles *E1-E4* predominate in the studied accessions that proves their key role in soybean adaptation to northern latitudes according with previous studies [1]. Rare for foreign soybean collections *e1-nl*, *e3-fs* and *e4-kes* alleles were identified in a large number of accessions. Three cultivars carry the rare *GmFT5a-H4* allele and can be used as a source of this rare allele in breeding. Different flowering time between genotypes (*e1-nl/e1-as*)/*e2/e3/e4* and (*e1-nl/e1-as*)/*e2/e3/E4* can be explained by the influence of the *E4* gene on flowering time. In previous studies *E4*, under natural conditions, effects on the post-flowering development [2], possibly in the Novosibirsk conditions, *E4* also effects on flowering time, as in experiments with 20-hour artificial daylight [3]. For the Novosibirsk region, varieties with the shortest maturity time are required. Accessions with genotypes *e1-nl/e2/e3/e4* and *e1-as/e2/e3/e4* have the earliest maturity time, so these genotypes can be used as a “molecular model” in soybean breeding to create varieties adapted to Novosibirsk conditions.

**Acknowledgements:** This work was supported by the Russian Science Foundation, project 21-76-30003.

## References

1. Jia H. et al. Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. *PLoS One*. 2014;9(4):e94139.
2. Xu M. et al. Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biol*. 2013;13:91.
3. Cober E.R. et al. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci*. 1996;36:601-605

## GBS assaying of F1 population obtained from self-pollination of muscadine grapevine cv. Dixie (*Muscadinia rotundifolia* Michaux.)

Potokina E.K.<sup>1,2\*</sup>, Karzhaev D.S.<sup>1,2</sup>, Lytkin K.F.<sup>1,2</sup>, Grigoreva E.A.<sup>1</sup>, Volkov V.A.<sup>1,2</sup>, Volynkin V.A.<sup>1</sup>, Nosulchak V.A.<sup>1</sup>, Lushchay E.A.<sup>1</sup>, Likhovskoi V.A.<sup>1</sup>

<sup>1</sup> All-Russian National Research Institute of Viticulture and Winemaking "Magarach", RAS, Yalta, Russia

<sup>2</sup> St. Petersburg Forest Technical University, St. Petersburg, Russia

\* e.potokina@yahoo.com

**Key words:** muscadine grapevine, F1 population, GBS, linkage map

Since phylloxera and mildew were brought from North America to Europe in the second half of the 19th century, breeders around the world have been introducing resistance genes from American and Asian wild species into the gene pool of European grapevines. During this time, significant success was achieved, and new varieties were obtained that give high quality vines and have a high degree of resistance to powdery mildew. Sources of pathogen resistance genes were mostly *Vitis* species from North America (*V. riparia*, *V. labrusca*, *V. aestivalis* and *V. berlandieri*) and, more recently, the Far East species (*V. amurensis*). All these sources provide only partial resistance of the cultivated grapevine, while loci introgressed from *Vitis* (*Muscadinia*) *rotundifolia* confer complete resistance to pathogens in all studies known to date. Recently, a high-resolution genetic map conferring 2069 SNP markers has been published for the *Muscadinia rotundifolia* (cv Traished), after GBS (Genotyping-By-Sequencing) assaying of two populations of F1 hybrids.

Only one variety of *M. rotundifolia* (cv Dixie) was introduced in Russia, where it has been successfully propagated in Krasnodar region and Crimean Peninsula. There, cv. Dixie bears fruit annually as a result of self-pollination, providing a natural source of F1 hybrids. Evaluating of the F1 hybrid population revealed biological variability of some morphological traits, e.g., grown power, weight and shapes of berries.

In this study we present result of GBS assaying of 134 F1 hybrids obtained from self-pollination of cv. Dixie. Number of detected heterozygous loci and general heterozygosity level is described. During self-pollination of Dixie variety, the process of formation of female and male gametes occurs leading to recombination events and producing new combinations of homozygous and heterozygous loci, different from the parent genotype. Counting the frequency of such recombination events between heterozygous loci in a population of Dixie F1 hybrids allowed us to calculate the genetic distances between SNP markers and construct a linkage map.

**Acknowledgement:** This work was supported by the Russian Science Foundation (Project No. 20-16-00060).

## Transcriptome analysis of flax seeds and capsules

Povkhova L.V.<sup>1,2\*</sup>, Novakovskiy R.O.<sup>1</sup>, Rozhmina T.A.<sup>1,3</sup>, Dvorianinova E.M.<sup>1,2</sup>, Frykin R.I.<sup>1,4</sup>, Pushkova E.N.<sup>1</sup>, Dmitriev A.A.<sup>1</sup>, Melnikova N.V.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup> Federal Research Center for Bast Fiber Crops, Torzhok, Russia

<sup>4</sup> Lomonosov Moscow State University, Moscow, Russia

\* povhova.lv@phystech.edu

**Key words:** flax, fatty acids, lignans, transcriptome sequencing

*Motivation and Aim:* Seeds of flax (*Linum usitatissimum* L.) are a rich source of unsaturated fatty acids (linolenic, linoleic, and oleic acids) and lignans. Linseed varieties differ significantly in fatty acid and lignan content. The pathways of lignan and fatty acid biosynthesis are known to include several key genes, however, molecular mechanisms of their regulation have yet to be understood. This work aims to identify the diversity of molecular mechanisms that determine the fatty acid composition of linseed oil and lignan content in flax seeds. An effective approach to resolve multiple issues of fatty acid and lignan synthesis is transcriptomic studies of different flax varieties with contrasting fatty acid and lignan content. Here, we perform such an analysis for several representative flax varieties at different stages of seed development.

*Material and Methods:* Cultivars and lines for the research were obtained from the collection of the Institute for Flax (Torzhok, Russia). RNA was extracted independently from seeds and capsules using the Quick-RNA Miniprep Kit (Zymo Research, USA). The concentration and quality of the isolated RNA were assessed using a Qubit 4 fluorometer (Thermo Fisher Scientific, USA) and 2100 Bioanalyzer (Agilent Technologies, USA). cDNA libraries for transcriptome sequencing were prepared using the QIAseq Stranded mRNA Select Kit (Qiagen, USA). The quality of the cDNA libraries was assessed with 2100 Bioanalyzer, and high-quality libraries were sequenced on the Illumina platform.

*Results:* We studied a representative set of flax varieties, covering the diversity of this agriculture in terms of the fatty acid composition of oil and the content of lignans in seeds. Nine varieties of flax were grown in climatic chambers under three different temperature and watering conditions: control ones; low temperatures and abundant watering; high temperatures and poor watering. For the analysis, we used only typical plants (at least 10 for each cultivar/line) selected using the data on key polymorphisms obtained by targeted sequencing or Cleaved Amplified Polymorphic Sequence (CAPS) analysis. This step helped to avoid erroneous results due to varietal heterogeneity, which is extremely important for an experiment to be reliable. The capsules and seeds were gathered on the 3<sup>rd</sup>, 7<sup>th</sup>, 14<sup>th</sup>, 21<sup>st</sup>, and 28<sup>th</sup> day after flowering. We performed a transcriptome sequencing of 270 cDNA libraries: 9 cultivars/lines, 3 temperature/watering modes, 2 types of material (capsules and seeds), 5 collection periods). We analyzed the changes in gene expression profiles during the development of flax seeds and revealed genes with differential expression between varieties, development stages, and growth conditions.

*Conclusion:* The obtained results will make it possible to identify the diversity of mechanisms determining the fatty acid composition of linseed oil and lignan content in flax seeds, suggest molecular markers associated with the seed characteristics, and develop recommendations for their use in marker-assisted flax selection for the use in food, pharmaceutical, or other industries.

*Acknowledgements:* This work was funded by the Russian Science Foundation according to the research project 21-16-00111.



## Transcriptomic analysis of generative and vegetative organs of male and female poplar plants

Pushkova E.N.<sup>1\*</sup>, Novakovskiy R.O.<sup>1</sup>, Povkhova L.V.<sup>1</sup>, Kostina M.V.<sup>2</sup>,  
Melnikova N.V.<sup>1</sup>, Dmitriev A.A.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Moscow Pedagogical State University, Institute for Biology and Chemistry, Moscow, Russia

\* pushkova18@gmail.com

**Key words:** *Populus*, poplar, transcriptome sequencing, gene expression, generative organs, vegetative organs

**Motivation and Aim:** Poplars (genus *Populus*) belong to dioecious species, in which male and female flowers are formed on separate plants. In recent years, large-scale studies of the genus *Populus* species were performed, and genes that play a key role in sex determination were identified. However, the data on the number of these genes diverge in the works of different authors. It is also not clear which pathways are involved in the development of male and female flowers. Moreover, contradictory data have been obtained on the differences between male and female plants, which are not limited to generative organs. This work aims at identifying the key mechanisms involved in sex determination of *Populus* species and the development of generative and vegetative organs of male and female poplar plants.

**Methods and Algorithms:** A set of generative and vegetative organs at 7 time points of two male and two female *Populus × sibirica* plants (total 56 samples) was collected in Moscow. Harvested generative and vegetative organs were immediately frozen in liquid nitrogen and then stored at -70 °C until RNA was extracted. Using the MagNA Lyser system (Roche, Switzerland), we homogenized 0.2 g of each *P. × sibirica* leaf, flower, and catkin axis samples during 90 s at 700 rpm in 600 µl of RNA Lysis Buffer from the Quick-RNA Miniprep Kit (Zymo Research, USA) and in the presence of Solid-glass beads (Sigma-Aldrich, USA) being 2 mm in diameter and having a mass of 0.14 g. Then, the homogenate was cooled. Total RNA was extracted using the Quick-RNA Miniprep Kit (Zymo Research) according to the standard protocol of the manufacturer, including the step of treatment by DNase I. The quality and concentration of the RNA were evaluated using a 2100 Bioanalyzer instrument (Agilent Technologies, USA) and Qubit 4 fluorometer (Thermo Fischer Scientific, USA) respectively. cDNA libraries for the high-throughput sequencing were prepared with the RNeasy Plant Mini Kit (Qiagen, USA) for the isolation of mRNA fraction from 1 µg of the total RNA and the QIAseq Stranded Total RNA Library Kit (Qiagen) for the subsequent RNA fragmentation and adapter ligation according to the manufacturer's protocols. The quality of the cDNA libraries was evaluated using a 2100 Bioanalyzer instrument (Agilent Technologies), and concentration – using a Qubit 4 fluorometer (Thermo Fischer Scientific). The cDNA libraries were sequenced on NextSeq500 (Illumina, USA) with a read length of 86 nucleotides.

**Results:** On the Illumina platform, we performed transcriptome sequencing of 56 cDNA libraries: two male and two female *P. × sibirica* plants, generative and vegetative organs, 7 stages of development: the end of June-beginning of July (the next-year bud begins to

form), the end of July-beginning of August (flower bud has already appeared, but there is no visible difference between male and female inflorescences), end of August-beginning of September (differences between male and female inflorescences are visible), end of September-beginning of October (rudiments of gynoecium and stamens become visible), end of October-beginning of November (carpels and stamens are formed, but have a small size), end of March-beginning of April (generative organs slightly increase in size), end of April-beginning of May (flowers are fully formed). The collection of plant material during these time points covers the main development stages of generative and vegetative poplar buds. We carried out a bioinformatics analysis of the obtained data and identified genes that are expressed at different development stages in the generative and vegetative organs of male and female *P. × sibirica* plants, as well as genes with sex-differential expression. Expression profiles were more similar for vegetative buds of male and female plants than for generative ones.

*Conclusion:* We obtained the expression profiles for the main development stages of generative and vegetative buds of male and female *P. × sibirica* plants and identified differentially expressed genes. The comparison of our results with the results of studies on other species of the genus *Populus* will allow establishment of the gene pathways involved in the development of generative and vegetative organs in male and female poplars.

*Acknowledgements:* This research was funded by RFBR according to the research project 20-34-90159.

# The divergence of wheat 5B chromosomes during evolution and its impact on recombination suppression

Salina E.<sup>1,2\*</sup>, Muterko A.<sup>1</sup>, Kiseleva A.<sup>1,2</sup>, Korol A.<sup>3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* salina@bionet.nsc.ru

**Key words:** wheat, dicoccoides, recombination suppression, 5B pseudomolecules, transposable elements

*Motivation and Aim:* Chromosomal rearrangements that lead to recombination suppression can have a significant impact on speciation, and they are also important for breeding. In the present work, we evaluated the key points of chromosome 5B reorganization throughout evolution by identifying regions of recombination suppression in interspecific crosses. Sequence analysis of 5B pseudomolecules of *T. aestivum* ( $2n = 6x$ ) and *T. dicoccoides* ( $2n = 4x$ ) in the studied regions on 5B allowed us to identify possible determinants of recombination suppression.

*Methods and Algorithms:* Multiple sequence alignment was carried out using MAFFT 7.311 software.

*Results:* The regions of recombination suppression in wheat chromosome 5B were identified based on comparisons of the 5B map of a cross between the Chinese Spring (CS) variety of hexaploid wheat and CS-5Bdic (genotype CS with 5B substituted with its homologue from tetraploid *Triticum dicoccoides*) with several 5B maps of tetraploid and hexaploid wheat. In total, two regions were selected in which recombination suppression occurred in cross CS x CS-5Bdic when compared with other maps: one on the short arm, 5BS\_RS, limited by markers BS0009810/BS00022336, and the second on the long arm, 5BL\_RS, between markers Ra\_c10633\_2155 and BS00087043. The regions marked as 5BS\_RS and 5BL\_RS, with lengths of 5 and 3.6 Mb respectively, were mined from the 5B pseudomolecule of CS and compared to the homoeologous regions (7.6 and 3.8 Mb respectively) of the 5B pseudomolecule of Zavitan (*T. dicoccoides*). It was shown that, in the case of 5BS\_RS, the local heterochromatin islands determined by the satellite DNA (119.2) and transposable element arrays, as well as the dissimilarity caused by large insertions/deletions (chromosome rearrangements) between 5BSs *aestivum/dicoccoides* are likely the key determinants of recombination suppression in the region. Two major and two minor segments with significant loss of similarity were recognized within the 5BL\_RS region.

*Conclusion:* The loss of similarity, which can lead to suppression of recombination is caused by chromosomal rearrangements, driven by the activity of mobile genetic elements (both DNA transposons and long terminal repeat retrotransposons) and their divergence during evolution. The revealed divergence may be a consequence of interspecific hybridization, plant genetic adaptation, or both.

*Acknowledgements:* The research was funded by the Kurchatov Genomic Center of ICG SB RAS (075-15-2019-1662). Genetic linkage maps and comparative chromosome analysis were supported by RSF (Russian Science Foundation) project No. 21-76-30003.

## Natural variations in telomere lengths in different Bryophytes

Sannikova A.V.<sup>1\*</sup>, Sharipova M.R.<sup>1</sup>, Shakirov E.V.<sup>1,2</sup>, Valeeva L.R.<sup>1</sup>

<sup>1</sup> Kazan Federal University, Kazan, Russia

<sup>2</sup> Department of Biological Sciences, College of Science, Marshall University, Huntington, USA

\* [anastasya.sannikova@bk.ru](mailto:anastasya.sannikova@bk.ru)

**Key words:** telomeres, bryophytes, moss, *Physcomitrella patens*, *Ceratodon purpureus*, *Sphagnum*

*Motivation and Aim:* Telomeres are specialized nucleoprotein structures at the ends of linear chromosomes in most of eukaryotes. Telomeres play an important role in protecting of the chromosomes termini from exonucleases and cell repair systems, as well as from chromosomal aberrations. In this case, the tandem telomere repeats acts as a "buffer zone" to prevent the loss of coding DNA. Telomeric DNA is shortened by the every cell division that leads to the irresistible approaching of the critical telomere length for the telomere protein complex formation and their functioning. Therefore, the telomere length is the main limitation and determination factor of chromosome protection and cell fate. On the other hand, the length of telomeres is a species-specific feature that varies in populations and individual organisms, as well as in various tissues and even on different arms of chromosomes. Plants are actually highly promising group of organisms for molecular genetics studies of evolutionary conservative mechanisms due to their sessile lifestyle and arising under it extremely flexible genetic regulation. The telomeric DNA of most plants consists of a palindrome motif (TTTAGGG)<sub>n</sub> that differs from the human telomeric repeat only for one nucleotide. In that case, bryophytes (mosses, liverworts, and hornworts) are one of the perspective groups of plants for telomere regulation studies also due to their phylogenetic position. Their study provides a unique opportunity to trace the evolution of genes associated with telomeres from the first primitive terrestrial plants to the most developed angiosperms. In addition, the slow evolution of bryophytes' genomes and its less redundancy, the predominance of the haploid gametophyte over the diploid sporophyte in lifecycles, and also simple handling in laboratory conditions, makes bryophytes one of the best model plant organisms. The aim of the current work was to determine and characterize the natural variations of telomere lengths in different bryophyte species and ecotypes.

*Methods and Algorithms:* In the study we used axenic cultures of mosses: *Physcomitrium patens*, 4 ecotypes (Gransden, Great Britain; Reute, Germany; Villersexel, France; Kaskaskia, USA); *Ceratodon purpureus* (male R40 and female GG1 lines), and natural isolates of peatmoss *Sphagnum fallax* MW (USA), *Sphagnum girgensohnii* (Russia, Sverdlovsk region) and *Sphagnum sp.* (Russia, Republic of Mari-El). Mosses were grown on a solid mineral medium BCD with the addition of ammonium tartrate. To analyze the telomere length, genomic DNA isolated from 14-day-old protonema of *P. patens* and *C. purpureus* or *Sphagnum* gametophores were used. Telomere length analysis was performed by TRF (Terminal Restriction Fragment analysis) coupled with Southern blot analysis. The average telomere length was calculated using the TeloTool software.

*Results:* We have shown that different ecotypes of *P. patens* plants have a telomere length in the range from 1000 to 1500 bp. It was 1.5-3 times shorter than the telomere

length of the model angiosperm plant *A. thaliana* Col-0 (2500-4500 bp in the average). The average telomere length in the Gransden, Reute and Villersexel ecotypes was ~ 1300 bp. However, the average telomere length in the Kaskaskia ecotype was slightly shorter (~ 1200 bp). In addition, we have shown that all the ecotypes studied have specific telomeric sequences, presumably of intra-chromosomal localization, and differing in their location and length within the ecotypes.

It was shown that the telomere length of the female plant *C. purpureus* GG1 ranges from 480 bp to 500 bp, while in the male plant *C. purpureus* R40 ranges from 900 bp to 1000 bp. Thus, the telomere length in male and female plants significantly differs for about two times, which, presumably, may be due to the genetic factors associated with U or V sex chromosomes. In addition, we have shown that only male line of *C. purpureus* R40 had abundant telomeric DNA repeats out of the main telomere length signal on chromosomes, which may be intrachromosomal telomeric DNA. Also the dynamics of telomere length maintaining was studied both for *C. purpureus* GG1 and *C. purpureus* R40 protonema (56 and 42 days respectively). It was shown, that the telomere length did not change in both the female line GG1 and the male line R40, which indicates the stability of the telomeric complex and the system of regulation of the length of *C. purpureus* telomeres during active cell division at the protonema growth stage.

The telomere lengths of *S. fallax* MW and *S. girgensohnii* were about 1880–2000 bp. However, telomeres were slightly shorter in another isolate of *Sphagnum* sp (the Mari-El isolate), whose telomere length ranged from 1100 to 1500 bp. Identification of intra-chromosomal telomeric tandem DNA repeats was carried out by exonuclease restriction using the enzyme Bal31. A strong signal of intrachromosomal telomeric repeats was identified in *S. fallax* MW DNA (~ 2000 bp). However, such interstitial telomeric DNA has not been detected in *Sphagnum* sp. and *S. girgensohnii*. Therefore, it was shown that there are intraspecies variations in telomere length of *Sphagnum* and also differences in the telomeric repeats location on chromosomes.

**Conclusion:** Therefore, we have shown that the telomeric DNA of bryophytes is shorter than the telomeric DNA of angiosperms. The telomere length of bryophytes can vary both between species and within one species. The telomere length in *P. patens* did not vary within the species. The telomere length of *C. purpureus* differs between male and female plants. The telomere length of *C. purpureus* GG1 and R40 lines did not change during the life cycle at the protonema stage. Telomere length in various *Sphagnum* species ranged from 1100 to 2000 bp. However, we found differences in the localization of telomeric DNA repeats in *Sphagnum* species. Further study of the biology of bryophyte telomeres will provide new data on the mechanisms of regulation of telomere length and their functional activity in terrestrial plants.

**Acknowledgements:** This work was part of Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by grant No. 21-14-00147.

## Experimental *in vitro* characterization of TATA-binding protein to TATA-boxes affinity of *Arabidopsis thaliana* real promoter genes

Sharypova E.B., Drachkova I.A., Ponomarenko M.P., Kolchanov N.A., Savinkova L.K.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* sharypova@bionet.nsc.ru

**Key words:** *Arabidopsis thaliana*, TATA box, TBP/TATA affinity

**Motivation and Aim:** There are practically no works on studying the interaction of TBP with TATA boxes of real plant genes. In this work, we have for the first time determined the equilibrium dissociation constants characterizing the affinity of TBP to the TATA boxes of the model plant *Arabidopsis thaliana* genes promoters located in different plant tissues.

**Methods:** DNA sequences of *Arabidopsis t.* promoter genes TATA boxes took from the original works of the authors; used the EMSA method to determine the equilibrium dissociation constants.

Gene	Sequence ODN (5' – 3')	K <sub>D</sub> , nM
<i>TT8</i>	TCTTCTCACATATTAATAACAACCCCT	19 ± 3
<i>β-amylase</i>	TCTACTTCTATATAAAGAGCTCGTGA	1,3 ± 0,3
<i>Pur7</i>	AATTCGTCTCTACACTTCCCCCACC	60 ± 10
<i>PFL</i>	CTCGGTGTCTATATAACTCAACCTC	5 ± 1
<i>TFL1</i>	GGAAAACCCCTATAAATAGATGTCTC	0,9 ± 0,2
<i>Cer2</i>	CTCAATCCTTATATAACACATCCCAT	6 ± 1
<i>AMY1</i>	GGGTCTCAATATAAACACACAAAGCA	23 ± 2
<i>EDD1</i>	AAAATCATATAGTTTAGGTGCGAACCG	24 ± 4
<i>MS5</i>	GTGTTTCATCTATAAACTCAAAGCTC	8 ± 4
<i>AtMLH1</i>	GATGACAAATTATAACCCTAAATAAAA	2,5 ± 0,7

**Results:** As can be seen from the above sequences, they all contain similar TATA elements, more or less homologous to the TATA-box consensus. If we compare the influence of SNP in the TATA box of *Arabidopsis* and human – T1A2T3A4A5A6A7, then the substitutions in positions 2, 3 and 4 (A and T to G or C) are the ones with the greatest influence on transcription and expression, which unites them [1]. Substitutions in positions 1, 6 and 7 have a less significant impact. From the data given in the Table, it can be seen that despite the preservation of the most conservative nucleotides in positions 2, 3 and 4, the affinity of TBP/TATA differs from 3.8 to 26 times. The lowest affinity (60 nM) was determined for the *Pur7* gene, in the TATA box of which the conservative 2T was replaced with C. From the presented data, it can also be seen that the highest affinity of TBP/TATA is observed in the genes of the apical meristem of the roots, stem, shoots and inflorescences of the plant (*PFL*, *TFL1*), and *CER2*, the expression of which was observed only in the epidermis [2].

*Conclusions:* The expanded data obtained can be an important aid for the widely used genome editing technology of food-important agricultural plants to regulate their yield.

*Acknowledgements:* The work was supported by Russian government projects No. FWNR-2022-0016 and No. FWNR-2022-0020 and by the Russian Federal Science & Technology Program for the Development of Genetic Technologies.

### *References*

1. Mukumoto F. et al. DNA sequence requirement of a TATA element-binding protein from Arabidopsis for transcription *in vitro*. *Plant Mol Biol.* 1993;23:995-1003.
2. Haslam T.M. et al. Arabidopsis ECERIFERUM2 is a component of the fatty acid elongation machinery required for fatty acid extension to exceptional lengths. *Plant Physiol.* 2012;160:1164-1174.

## Plant immunity genes of *Solanum tuberosum* L. cultivars associated with resistance to potato late blight

Shmakov N.<sup>1\*</sup>, Kochetov A.<sup>1,2,3</sup>, Afonnikov D.<sup>1,2</sup>, Vasiliev G.<sup>1</sup>, Antonova O.<sup>4</sup>, Shatskaya N.<sup>1</sup>, Glagoleva A.<sup>1</sup>, Ibragimova S.<sup>1</sup>, Khiutti A.<sup>5</sup>, Afanasenko O.<sup>5\*</sup>, Gavrilenko T.<sup>4</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Faculty of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State Agrarian University, Novosibirsk, Russia

<sup>4</sup> N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia

<sup>5</sup> All Russian Institute of Plant Protection, St. Petersburg, Russia

\* [shmakov@bionet.nsc.ru](mailto:shmakov@bionet.nsc.ru); \* [olga.s.afan@gmail.com](mailto:olga.s.afan@gmail.com) (O.S.A.)

**Key words:** Plant immunity, NBS-LRR genes, *Solanum tuberosum*, *Phytophthora infestans*

**Motivation and Aim:** Tuber potato *Solanum tuberosum* L. is one of the most important agricultural crops in the world. Its yield can be affected by various factors, including pathogens and parasites. One of the most dangerous pathogens of *S. tuberosum* is *Phytophthora infestans* (Mont.), an oomycete that causes late blight in potato and tomato. Mechanisms of plant immunity involve proteins of the family of Nucleotide-Binding Site – Leucine Rich Repeat (NBS-LRR, NLR for short). Family of genes encoding NLR proteins is believed to be one of the largest gene families in plants, with more than 700 NLR-coding genes identified in *S. tuberosum* genome [1]. Once NLR protein-coding genes associated with resistance to a specific pathogen are identified, marker-assisted selection can be used to create new crop varieties bearing resistance genes.

**Methods and Algorithms:** Three Russian potato cultivars – Sudarynya, Siverskij and Evraziya, – cultivar Bintje originating in Netherlands, and wild potato relative *Solanum stoloniferum* L. were used in present study. Studied cultivars were examined for their resistance to late blight infection.

Samples of leaves before inoculation and 24 hours after inoculation with *P. infestans* were taken. RNA-seq analysis using Illumina Nextseq 550 platform was performed in order to study transcriptomic changes of *Solanum* samples. Potato transcriptomes were reconstructed *de novo*, and NLR-encoding transcripts were identified in the transcriptomes of potato cultivars. *De novo* reconstructed transcriptomes were aligned to the reference potato genome, and transcripts that did not align were considered novel transcripts, specific for potato cultivars, or otherwise not observed in potato genome previously. Thus, NLR-encoding potato transcripts and novel NLR-encoding potato transcripts were identified. Finally, transcripts expression levels were estimated, and differentially expressed NLR-encoding transcripts were identified.

**Results:** Sudarynya, Siverskij cultivars are shown to be resistant to late blight, while Evraziya and Bintje are susceptible; *S. stoloniferum* is highly resistant to late blight. Five individual *Solanum* transcriptomes contain from 334 to 504 full-length NLR-encoding transcripts. Furthermore, from 103 to 133 novel NLR-encoding transcripts are present in in *S. tuberosum* transcriptomes, and 160 such transcripts are present in *S. stoloniferum* transcriptome. These transcripts are not present in *S. tuberosum* genome.



Sequences of *R3b* and *Rpi-b1b1* genes, known for association with resistance to late blight, are present in transcriptomes of Siversky and Sudarynya cultivars; sequence of *R3b* gene is also present in transcriptome of Evraziya cultivar. Expression of gene *Rpi-b1b1* is induced in presence of *P. infestans*.

*Conclusion:* Novel transcripts observed in *Solanum* genotypes resistant to late blight represent a valuable resource for selection. These sequences can be used to create new potato cultivars highly resistant to *P. infestans* attack, using technologies such as marker-assisted selection. Regulation of NLR genes expression sheds light on mechanisms of plant immunity and reaction to pathogens as well as evolution of plant immunity genes and host-pathogen coevolution.

*Acknowledgements:* This work was supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (agreement No. 075-15-2019-1662).

### *References*

1. Jupe F. et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013;76:530-544. doi: 10.1111/tpj.12307.
2. Kochetov A. V. et al. NLR Genes Related Transcript Sets in Potato Cultivars Bearing Genetic Material of Wild Mexican *Solanum* Species. *Agronomy.* 2021;11:2426. doi: 10.3390/agronomy 11122426.

## Transcriptional profiling of fusarium wilt in flax

Stanin V.<sup>1\*</sup>, Bankin M.<sup>1</sup>, Logachev A.<sup>1</sup>, Rozhmina T.<sup>2</sup>, Samsonova A.<sup>3</sup>, Kanapin A.<sup>3</sup>, Samsonova M.<sup>1</sup>

<sup>1</sup> *Mathematical Biology and Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

<sup>2</sup> *Laboratory of Breeding Technologies, Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

<sup>3</sup> *Centre for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

\* *staninvladislav@mail.ru*

**Key words:** Fusarium wilt, flax, transcriptional profiling, cytochrome, host-parasite interaction

*Motivation and Aim:* Fusarium wilt of flax is an aggressive disease caused by soil-borne fungal pathogen *Fusarium oxysporum* f. sp. *lini* posing a major threat to flax production worldwide [1–3]. Recently we reported chromosome-level assemblies of 5 highly pathogenic *F. oxysporum* f. sp. *lini* strains, including a highly virulent MI39 isolate, thus providing a background for detailed genomic studies of host-pathogen interaction mechanisms on multiple levels [4, 5]. Here, we present a first insight into molecular mechanisms involved in response to infection. Such data may be of use when integrated with the infection resistance markers emerging from GWAS studies, thus opening up new possibilities for disease control and prevention.

*Methods and Algorithms:* In this study we examined susceptible and resistant flax varieties, namely, LM98 and Atalante, infected by MI39 *Fusarium* isolate. Uninfected plants and pure culture of the fungus served as controls. Flax samples were harvested on the third and the fifth day post inoculation (dpi). A total RNA mixture (i.e., flax and fungus) was extracted from infected plant roots and sequenced. The RNASeq data was processed in parallel with kallisto/sleuth suite [6, 7] using flax and *Fusarium* gene models. The functional annotation and GO enrichment analyses of resulting sets of differentially expressed genes was done with XGR package [8].

*Results:* *Fusarium* infection produces a diverse spectrum of transcriptional states both of the host and the pathogen. The resistant plants exhibited a sharp, prolonged response to the infection as early as the third dpi. In line with previous work [9–11], we observed changes in expression of genes controlling membrane remodelling, secondary metabolite synthesis, synthesis of lipids and cytochrome-related proteins, transcription factors in response to infection. Furthermore, an activation of Germin-like genes provides a clear indication of flax response to the biotic stress associated with disease progression. In contrast, the susceptible plants provided no sign of significant transcriptional response on the third dpi. However, the delayed changes in transcriptional profiles were visible on the fifth dpi, manifesting in expression changes of genes controlling protein kinases, transmembrane transport, cytochrome-related genes and glycoproteins associated with plant resistance to diseases.

Upon invasion the fungus exhibits upregulation of genes associated with pathogen penetration through the host's cell walls and further progression (e.g., proteases and cellular energetics) irrespective of the host's resistance status.

**Conclusion:** Previous research [12] has highlighted the significance of cytochrome-related genes in plant defense. Also, earlier we reported QTN in a gene encoding flax' cytochrome P450 protein associated with resistance to *Fusarium* infection [13]. Our findings further emphasize the indispensable role the cytochrome superfamily in plant pathogen response. Furthermore, our results also provide first insights on the whole genome level into transcriptional activity of the plant pathogen upon penetration into host. The data generated in the framework of this project lay a solid foundation for in-depth delineation of the regulatory mechanisms of plant-pathogen interaction.

**Acknowledgements:** This work is supported by Peter the Great St. Petersburg Polytechnic University in the framework of Russian Federation's Priority Strategic Academic Leadership Programme (Agreement 75-15-1038-2021-1333).

## References

1. Nair P., Kommedahl T. The establishment and growth of *Fusarium lini* in flax tissues. *Phytopathology*. 1957;47(1):25.
2. Nelson P.E., Toussoun T.A., Cook R.J. (Eds.). *Fusarium: diseases, biology, and taxonomy*. Pennsylvania State Univ. 1981.
3. Michielse C.B., Rep M. Pathogen profile update: *Fusarium oxysporum*. *Mol Plant Pathol*. 2009;10(3):311.
4. Kanapin A. et al. The genome sequence of five highly pathogenic isolates of *Fusarium oxysporum* f.sp. *lini*. *Mol Plant-Microbe Interact*. 2020;33(9):1112-1115.
5. Samsonova A. et al. A genomic blueprint of flax fungal parasite *Fusarium oxysporum* f. sp. *lini*. *Int J Mol Sci*. 2021;22(5):2665.
6. Bray N.L. et al. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525-527.
7. Pimentel H. et al. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods*. 2017;14(7):687-690.
8. Fang H. et al. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*. 2016;8(1):129.
9. Galindo-González L., Deyholos M.K. RNA-seq transcriptome response of flax (*Linum usitatissimum*) to the pathogenic fungus *Fusarium oxysporum* f. sp. *lini*. *Front Plant Sci*. 2016;7:1766.
10. Dmitriev A.A. et al. Differential gene expression in response to *Fusarium oxysporum* infection in resistant and susceptible genotypes of flax (*Linum usitatissimum* L.). *BMC Plant Biol*. 2017;17(2):253.
11. Boba A. et al. Transcriptomic profiling of susceptible and resistant flax seedlings after *Fusarium oxysporum lini* infection. *PLoS One*. 2021;16(1):e0246052.
12. Xu J. et al. The cytochrome P450 superfamily: Key players in plant development and defense. *J Integr Agric*. 2015;14(9):1673-1686.
13. Kanapin A. et al. Genomic regions associated with fusarium wilt resistance in flax. *Int J Mol Sci*. 2021;22(22):12383.

## The molecular response to low-temperature stress in peach cultivars with different cold tolerance

Tsyupka V.<sup>1,2\*</sup>, Smykov A.<sup>1,2</sup>, Grebennikova O.<sup>1,2</sup>, Chelebieva E.<sup>1</sup>, Ibadullaeva E.<sup>1</sup>, Bulavin I.<sup>1,2</sup>, Vodiasova E.<sup>1</sup>

<sup>1</sup> Kurchatov Genomic Centre of the Nikita Botanical Garden – National Scientific Center of RAS, Yalta, Russia

<sup>2</sup> Nikita Botanical Garden – National Scientific Center of RAS, Yalta, Russia

\* valentina.brailko@yandex.ru

**Key words:** antioxidant enzyme, gene expression, cold stress, peach cultivars

*Motivation and Aim:* The *Prunus persica* (L.) Batsch (peach) is a vital fruit crop that tends to be injured during early spring frost. Such stress could depress growth, development and distribution. The peach varieties have been developed using breeding methods, which is better adapt to the low temperature. At the same time, there are very few studies on the molecular mechanisms that make varieties tolerant to low temperatures. Most of these studies focus on one or two varieties of peach only, which is insufficient to understand the general patterns. Cold stress has a negative effect on many cellular processes, the most common one is the cell membrane damage and cell destruction due to reactive oxygen species (ROS) generation. It can be assumed that varieties with cold tolerance should have effective antioxidant protection. Thus, the aim of this work was to study cellular damage to lower temperatures by microscopy and to analyze the activity and expression of antioxidants in cold-resistant and cold-sensitive varieties.

*Methods and Algorithms:* Six cultivars were used for this experiment: three cold-tolerant (Krymskiy feyerverk, Podarok Like, Temisovskiy) and three cold-sensitive (Loadell, Springtime, Springold). The 9 healthy annual branches of each cultivar were collected from tree on 2<sup>nd</sup> of the March. Then they were divided into three group: control and two experimental groups. The experimental groups were placed in the growth chamber with temperature  $-6^{\circ}\text{C}$  for 24 h (first step) and then the temperature was raised to the control for 24 h (second step). Three stigmas were sampled from each branch and fixed in RNALater for genetic analyses and frozen at  $-80^{\circ}\text{C}$  for biochemistry. Catalase and peroxidase were selected as the key antioxidant enzymes.

*Results:* The cell damages were detected during cold stress by microscopy. Also, as the activity changes of antioxidant enzymes as gene expression changes of their encoding genes were observed.

*Conclusion:* It is assumed that varieties that are not frost-resistant will not have activated antioxidant defense mechanisms, which in turn can lead to an increase in ROS and cell damage. This, in turn, leads to the death of flower buds and inhibited growth and development.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Centre of the Nikita Botanical Garden – National Scientific Center of RAS (075-15-2019-1670).

## Protein prenyltransferases regulate the multicellular thallus development of liverwort *Marchantia polymorpha*

Valeeva L.R.<sup>1\*</sup>, Ibragimova S.<sup>1</sup>, Kulazhenko M.S.<sup>1</sup>, Yoshitake Y.<sup>2</sup>, Kohchi T.<sup>2</sup>, Sharipova M.R.<sup>1</sup>

<sup>1</sup> Institute of Fundamental Medicine and Biology, Kazan (Volga region) Federal University, Kazan, Russia

<sup>2</sup> Graduate School of Biostudies, Kyoto University, Kyoto, Japan

\* lia2107@yandex.ru

**Motivation and Aim:** Protein prenylation plays a crucial role in the development of all multicellular organisms. Multicellularity transition is one of the major events in the evolution of living organisms. Considering that in Plant kingdom the multicellularity transition occurred several times independently the study of protein prenylation role in the development of plants allow us to find out new data on the multicellularity regulation. In this study, the role of three different prenyltransferases in the development of liverwort *Marchantia polymorpha* was studied.

**Methods and Algorithms:** Knockout lines of *M. polymorpha* protein prenyltransferases' genes were obtained via CRISPR/Cas9 technology. Complement lines were developed using pMpGW301-305 plasmids. Expression pattern were detected by CLSM. Proteome analysis were performed by MS.

**Results:** Mutant lines of *Δplp* and *Δggb* demonstrated callus-like phenotype with undifferentiated cells, formed solid structures, whereas *Δeral* knockout plants could form wild type dorso-ventral thallus phenotype. However, that *ΔRabggb* mutant plants were viable and remained the WT phenotype. The complementation of KO *M. polymorpha* resulted in the recovering of WT phenotype. The proteome analysis of *Δplp* and *Δggb* mutant plants allowed us to find out the correlation in the protein expression in these two lines comparing to WT.

**Conclusion:** Therefore, we showed the influence of protein prenylation on the development of the multicellular thallus of *M. polymorpha*. The following study of the expression of found genes will allow us to determine new gene pathways involved in multicellularity regulation.

**Acknowledgements:** The work was supported by the Russian Federation Presidential Scholarship No. CII-3391.2021.4.

## Towards identification of genetic factors underlying curly birch wood patterning using high-throughput genotyping

Zhigunov A.V.<sup>1\*</sup>, Vetchinnikova L.V.<sup>1,2</sup>, Volkov V.A.<sup>1</sup>, Potokina E.K.<sup>1</sup>

<sup>1</sup> St. Petersburg Forest Technical University, St. Petersburg, Russia

<sup>2</sup> Karelian Research Center RAS, Petrozavodsk, Republic of Karelia, Russia

\* a.zhigunov@bk.ru

**Key words:** curly birch, sibling populations, GBS, high-throughput genotyping

We present the project aimed at identifying genetic factors responsible for the manifestation of the wood "patterning" trait in curly birch *Betula pendula* Roth var. *carelica* (Mercklin) Hämet-Ahti. Earlier it was found that the wood "patterning" trait is inherited in the seed progeny of curly birch and preserved during vegetative propagation both in natural populations and in man-made plantations. Experimental evidence for the monogenic nature of inheritance of this trait has recently been obtained. In populations of complete siblings obtained from controlled crosses of contrasting parental genotypes (curly and normal), a classical Mendel segregation was obtained, indicating that the "patterned" allele of the as yet unidentified gene is dominant and in the heterozygous state causes the phenotype of "patterned" wood of the curly birch. In the homozygous state, this allele is fatal. Hence, it is become possible to identify the "patterned" wood gene in curly birch or to identify polymorphic loci closely linked to it.

The project is focused on identifying molecular markers linked with the "patterned" wood gene in curly birch through high-throughput genotyping of trees from sibling's or half-sibling's progeny obtained from different crosses of curly birch, where segregating by the studied phenotype is observed. The latter becomes possible due to availability of high-throughput "genotyping by sequencing" technology (GBS, Genotyping By Sequencing). Based on the results of identification of a large number of polymorphic loci, a search for associations between polymorphic molecular markers (SNPs) and variability of the wood "patterning" trait will be performed. The accessibility of a published "reference" genome for *Betula pendula* will make it possible to use the obtained information of high-throughput genotyping and association analysis to search for the candidate gene responsible for the formation of the studied phenotype.

Identification of the genetic determinants of this economically valuable trait will allow to develop molecular markers of the patterned wood texture trait of curly birch, which can be used at the early stages of plant development. Morphologically, indirect signs of wood "patterning" in curly birch can appear, on average, at the age of 8 years.

*Acknowledgements:* This work was supported by the Russian Science Foundation (Project No. 22-16-00096).

## 6.2 Section “Developmental biology of plants: computational and experimental approaches”



## Condensation of STM empowers meristem activity and enhances salt tolerance

Cao X.<sup>1\*</sup>, Wang Y.<sup>2</sup>, Jiao Y.<sup>1,2,3</sup>

<sup>1</sup> Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

<sup>2</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> School of Life Sciences, Peking University, Beijing, China

\* caoxiuwei12@genetics.ac.cn

**Key words:** condensate, STM, meristem activity

In plants, shoot meristem generates the entire shoot system, and is precisely maintained throughout the life cycle under various environmental challenges. Here, we identified a prion-like domain (PrD) in the key shoot meristem regulator SHOOT MERISTEMLESS (STM), that distinguish STM from other related KNOX1 proteins. We demonstrate that PrD promotes STM to form nuclear condensates, which is required for maintaining the shoot meristem. STM nuclear condensation formation is stabilized by selected PrD containing STM-interacting BELL proteins *in vivo* and *in vitro*. Moreover, condensation promotes interaction between STM and Mediator to enhance transcription. Thus, condensation formation emerges as a novel mechanism of regulating shoot meristem function. Furthermore, we find that the formation of STM condensates is enhanced upon salt stress, which allows enhanced salt tolerance and increased shoot branching. Our findings highlight the importance of partitioning of transcription factors in cell fate determination, and as a tunable environmental sensing mechanism.



## The real function of the gene *TERMINAL FLOWER 1*

Kharchenko V.E.

*SEI HE LPR Lugansk State Agrarian University, Lugansk, Lugansk People's Republic*  
*viktoriaharchenko@rambler.ru*

**Key words:** TERMINAL FLOWER, inflorescence, heterochrony, pedomorphosis, anaboly, morphogenesis, harvest

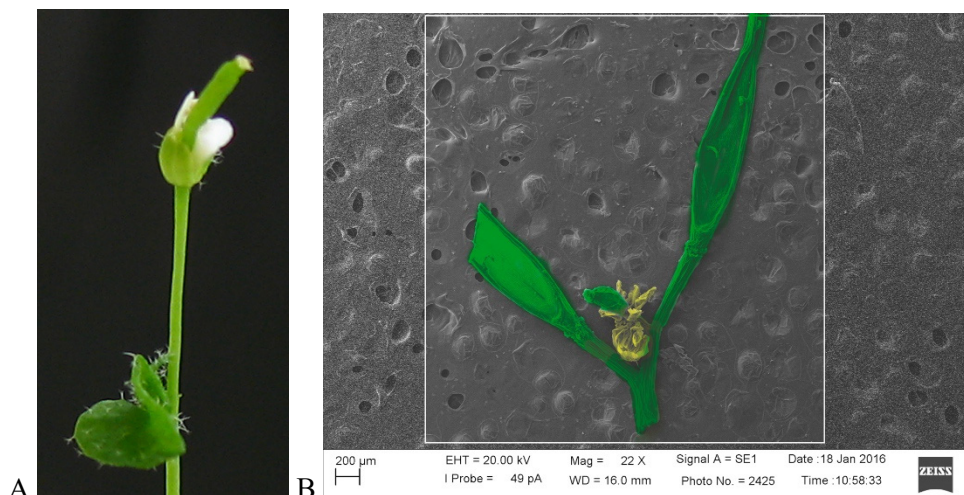
*Motivation and Aim:* *TERMINAL FLOWER 1 (TFL1)* is a key regulator of *Arabidopsis thaliana* plant architecture that is of paramount importance to plant adaptation and crop improvement. Published research denotes that *TFL1* suppresses transition to flowering [1]. A protein homologous of *TFL1* were found in virtually all seeds plants [2]. Despite the importance for plant breeding, limited research data on the *TFL1* and its regulation of flowering is available. The primary objective of this study to clarify the functions of the gene *TFL1*.

*Methods and Algorithms:* The functions of the *TFL1* gene were tested using a detailed analysis of *Arabidopsis thaliana* phenotype variability with its normal (line *Ler*) and abnormal expression (the mutant line *tfl1-2*). Obtained results were analyzed using ANOVA from the STATISTICA package. The shoots' structure was analyzed using the fractal approach. Certain elements were isolated on the shoots and duplicated in accordance with the principle of decreasing proportion. Modification variability of inflorescences was evaluated by different the level of illumination (from 6000 to 800 lux). The structure of inflorescences was analyzed during morphogenesis using an electron scanning microscope and a laser confocal microscope.

*Results:* Plants with a single terminal flower were obtained as a result of modification variability in the parent line *Ler Arabidopsis thaliana* (Fig. 1, A). Indeed, the terminal flower had a structure typical to *Arabidopsis thaliana* [2]. Microscopic analysis of the inflorescences revealed that the complex of fused flowers at the top of *Arabidopsis thaliana tfl1-2* shoots is not identical to the "terminal flower", since the last flower in its inflorescence can be in a lateral position and is usually underdeveloped (see Fig. 1, B). Therefore, mutations in the *TFL1* gene will not switch the inflorescence development program from indeterminate to determinate but only reduce it. The obtained results are consistent with the effects described in experiments on ectopic expression of *TFL1*-like genes [2, 4, 5]. With insufficient expression of the *TFL1* gene in plants, the structure of the inflorescence turned out to be underdeveloped. When *TFL1* is overexpressed, new stages of inflorescence morphogenesis arise. However, they do not form apical flowers, and their flower symmetry does not change. A new stage of morphogenesis arises in the inflorescence when the *TFL1* gene is overexpressed, similarly to the way it is observed in polyploids. Therefore, *TFL1* regulates the process of morphogenesis in general and is of key importance for harvest plants.

*Conclusion:* The results obtained in the model suggest that *TERMINAL FLOWER 1* regulates not only the time of transition to flowering in plants, but also the stages of inflorescence morphogenesis and harvest of plants. The insufficient expression leads to reduced or lost stages of morphogenesis (pedomorphosis). On the contrary, overexpression causes new stages of morphogenesis (anaboly, hypermorphosis) to

appear. It is probable, that the misunderstanding of the function of the *TFL1* is associated with an attempt to express it through a specific phenotypic effect, and not through the modus of action.



**Fig. 1.** Terminal flowers in *Arabidopsis thaliana*

This terminal flower arose in *Ler Arabidopsis thaliana* as a result of modification variability (A). This terminal flower arose in *tfl1-2 Arabidopsis thaliana* as a result of mutation variability (B).

#### References

1. Wickland D.P., Hanzawa Y. The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. *Mol Plant*. 2015;8(7):983-997. doi: 10.1016/j.molp.2015.01.007.
2. Danilevskaya O.N., Meng X., Ananiev E.V. Concerted Modification of Flowering Time and Inflorescence Architecture by Ectopic Expression of TFL1-Like Genes in Maize. *Plant Physiology* 2022;153:238-251.
3. Kharchenko V.E. Terminal flower and development of inflorescence structure. Novosibirsk: SibAK LLC, 2021.
4. Nakagawa M., Shimamoto K., Kyojuka J. Overexpression of RCN1, RCN2, rice TERMINAL FLOWER 1/CENTRORADIALIS homologs, confers delay of phase transition, altered panicle morphology in rice. *Plant J*. 2002;29(6):743-750.
5. Mohamed R. et al. *Populus CEN/TFL1* regulates first onset of flowering, axillary meristem identity, dormancy release in *Populus*. *Plant J*. 2010;62(4):674-688.

## Does *DEEPER ROOTING 1 (DRO1)* gene of cucumber (*Cucumis sativus* L.) regulate a lateral root angle?

Kiryushkin A.S.\*, Ilina E.L., Guseva E.D., Demchenko K.N.

Laboratory of Cellular and Molecular Mechanisms of Plant Development, Komarov Botanical Institute, RAS, St. Petersburg, Russia

\* akiryushkin@binran.ru

**Key words:** CRISPR/Cas, root system architecture, genome editing, *DRO1*, lateral root, cucumber

**Introduction:** The orientation or gravitropic set point angle (GSA) of underground plant organs are regulated by different gene families. The IGT (*LAZY1/TAC1/DRO1*) gene family is one of them. The representative of the IGT family, the *DRO1* gene, was initially described as a regulator of rice rooting depth. *DRO1* homologues were subsequently identified in other plant species, both monocots and dicots. The aim of this study was to elucidate whether the *DRO1* gene of cucumber (*Cucumis sativus* L., Cucurbitaceae) regulates a change in lateral root angle.

**Methods:** The results were obtained using phylogenetic analysis of cucumber IGT protein family, quantitative real time PCR (qPCR), confocal laser scanning microscopy (CLSM), and CRISPR/Cas9-mediated editing of cucumber *DRO1* genes.

**Results:** Three putative orthologues of rice and Arabidopsis *DRO1* proteins encoding by *CsDRO1a*, *CsDRO1b*, and *CsDRO1c* genes were identified in cucumber. The phytohormone auxin plays a central role in formation of root GSA, so using qPCR the effect of exogenous auxin (1-Naphtaleneacetic acid) on *CsDRO1a*, *CsDRO1b*, and *CsDRO1c* expression was examined. The expression of *CsDRO1a* gene was significantly increased, while transcript levels of *CsDRO1b* and *CsDRO1c* were not changed. Study of the expression of p*CsDRO1a*, p*CsDRO1b*, and p*CsDRO1c* promoter-reporter fusions in the cucumber root tips was conducted using Agrobacterium-mediated hairy root transformation following by analysis of root sections using CLSM. Expression of *CsDRO1a*, *CsDRO1b*, and *CsDRO1c* was observed in central cylinder, pericycle, endodermis, and cortex up to an average distance of 700  $\mu\text{m}$  from the initial cells. Then, expression maxima of *CsDRO1a* was retained only in endodermis cells involved in lateral root development, whereas *CsDRO1b* and *CsDRO1c* continued to express in cells layers of central cylinder, in pericycle, and in endodermis. The role of the single *CsDRO1* gene (*CsDRO1a*, *CsDRO1b* or *CsDRO1c*) in regulation of lateral root angle were clarified using CRISPR/Cas9-mediated knockout. The efficiency of *CsDRO1a* editing was 94 %, for *CsDRO1b* – 75 %, and for *CsDRO1c* – 50 %. Although editing events resulted in complete knockout of the single *CsDRO1* gene in individual transgenic roots, no statistically significant changes in lateral root angle were detected.

**Conclusions:** The results suggest that *CsDRO1a*, *CsDRO1b*, and *CsDRO1c* may have degenerate function to regulate lateral root angle in cucumber parental root.

**Acknowledgements:** The Ministry of Science and Higher Education of the Russian Federation supported this study in the framework of Bio-resource collections (Grant No. 075-15-2021-1056).

## Salicylic acid and abiotic stress: nature of interaction

Kovrizhnykh V.V.\* , Bagautdinova Z.Z.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\**vasilinakovr@gmail.com*

**Key words:** salicylic acid, root growth, abiotic stresses

*Motivation and Aim:* A plant hormone salicylic acid (SA) has attracted attention of plant physiologist as a modulator of plant immunity under biotic stress conditions [1]. However, over the past decades, a large amount of data has accumulated describing the SA role in plant growth under various abiotic stresses [2]. Despite SA application in agriculture as plant growth regulator has great potential, the requirements, which would provide an optimal growth-defense balance, are poorly understood. The aim of the current study is to compile a general map of characteristic changes in root morphology after SA treatment under stress conditions such as chilling, salinity, drought and heavy metals, and retrieve the conditions, which facilitate protective effects of SA.

*Methods and Algorithms:* In the study, we analyzed more than 30 publications devoted to the simultaneous stress and SA effects on the root system architecture in 17 plant species. Systematization of information in tables and separation of data by process (seed germination, root length, root biomass), abiotic stress types, plant species and SA treatment conditions were used for comparative data analysis.

*Results:* SA treatment alleviates the abiotic stress effects on the root system. SA can restore root length and biomass suppressed by salt, drought, cold, nickel, cadmium, arsenium, zinc, lead, chromium and aluminium stresses [3]. The effective SA concentration depends on the plant species and the stress type. For example, for some species, low SA concentrations promote seed germination under salinity, while higher SA concentrations lead to an increase of stress effects.

The SA protective effect may rely on a decrease in the toxicity of a stress factor but mainly it is related to SA stimulation of plant growth and development, which are inhibited by stress conditions. This recovery is especially promoted by SA protection of cell divisions. The restoration of root growth under stresses needs treatment with those SA concentrations, which under normal conditions enhance growth or at least do not inhibit it [4–6]. Such treatment provides the optimal level of endogenous SA through the regulation of SA metabolism [3]. Further SA mitigation of abiotic stress effects occurs in crosstalk with other plant hormones (in particular, IAA, ethylene, ABA).

*Conclusion:* SA treatment completely eliminates the effect of weak and moderate stress and partially restores root growth under severe ones. To ensure optimal plant growth under abiotic stress plants have developed various mechanisms to control balance between growth and defense, in which SA plays an important role.

*Acknowledgements:* This work was supported by the Grant of the President of the Russian Federation MK-3470.2021.1.4.

### References

1. Zhang Y. et al. Salicylic acid: Biosynthesis, perception, and contributions to plant immunity. *Curr Opin Plant Biol.* 2019;50:29-36.

2. Rivas-San Vicente M. et al. Salicylic acid beyond defence: Its role in plant growth and development. *J Exp Bot.* 2011;62:3321-3338.
3. Bagautdinova Z. et al. Salicylic Acid in Root Growth and Development. *Int J Mol Sci.* 2022;23(4):2228.
4. Zadehbagheri M. Salicylic acid priming in corn (*Zea mays* L. var. Sc. 704) Reinforces NaCl tolerance at germination and the seedling growth stage. *Int J Biosci.* 2014;4:187-197.
5. Alyemeni M. et al. Effect of salicylic acid on the growth, photosynthetic efficiency and enzyme activities of leguminous plant under cadmium stress. *Not Bot Horti Agrobot Cluj-Napoca.* 2014;42(2):440-445.
6. Espanany A., Fallah S. Seed germination of dill (*Anethum graveolens* L.) in response to salicylic acid and halopriming under cadmium stress. *Iran J Plant Physiol.* 2016;6(3):1701-1713.

## Methods to investigate the biomechanics of the plant cell wall

Kozlova L.\*, Petrova A., Gorshkova T.

*Kazan Institute of Biochemistry and Biophysics, FRC KazSC RAS, Kazan, Russia*

\* *kozlova@kibb.knc.ru*

**Key words:** plant, cell wall, polysaccharides, atomic force microscopy, development

The plant cell wall is the envelope of every plant cell. It consists mainly of polysaccharides. The cell wall resists the internal hydrostatic pressure of the cell, also known as turgor pressure. Turgor and the cell wall cooperate to create the hydrostatic skeleton of the plant. Young plants and young parts of the plant, which do not yet have rigid mechanical tissues, rely only on this hydrostatic skeleton to maintain their architecture. This is the reason why wilting in the shortage of water is so noticeable in young plants.

On the one hand, the cell walls must be strong enough to resist the turgor pressure. On the other hand, it must be extensible to allow the growth of the plant cell. It should be noted that the enlargement of plant cells is much more prominent than that of animal cells. A plant cell can increase its size to 1000 times its initial size. During the whole process, no cracks or tears should appear in the cell wall and its thickness should be maintained.

Another aspect that is directly regulated by the mechanical properties of the cell wall is the shape of the cell. In order to fulfill their functions, plant cells sometimes acquire rather sophisticated shapes. However, even relatively simple shapes require the cell wall to have different properties on different sites. Turgor pressure acts uniformly in all directions, so the acquisition of any shape other than spherical requires local heterogeneity of cell wall properties.

Thus, the plant cell wall is a unique material that combines strength, extensibility and heterogeneity. The study of the plant cell wall in terms of mechanics is important both for developmental biology and for the design of smart materials as a source of inspiration.

This talk will focus on the methods used to study the biomechanics of plant cell walls, with an emphasis on atomic force microscopy as the primary and most widely used technique.

*Acknowledgements:* The work was partially supported by the Russian Science Foundation (18-14-00168).

## CisCross web service: a gene set enrichment analysis to predict the upstream regulators for *Arabidopsis thaliana*

Lavrekha V.V.<sup>1,2\*</sup>, Tsukanov A.V.<sup>1</sup>, Bogomolov A.G.<sup>1</sup>, Omelyanchuk N.<sup>1</sup>, Zemlyanskaya E.V.<sup>1,2</sup>, Levitsky V.G.<sup>1,2</sup>, Mironova V.V.<sup>1,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Faculty of Science, Radboud University Nijmegen, Nijmegen, Netherlands*

\* [vvl@bionet.nsc.ru](mailto:vvl@bionet.nsc.ru)

**Key words:** DAP-seq, gene set enrichment, transcription factor binding sites, *Arabidopsis thaliana*

**Motivation and Aim:** Modern transcriptomic technology RNA-seq provides lists of differentially expressed genes (DEG), i.e. genes for which expression levels between two experimental conditions of cell types, tissues or specific treatments are statistically significant [1]. Although consequent standard analyses uncover gene networks trying to explain observed biological responses, molecular mechanisms of DEG transcription regulation are still far from being revealed. Transcription factors (TFs) are proteins that function as major players involved in this regulation. They bind genomic target sequences known as transcription factor binding sites (TFBS) to initiate gene activation or repression. Recently developed DAP-Seq (DNA Affinity Purification and sequencing) high-throughput method detects the DNA fragments from genomic DNA libraries that are bound by an *in vitro* expressed TF [2]. O'Malley and colleagues generated DAP-Seq profiles for 387 *Arabidopsis thaliana* TFs. The large collection of similarly processed in [2] DAP-Seq sets of peaks, The Plant Cistrome database, has been used for systematic identification of TF upstream regulators [3, 4]. Two web-applications for gene set enrichment analysis in *Arabidopsis*, TF DEACoN [4] and EAT-UpTF [5], have been developed already, which use Plant Cistrome list of TFs's targets. However, the Plant Cistrome database consists of processed by GEM tool sets of peak [2, 6]. As shown in a benchmarking study MACS2 peak calling tool may give better quality of peak sets [7, 8]. That's why we re-annotated raw DAP-Seq data [2] by the MACS2 tool so it fits better for the task of the gene set enrichment analysis. The algorithm of TF enrichment analysis, CisCross algorithm, which we used previously in [3] and new collections of TF's sets of peaks we perform on the new CisCross web service.

**Methods and Algorithms:** All 931 raw datasets [2] were processed and aligned with snakePipes (v. 2.5.6) [9]. Peak calling was done by MACS2 [8] or GEM [7]. For all samples, we used control SRR2926068 or SRR2926069 depending on the data type. We filtered out the peak sets with a FRIP value of less than 0.01. If two DAP-seq replicas were obtained for a TF, then we joined them using the IDR (irreproducible discovery rate) tool [10]. Then we select the final variant of a set of peaks and removed sets of peaks, which contained less than 200 peaks. As a result, we've got 649/650 peak sets for 426/428 TFs for CisCross MACS2/GEM collections.

**CisCross algorithm.** 5'-regulatory regions of the requested list of genes are used as a foreground and those for the rest *Arabidopsis* genes as a background. For each TF peak set, CisCross counts the number of genes in foreground/background which 5'-regulatory

regions possess or do not possess the TF binding peak. The gene promoter enrichment for TF binding peaks is assessed using Fisher's exact test. CisCross runs through all the peak sets in the selected DAP-Seq collection and applies the correction for multiple comparisons.

**Results:** The CisCross web service (<https://plamorph.sysbio.ru/ciscross/>) includes three modes: (1) CisCross-Main, (2) CisCross-Light, and (3) TF Targets. CisCross-Light gives the list of DAP-Seq binding peaks detected in the 5'-regulatory region of the requested gene. CisCross-Main implements the algorithm written above with the following options. For the background one can choose between Araport11 and TAIR10 genome versions; 500, 1000, or 1500 bp-long 5'-regulatory regions. For the DAP-Seq peak set collections: (1) Plant Cistrome [2]; (2) CisCross GEM; (3) CisCross MACS2. For the Multiple testing procedure: Benjamini-Hochberg or Bonferroni methods. TF Target mode is aimed at finding all targets of input TF among the genes submitted for input.

**Conclusion:** As CisCross collections contains the peak sets for a greater than Plant Cistrome amount of TFs, it resulted in detection of novel regulators for gene sets. Using different backgrounds and length of upstream region gives the researcher more opportunities to investigate regulation of genes by CisCross-Main. In addition, CisCross web service modes help to explore the output of the CisCross-Main in more detail.

**Acknowledgements:** The work was supported by Russian Science Foundation (21-14-00240).

## References

1. Oshlack A., Robinson M.D., Young M.D. From RNA-seq reads to differential expression results. *Genome Biol.* 2010;11(12):220. doi: 10.1186/gb-2010-11-12-220.
2. O'Malley R.C., Huang S.S.C., Song L., Lewsey M.G., Bartlett A., Nery J.R., ..., Ecker J.R. Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell.* 2016;165(5):1280-1292.
3. Shi D., Jouannet V., Agustí J., Kaul V., Levitsky V., Sanchez P., ..., Greb T. Tissue-specific transcriptome profiling of the Arabidopsis inflorescence stem reveals local cellular signatures. *Plant Cell.* 2021;33(2):200-223.
4. Harkey A.F., Sims K.N., Muday G.K. A new tool for discovering transcriptional regulators of co-expressed genes predicts gene regulatory networks that mediate ethylene-controlled root development. *In silico Plants.* 2020;2(1):diaa006.
5. Shim S., Seo P.J. EAT-UpTF: Enrichment Analysis Tool for Upstream Transcription Factors of a group of plant genes. *Front. Genet.* 2020;11:1058.
6. Guo Y., Mahony S., Gifford D.K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* 2012;8(8):e1002638. doi: 10.1371/journal.pcbi.1002638.
7. Thomas R., Thomas S., Holloway A.K., Pollard K.S. Features that define the best ChIP-seq peak calling algorithms. *Brief Bioinform.* 2017;18(3):441-450. doi: 10.1093/bib/bbw035.
8. Zhang Y., Liu T., Meyer C.A., Eeckhoutte J., Johnson D.S., Bernstein B.E., Nusbaum C., Myers R.M., Brown M., Li W., Liu X.S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):r137. doi: 10.1186/gb-2008-9-9-r137.
9. Bhardwaj V., Heyne S., Sikora K., Rabbani L., Rauer M., Kilpert F., ..., Manke T. snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics.* 2019;35(22):4757-4759.
10. Li Q., Brown J.B., Huang H., Bickel P.J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 2011;5(3):1752-1779. doi: 10.1214/11-AOAS466.



## Processing of serial microscopic images for 3D reconstruction of plant tissues

Mursalimov S.<sup>1\*</sup>, Ohno N.<sup>2,3</sup>, Deineko E.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Jichi Medical University, Shimotsuke, Japan*

<sup>3</sup> *National Institute for Physiological Sciences, Okazaki, Japan*

\* *mursalimovsr@gmail.com*

Volume microscopy is a new rapidly developing technique used in cell and developmental biology. Both light and electron volume microscopes generate hundreds and thousands of serial microscopic images. Regions of interest (ROIs) containing particular cells, organelles or structures should be analyzed on every single image. Thus, there is a need to develop protocols for the fast and precise ROI detection and visualization. We used both light and electron volume microscopies to analyze male meiocytes inside plant anthers. ImageJ, Microscopy Image Browser, Dragonfly, UNI-EM and Amira software were used for the image alignment, segmentation and 3D modeling. Manual, semi-automatic and automatic types of image segmentation were used. Unusual behavior of nuclei in the developing male meiocytes of tobacco and rye was detected on 3D models. It was shown that at certain meiotic stages intercellular nuclear migration (cytomixis) occurs in most normal tobacco and rye male meiocytes. The obtained results suggest that there are still many unexplored features of meiosis hidden by limitations of common types of microscopy and that volume microscopy can turn over a new leaf in meiosis research. Development of deep learning image analysis tools is crucial for this research.

*Acknowledgements:* The reported study was funded by the Russian Foundation for Basic Research and the Japan Society for Promotion of Science according to the research project No. 21-54-50001 and JPJSBP120214813.

# Salicylic acid treatment changes plant root system patterning

Omelyanchuk N.A.<sup>1\*</sup>, Bagautdinova Z.Z.<sup>1</sup>, Zemlyanskaya E.V.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *nadya@bionet.nsc.ru*

**Key words:** salicylic acid, root growth, lateral roots, adventitious roots, auxin

*Motivation and Aim:* Plant hormone salicylic acid (SA) mediates responses to pathogen and pest attacks and abiotic stresses, including drought, salinity, chilling, heavy metals and others [1]. Besides, to date, numerous data have been accumulated that favor a versatile role of SA in plant development. The SA content, the ratio of free and conjugated forms and their dynamics differ in shoots and roots, and thus can potentially cause differences in SA functions there. However, unlike in shoots, morphogenetic role of SA in roots has been poorly generalized so far [2]. Here we infer the patterns of root system response to SA across different species.

*Methods and Algorithms:* We performed a systematic analysis of more than 100 papers on SA treatments in about 40 plant species, with SA concentration ranging from 10 fM to 10 mM to summarize data on root system response to this hormone. A simple root structure facilitates description of treatment effects.

*Results:* SA regulates root morphology from radicle emergence during seed germination to middle cortex formation in aged plants. SA influences most of these processes in a concentration-dependent manner with opposite effects manifested at different concentrations [3]. Wherein, concentration ranges, which cause a certain effect, are genotype-specific. SA concentration dependence was found for following processes: radicle emergence (in cucumber, wheat, maize, carrot, and Arabidopsis); root elongation (in cucumber, wheat, beans, lens, *Trigonella foenum-graceum* L., and *Pennisetum glaucum* L.) and adventitious rooting (in mungbean, rhododendron, and Arabidopsis). For many species, SA concentrations that only activate or only inhibit the aforementioned processes have been found yet. The same unidirectional patterns are observed for SA impact on lateral root development in Arabidopsis and *Catharanthus roseus* hairy root tissue culture obtained from *Agrobacterium rhizogenes* infected leaves [4-6]. In Arabidopsis, SA (3-250  $\mu$ M) inhibits lateral rooting whereas in *Catharanthus roseus* hairy root tissue culture very low SA concentrations (10 fM) stimulate this process, pointing out that any case of the unidirectional SA inhibition of root growth needs to be tested with nano and femto mole SA concentrations. In its turn, unidirectional SA stimulation of root growth requires verification with higher SA concentrations.

*Conclusion:* At least in some species, it was demonstrated that SA is a hormetic regulator, having a biphasic dose response, where low and high concentrations promote and inhibit growth respectively. This allows SA to control the growth-stress response balance in root system development.

*Acknowledgements:* This work was supported by the RSF grant No. 20-14-00140.

## References

1. Zhang Y. et al. Salicylic acid: Biosynthesis, perception, and contributions to plant immunity. *Curr Opin Plant Biol.* 2019;50:29-36.

2. Rivas-San Vicente M. et al. Salicylic acid beyond defence: Its role in plant growth and development. *J Exp Bot.* 2011;62:3321-3338.
3. Bagautdinova Z. et al. Salicylic Acid in Root Growth and Development. *Int J Mol Sci.* 2022;23(4):2228.
4. Pasternak T. et al. Salicylic acid affects root meristem patterning via auxin distribution in a concentration-dependent manner. *Plant Physiol.* 2019;180(3):1725.
5. Armengot L. et al. Functional interplay between protein kinase CK 2 and salicylic acid sustains PIN transcriptional expression and root development. *Plant J.* 2014;78(3):411.
6. Echevarria-Machado I. et al. Responses of transformed *Catharanthus roseus* roots to femtomolar concentrations of salicylic acid. *Plant Physiol Biochem.* 2007;45(6-7):501.

## **Role of the formation of fibers with a tertiary cell wall in the biomechanical system of the whole plant**

Petrova A.\*, Kozlova L., Gorshkova T.

*Kazan Institute of Biochemistry and Biophysics, FRC KazSC RAS, Kazan, Russia*

\* *anna.an.petrova@gmail.com*

The morphogenesis of the plant is based on changes in biomechanical properties of its cell walls. In developing organs where cells that differ in their geometry and physical properties of their cell walls increase their size, a complex and dynamic picture of stress distribution is observed. However, this picture is not much simpler in mature organs where none or few cells are still capable of growing. One of the basic approaches which are used by plants to reshape their grown body parts is the formation of fibers with the gelatinous or tertiary cell walls. The fibers with tertiary cell walls in the phloem of many species are not inducible but constitutive. They are especially well pronounced in so-called bast fiber crops like flax or ramie, though their role in the mechanical performance of these herbaceous plants is much less studied than that of tension wood fibers. Using the set of biomechanical approaches with the examination of different mechanical characteristics at macro and nano-scales and considering the flax stem as the compositional material we have shown that the tertiary cell wall fibers indeed develop tension within the stem. The presence of such tension is one of the reasons of fiber deformations that occur upon fiber processing. We also investigated the biomechanical interplay between xylem and phloem during stem development and tracked down their relationships in mutant plants with stunted fibers development. In flax, the tension generated by phloem fibers is partially compensated by the xylem. The leading role in the stem stiffness formation shifts from phloem to xylem along with the plant development. Evidently, these particular complex relationships allow for flax to achieve its exceptional ratio of stem height to diameter remaining sustainable. Similar principles of composite architecture and the mechanical properties of its components can be adopted in artificial biomimetic constructions.

*Acknowledgements:* Work was partially supported by the Russian Science Foundation No. 21-74-10131.

## ***NF-Y* genes in the somatic embryogenesis**

Potsenkovskaia E.<sup>1, 2, 3</sup>, Tvorogova V.<sup>1, 2, 3\*</sup>, Lutova L.<sup>1, 2, 3</sup>

<sup>1</sup> *Department of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia*

<sup>2</sup> *Plant Biology and Biotechnology Department, Sirius University of Science and Technology, Sochi, Russia*

<sup>3</sup> *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

\* *epots556@gmail.com*

**Key words:** transcription factors, *NF-Y*, somatic embryogenesis

**Motivation and Aim:** *NF-Y* are conserved transcription factors of eukaryotes, consisting of three subunits, *NF-YA*, *B* and *C*. In plants, each subunit is encoded by a family of genes, and combinations of different subunits are able to regulate various processes in plant development. One of the *NF-YB* genes in *Arabidopsis thaliana*, the *LEC1* gene, regulates many processes, most of which are associated with embryogenesis [1]. For example, it is involved in the morphogenesis of the embryo, the accumulation of storage substances in the seed, and can also stimulate the development of somatic embryos.

Somatic embryogenesis – the embryo development from vegetative tissues – is an important process for biotechnology. As a regeneration process, it is used for reproduction and for obtaining transgenic plants. For most leguminous plants, despite their importance for agriculture, protocols for obtaining somatic embryos have not been developed and their regeneration, as well as transformation, is not always possible. We hypothesized that, since the *LEC1* protein is involved in embryogenesis in *A. thaliana*, its orthologues in legumes can also stimulate somatic embryogenesis functioning in combination with specific *NF-YA* and *C* subunits. The aim of our study was to find new somatic embryogenesis regulators among *NF-Y* genes in legumes

**Methods and Algorithms:** We used *Medicago truncatula* as model species and applied qPCR analysis, as well as transcriptomic databases analysis, to find *NF-Y* genes with high expression levels during somatic embryogenesis. We also used yeast two-hybrid system to find potential partners among *M. truncatula* *NF-Y* proteins.

**Results:** We found several genes from *NF-Y* family which are expressed during somatic embryogenesis and during seed development. T0 calli with overexpression of *LEC1* orthologues did not demonstrate increased somatic embryogenesis capacity, but overexpression of one of *NF-YA* genes increased callus weight.

**Conclusion:** The data about *NF-Y* genes participating in the somatic embryogenesis can be used for the search of *NF-Y* complexes involved in this process. We are planning to investigate the effect of co-overexpression of *NF-YA*, *B* and *C* genes on the regeneration processes.

**Acknowledgements:** The study is supported by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2020-922 date 16.11.2020 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”.

### **References**

1. Jo L. et al. Combinatorial interactions of the *LEC1* transcription factor specify diverse developmental programs during soybean seed development. *PNAS*. 2020;117(2):1223-1232.

## Understanding the role of *MAKR6* in *Arabidopsis thaliana* L. root development

Sidorenko A.D.<sup>1,2\*</sup>, Novikova D.D.<sup>3</sup>, Mironova V.<sup>1,2</sup>, Zemlyanskaya E.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *University of Lausanne, Lausanne, Switzerland*

\* *a.sidorenko1@g.nsu.ru*

**Key words:** root, vascular tissue, xylem, auxin, MEMBRANE-ASSOCIATED KINASE REGULATOR

**Motivation and Aim:** The MEMBRANE-ASSOCIATED KINASE REGULATOR (MAKR) family is a recently discovered protein family, which includes seven members [1]. Its constituent proteins are essential regulators of plant development. BKI1 and MAKR1 regulate brassinosteroid signaling. MAKR2, MAKR4 and MAKR5 control root gravitropism, lateral root formation, and formation of root protophloem respectively. At the same time, biological functions of the remaining paralog, MAKR6, are still unknown. Here we infer *MAKR6* function based on the analysis of its expression pattern and the promoter analysis.

**Methods and Algorithms:** We measured the overall expression level of *MAKR6* using RT-qPCR. Using Agrobacterium-mediated transformation, we generated *Arabidopsis thaliana* reporter lines *pMAKR6:nls3GFP*, in which *pMAKR6* promoter regulated green fluorescent protein (GFP) expression. To detect the reporter signal, we used epifluorescent and confocal microscopy techniques. The list of transcription factors (TFs) – potential regulators of *MAKR6* expression – was compiled by superimposing the coordinates for *MAKR6* promoter and publicly available DAP-seq peaks for 365 TFs [2]. **Results:** We demonstrated that *pMAKR6* promoter is active in above and below ground parts of *A. thaliana* seedlings with cell-type specificity. Using RT-qPCR, we demonstrated that *MAKR6* is an early auxin response gene, which is upregulated after 30 minutes of auxin application. In the reporter line, auxin treatment enhanced GFP signal in both roots and shoots, and induced ectopic expression of GFP in the root differentiation zone. The potential binding regions for a number of known auxin response regulators (including ARF5) were found in *pMAKR6* promoter.

**Conclusion:** Taken together, our results evidence that *MAKR6* might be a new auxin responsive regulator of plant development in *A. thaliana*.

**Acknowledgements:** This work was supported by the RSF grant No. 21-14-00240.

### References

1. Novikova D.D. et al. Meet your MAKR: the membrane-associated kinase regulator protein family in the regulation of plant development. *FEBS J.* 2021. Epub ahead of print.
2. O'Malley R.C. et al. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell.* 2016;165(5):1280-1292.
3. De Rybel B. et al. Plant vascular development: from early specification to differentiation. *Nat Rev Mol Cell Biol.* 2016;17(1):30-40.

## Control of *Arabidopsis thaliana* shoot gravitropism at the level of cell wall biomechanics and metabolomics

Suslov D.<sup>1\*</sup>, Pozhvanov G.<sup>1,2</sup>, Lipchinsky A.<sup>1</sup>

<sup>1</sup> St. Petersburg State University, St. Petersburg, Russia

<sup>2</sup> Komarov Botanical Institute, RAS, St. Petersburg, Russia

\* d.suslov@spbu.ru

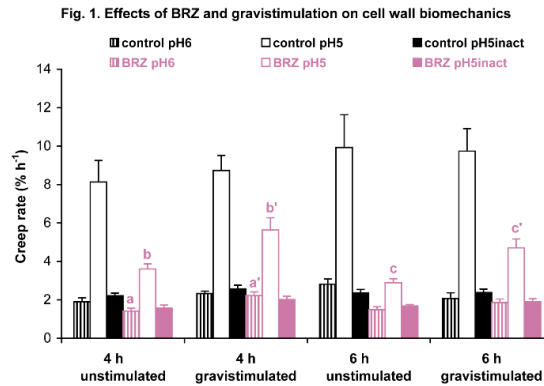
**Key words:** gravitropism, cell wall, extensibility, mannans, biomechanics, creep test, metabolomics

**Motivation and Aim:** Gravitropism is directed growth of a plant or plant organ in response to gravity, resulting in their bending. Plant shoots typically demonstrate negative gravitropism after inclination, i.e. bending upward in the direction opposite to the gravity vector. The bending results from different cell expansion rates on upper and lower flanks of the inclined shoot. It is controlled not only by cell wall extensibility, the main biophysical growth-limiting factor, but also by the flexibility of cell wall polymers and their compression resistance on the concave side of the gravistimulated shoot. Thus, plant shoot gravitropism is a convenient model to decipher the functions of cell wall polymers in plant growth and development. We have recently demonstrated that brassinazol (BRZ), an inhibitor of brassinosteroid phytohormone biosynthesis, stimulated *Arabidopsis* hypocotyl gravitropism, which was accompanied by a decrease in the level of mannans, minor hemicellulosic polysaccharides of cell walls [1]. Now we tried to establish if the effect of BRZ of shoot gravitropism was actually mediated by mannans. With this aim we compared cell wall biomechanics and metabolomics of hypocotyls in BRZ-treated Col-0 *Arabidopsis* plants and *csla2csla3csla9* (*csla2,3,9*) triple mutants, which are devoid of cell wall mannans.

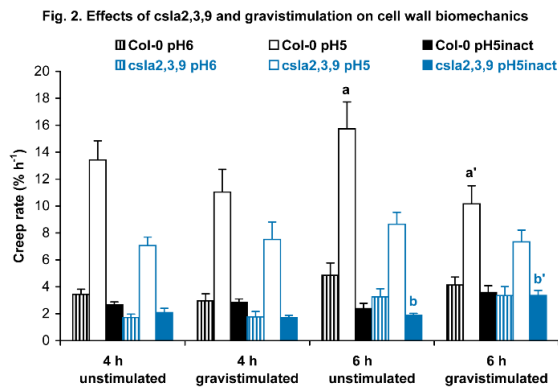
**Methods and Algorithms:** Etiolated *Arabidopsis* seedlings were grown on vertical Petri plates with and without gravistimulation by a 90° counterclockwise rotation of the plates. Metabolites extracted from hypocotyls were analyzed by GC/MS. Cell wall biomechanics were studied by the creep method that measures creep, i.e. time-dependent deformation of the wall samples under a constant load. Creep of frozen/thawed hypocotyls was measured at pH 5 and 6 that activate different groups of cell wall proteins regulating its extensibility, and at pH 5 with heat-inactivation where cell wall proteins are heat-killed, and the wall extension is defined by the mechanics of its structural polymers and the character of static interactions between them.

**Results:** BRZ (0.2 μM) inhibited elongation of Col-0 *Arabidopsis* hypocotyls grown vertically, but greatly increased their bending after gravistimulation. In the presence of BRZ gravistimulated bent hypocotyls were significantly longer than unstimulated straight hypocotyls, suggesting that BRZ activates *Arabidopsis* shoot gravitropism via an overall increase in cell expansion rate. BRZ-treated gravistimulated hypocotyls had significantly higher cell wall creep rates than their unstimulated counterparts at pH 5 and 6 (Fig. 1). This indicates that cell wall loosening proteins active at pH 5 and pH 6 were involved in the mechanism of BRZ action on shoot gravitropism. BRZ changed the levels of many metabolites including the building blocks of cell wall structural polysaccharides (arabinose, xylose, galacturonic acid) and glycoproteins (oxyproline, proline). However, we observed no significant differences in metabolites between BRZ-

treated gravistimulated and unstimulated Col-0 hypocotyls. Apparently, the profound effect of brassinosteroid biosynthesis inhibition strongly outweighed and masked any changes in metabolites associated with gravistimulation.



Hypocotyls of *csla2,3,9* mannan-deficient mutants had the same length as in wild type Col-0 plants. By analogy with BRZ, the triple mutation stimulated Arabidopsis shoot gravitropism, although to a lower extent than BRZ. Biomechanical studies of *csla2,3,9* hypocotyls revealed significant differences at pH 5 with heat-inactivation 6 h after gravistimulation (Fig. 2). Hence, the effect of *csla2,3,9* on shoot gravitropism is associated with a change in basic cell wall polymer structure. Metabolome analysis shows that structural cell wall glycoproteins could mediate this effect, as oxyproline, their hallmark component, was significantly decreased in gravistimulated *csla2,3,9* vs. gravistimulated Col-0 hypocotyls.



**Conclusion:** We established that BRZ and mannan depletion had very different effects on the wall biomechanics and metabolome of Arabidopsis hypocotyls. This shows that the effect of BRZ on shoot gravitropism is not mediated by mannans.

**Acknowledgements:** The study is supported by the Russian Foundation for Basic Research (grant 19-04-00424-a).

## References

1. Somssich M. et al. Brassinosteroids Influence Arabidopsis Hypocotyl Gravitropism through Changes in Mannans and Cellulose. *Plant Cell Physiol.* 2021;62(4):678-692.



# Transcriptomic analysis for the optimization of transformation of legumes

Tvorogova V.<sup>1,2,3\*</sup>, Baltin S.<sup>1</sup>, Efremova E.<sup>1</sup>, Dubenko T.<sup>3</sup>, Krasnoperova E.<sup>1</sup>, Kozlov N.<sup>1</sup>, Potsenkovskaia E.<sup>1,2,3</sup>, Simonova V.<sup>2</sup>, Yakovleva D.<sup>1</sup>, Zlydneva N.<sup>1,3</sup>, Lutova L.<sup>1,2,3</sup>

<sup>1</sup> Department of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia

<sup>2</sup> Plant Biology and Biotechnology Department, Sirius University of Science and Technology, Sochi, Russia

<sup>3</sup> N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia

\* [krubaza@mail.ru](mailto:krubaza@mail.ru)

**Key words:** transcription factors, plant regeneration, legumes

**Motivation and Aim:** Legumes have multiple benefits for agriculture due to their nutritional qualities and ability to fix atmospheric nitrogen. Although there is a need for genetic modification of different legume species, transformation of most of them is hampered by their low regeneration capacity. The aim of this study is to discover new genes which can affect regeneration, unravel the mechanisms of their functioning and check the possibility of their usage for the optimization of regeneration and/or transformation protocols.

**Methods and Algorithms:** We used *Medicago truncatula* as model leguminous species, because it has small sequenced genome and lines, differing in their regeneration capacity. For that species, somatic embryogenesis, i.e. development of embryo-like structures from somatic cells, is the main regeneration pathway. We apply transcriptomic analysis and CHIP-seq methods to embryogenic and non-embryogenic *M. truncatula* lines to find new potential regulators of somatic embryogenesis.

**Results:** We found several genes from *WOX* family which are able to affect somatic embryogenesis being overexpressed. Some of them, like the *MtWOX2* gene, induce callus formation, whereas other, like *MtWOX1* or *MtWOX9-1*, stimulate somatic embryogenesis itself. Interestingly, when compared with control, *MtWOX2* and *MtWOX9-1* overexpressing calli have partly opposite transcriptomic changes. We also found several participants of somatic embryogenesis among members of the NF-Y family of transcription factors.

**Conclusion:** The results obtained on *M. truncatula* allowed us to find several new potential regeneration regulators. Their ability to stimulate regeneration is currently being checked on non-model leguminous species, like *Lens culinaris*, *Pisum sativum* and *Trifolium repens*.

**Acknowledgements:** The study is supported by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2020-922 date 16.11.2020 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”.

## Ethylene assists auxin in the regulation of root cap renewal in *Arabidopsis thaliana* L.

Ubogoeva E.<sup>1\*</sup>, Savina M.<sup>1</sup>, Mironova V.<sup>1,2</sup>, Zemlyanskaya E.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* ubogoeva@gmail.com

**Key words:** auxin, ethylene, root cap, apical root meristem

**Motivation and Aim:** In plants, a root cap (RC) is an extraordinary organ, which continuously grows while maintaining a constant size. The tissue size homeostasis is due to a synchronization of columella stem cell proliferation and cell detachment at the periphery of RC, and it is essential for the root functioning. Auxin gradient in the RC is a key regulator, which couples the two processes. It is assumed that cell division occurs at an auxin maximum, and detachment – at a minimum [1]. However, the mechanisms of how auxin coordinates the two processes are unknown. Here we accomplish *in vivo* and *in silico* analysis of root cap self-renewal in *Arabidopsis thaliana* to get a new insight into the underlying mechanisms.

**Methods and Algorithms:** For *in silico* analysis, we developed a hybrid 1D model of auxin-regulated cell dynamics in columella, which implements “division at a maximum – detachment at a minimum” concept [2]. Confocal and light microscopy techniques were used to estimate RC cell dynamics *in vivo*.

**Results:** To formally describe RC cell dynamics observed both *in planta* and in a computer simulation, we used easily detectable structural parameters: the number of attached/detached/total RC cell layers. Comparing the results of *in vivo* experiments and mathematical modeling, we found that the model properly describes RC cell dynamics only under normal conditions but not under auxin excess, implying that “division at a maximum – detachment at a minimum” concept needs a refinement. A plant hormone ethylene is an essential regulator of root development, wherein the two hormones are known to reciprocally regulate each other’s activity [3]. Using 1-aminocyclopropane-1-carboxylic acid (ACC) treatment and a mutant assay, we demonstrated that alterations of both ethylene production and signaling affect root cap dynamics *in planta*. Next, utilizing mathematical modeling, we predicted that simultaneous fine-tuning of auxin effect on cell growth and division could provide for ethylene-related RC phenotypes. Finally, the combined results of *in vivo* and *in silico* analyses of auxin response in mutants deficient in ethylene signaling supported the role of ethylene in auxin-mediated regulation of RC cell dynamics.


**Conclusion:** Taken together, our results demonstrate that ethylene assists auxin in the regulation of root cap renewal in *Arabidopsis thaliana*.

**Acknowledgements:** This work was supported by the RSF grant No. 20-14-00140.

### References

1. Dubreuil C. et al. A local auxin gradient regulates root cap self-renewal and size homeostasis. *Curr Biol.* 2018;28(16):2581-2587.e3.
2. Savina M.S. et al. Cell dynamics in WOX5-overexpressing root tips: the impact of local auxin biosynthesis. *Front Plant Sci.* 2020;11:560169.
3. Zemlyanskaya E.V. et al. Deciphering auxin-ethylene crosstalk at a systems level. *Int J Mol Sci.* 2018;19(12):4060.

# 7 Symposium “Animal genetics, bioinformatics and systems computational biology”



- 7.1 Section “Animal models of pathologies” [667](#)
- 7.2 Section “Population and evolutionary genomics/genetics of wild and domestic animals” [687](#)
- 7.3 Section “Neurogenomics and genetics of behavior” [711](#)
- 7.4 Section “Neurogenomics of behavior” [726](#)

# The gene expression fluctuation asymmetry as an indicator of development instability in different MHC compatibility models of mother-fetus interactions in mouse strains

Babochkina T.I.\*, Gerlinskaya L.A., Anisimova M.V., Kontsevaya G.V., Feofanova N.A., Stanova A.K., Orbant M.O., Moshkin M.P., Moshkin Y.M.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* babochkinat@yahoo.com

**Key words:** fluctuation asymmetry (FA), MHC compatibility, development instability

*Motivation and Aim:* Development destabilization can be indicated by fluctuating asymmetry (FA) of bilateral traits [2, 4].

At the same time, similarity or difference between the mother and the fetus with respect to major histocompatibility complex (MHC) genes play key role in development [1, 3]. The aim of our work is to study the effect of the similarity/difference of the MHC genes of the mother and fetus on the destabilization of prenatal development on models of different MHC gene combinations between mother and fetus organisms which presented the different variants of immunogenetic dialogue between the mother and fetuses. As an indicator of development instability, we used the gene expression FA in bilateral structures of the fetus.

*Methods and Algorithms:* To model different variants of the MHC compatibility, we transferred two-cell embryos of the C57BL/6J inbred strain to surrogate mothers of either the same or different haplotype for MHC genes. This allowed us to model three variants of mother–fetus immunogenetic dialogue: 1) bidirectional immunogenetic dialogue between C57BL/6J embryos of H2b MHC haplotype and BALB/c surrogate mothers of a distinct H2d MHC haplotype; 2) unidirectional immunogenetic dialogue between C57BL/6J embryos and immunodeficient NOD SCID surrogate mothers of H2g7 MHC haplotype; and 3) reduced immunogenetic dialogue between same genotype C57BL/6J mother and fetuses.

We utilized the FA analysis of gene expression and PLGF, VEGF concentration and fetoplacental indexes and body weight of the embryos as indicators of development destabilization.

For gene expression FA analysis, an analytical algorithm was developed based on the decomposition of dispersions by the method of principal components (PCA).

*Results:* It was found that the highest level of FA occurred during intra line transplantations, that is in the absence of an immunogenetic dialogue between the fetus and mother, and the least when transferring C57BL/6J embryos to BALB/c mothers, that is full-fledged dialogue in the mother-fetus system. The transfer of allogeneic embryos to immunodeficient mothers (C57BL/6J in NodScid) gave an intermediate result.

At the same time, intra-individual gene expression FA variations as an indicator negatively correlated with the body weight of the embryos and the content of the Placental Growth Factor (PIGF) in the placenta.

The second important result is that the body mass of fetuses and the level of PLGF were negatively correlated to the values of the expression development noise.

*Conclusion:* The same inbred mouse strain carried by surrogate mothers (syngeneic pregnancies) with the same MHC haplotypes (histo-MHC-compatible pregnancies) demonstrated higher internal expression noise. On the contrary, the different inbred mouse strain carried by surrogate mothers (allogeneic pregnancies) with distinct MHC haplotypes demonstrated the lower internal expression noise and an advantage in the conditions of early embryo development. And the individual variations in internal genetic noise correlated with indicators reflecting the growth of embryos (body mass of fetus) and the functioning of the fetal-placental complex (PLGF concentration).

Thus, the gene expression FA in the bilateral structure of the fetus associated with developmental conditions depends significantly on MHC compatibility between fetus and surrogate mothers.

*Acknowledgements:* The research was supported by the Russian Science Foundation grant No. 20-14-00055, budget project of Ministry of Science and Higher Education of the Russian Federation No. 0259-2021-0015, 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).

### *References*

1. Billington W.D. Influence of immunological dissimilarity of mother and foetus on size of placenta in mice. *Nature*. 1964;202:317-318.
2. Carter A.J., Weier T.M., Houle D. The effect of inbreeding on fluctuating asymmetry of wing veins in two laboratory strains of *Drosophila melanogaster*. *Heredity*. 2009;102(6):563-572.
3. Gerlinskaya L.A., Litvinova E.A., Kontsevaya G.V., Feofanova N.A. et al. Phenotypic variations in transferred progeny due to genotype of surrogate mother. *Mol Hum Reprod*. 2019;25(2):88-99.
4. Meagher S., Penn D.J., Potts W.K. Male-male competition magnifies inbreeding depression in wild house mice. *Proc Natl Acad Sci U S A*. 2000;97(7):3324-3329.

# Chromosome synapsis and recombination in intraspecific and interspecific heterozygotes for chromosomal rearrangements in voles of the genus *Alexandromys*

Bikchurina T.I.<sup>1,2\*</sup>, Rubtsova D.V.<sup>1</sup>, Kartavtseva I.V.<sup>3</sup>, Sheremetyeva I.N.<sup>3</sup>, Pavlenko M.V.<sup>3</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup>Federal Scientific Center of the East Asia Terrestrial Biodiversity, FEB RAS, Vladivostok, Russia

\* [bikchurina@bionet.nsc.ru](mailto:bikchurina@bionet.nsc.ru)

**Key words:** meiosis, hybrid sterility, chromosomal rearrangements, voles

**Motivation and Aim:** In the early stages of speciation, an increase of genetic or karyotypic divergence can lead to the formation of postzygotic reproductive isolation. In mammals, one of the key mechanisms of reproductive isolation is hybrid sterility that occurs gradually. Closely related vole species of the genus *Alexandromys* (Rodentia; Arvicolinae) provide a convenient model for studying the formation of hybrid sterility between species with a high level of chromosomal polymorphism. Heterozygotes for chromosomal rearrangements form different synaptic configurations in prophase I of meiosis that leads to synaptic and recombination disorders. These aberrations, in turn, lead to the apoptosis of gametocytes and stop of gametogenesis.

**Methods and Algorithms:** To elucidate the cytological mechanisms of the formation of hybrid sterility between three species of the genus *Alexandromys*, we analyzed the dynamics of spermatogenesis in intraspecific (*A. evoronensis*) and interspecific male hybrids (between *A. evoronensis*, *A. maximowiczii* and *A. mujanensis*). Using immunolocalization of key meiotic proteins, we analyzed synapsis and recombination at pachytene stage in hybrids and parental species.

**Results:** Despite the formation of a large number of heteromorphic bivalents and trivalents, the frequency of recombination events per cell in male intraspecific hybrids of the *A. evoronensis* was comparable to that of the parental species. We also did not observe negative heterosis for this trait in interspecific hybrids between the *A. mujanensis* and *A. maximowiczii*. However, a number of different meiotic anomalies between interspecific hybrids of different interspecific crosses was found. According to the results of histological analysis, intraspecific hybrid males of the *A. evoronensis* were fertile, while interspecific hybrid males of three closely related species of *Alexandromys* voles were sterile with spermatogenesis stopped during meiosis.

**Conclusion:** Multiple disturbances in autosomal synapsis and an increased frequency of sex chromosome asynapsis seemed to be the main cause of meiotic arrest and spermatocyte apoptosis in interspecies hybrids. The severity of meiotic disorders in the vole hybrids analyzed varied significantly, which reflects the presence of chromosomal polymorphism in the two out of three species examined, as well as the phylogenetic similarity of all three species and the high rate of karyotypic divergence.

**Acknowledgements:** The study is supported by the grants from the Ministry of Science and Higher Education of the Russian Federation No. 0259-2021-0011 and No. 2019-0546 (FSUS-2020-0040).

## Production of fox FGF2 for fox pluripotent stem cell culture

Meshcheryakov N.<sup>1</sup>, Smirnov A.<sup>2</sup>, Menzorov A.<sup>1,2\*</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* menzorov@bionet.nsc.ru

**Key words:** FGF2, *Vulpes vulpes*, iPSC, HEK293-EBNA1

**Motivation and Aim:** Embryonic stem cells and induced pluripotent stem cells of various non-human species are widely used to study embryonic development and differentiation. Pluripotent stem cells usually require specific growth factors such as LIF (mouse pluripotent stem cells) and FGF2 (human pluripotent stem cells). Human or mouse growth factors are successfully used for cell culture of pluripotent stem cells of different species. However, cat induced pluripotent stem cells required species-specific feline FGF2 [1], thus species-specific LIF or FGF2 may be beneficial. In our attempts to produce fox induced pluripotent stem cells we observed more colonies after partial reprogramming with human FGF2 than with mouse LIF. Thus, we decided to develop a system for fox FGF2 production for fox pluripotent stem cell culture.

**Methods and Algorithms:** We utilized the HEK293-EBNA1-based system [2] for protein production. EBNA1 expressing 293 c18 cell line could be used for efficient production of proteins with expression vectors bearing the Epstein-Barr virus origin of replication, oriP. Proteins could be secreted into the cell culture medium and in case of high yield be used diluted for cell culture without protein purification. To produce fox FGF2 we used pCEP-Pu-BM40-6His-Myc-fX-mLIF plasmid, a kind gift from Dr. V. Wixler, the University of Munster, Germany. We extracted RNA from the fox frontal lobe (a kind gift from Dr. Yu. Herbek, Institute of Cytology and Genetics, Novosibirsk, Russia), produced cDNA, and amplified the 465 bp coding part of fox *FGF2* gene (Gene ID: 112933647) (primers 5'-GTG GGC TAG CCG CAG CCG GGA GCA TCA CCA C-3' and 5'-GTC TGG ATC CTC AGC TCT TAG CAG ACA TTG GAA-3'). PCR-product was ligated into pCEP-Pu-BM40-6His-Myc-fX-mLIF using *NheI* and *BamHI* endonucleases. We designated the plasmid as pCEP-Pu-BM40-6His-Myc-fX-foxFGF2.

**Results:** We transfected 293 c18 cells with pCEP-Pu-BM40-6His-Myc-fX-foxFGF2 and produced 10 clones. We have shown that the three clones with the highest growth rate all express fox *FGF2*. Next, we determined the fox FGF2 protein concentration in DMEM/F12 medium conditioned by the FGF2 producing cells using His Tag ELISA Detection Kit (GenScript, USA). The highest concentration was 19 ng/ml. Human pluripotent stem cells are cultured with 10 ng/ml, thus 19 ng/ml is most probably not enough for cell culture of fox pluripotent stem cells using a dilution of conditioned medium, protein concentration and purification are necessary

**Conclusion:** We produced a cell line with fox FGF2 expression. The protein yield in the conditioned medium is low for the direct usage with fox pluripotent stem cells, additional steps of concentration and purification are necessary.

**Acknowledgements:** The study was funded by RFBR, project No. 20-04-00369, and the Ministry of Education and Science of the Russian Federation, state project FWNR-2022-0019.

### References

1. Dutton L., Dudhia J., Guest D.J. *Stem Cells Dev.* 2019;28:1299-1309.
2. Tom R., Bisson L., Durocher Y. *CSH Protoc.* 2008;pdb.prot4976.

## Quantitative estimation of trisomy on chromosome 6 in embryonic fibroblasts of mice carrying duplications obtained using CRISPR/Cas9 technology

Minina J.M.<sup>1\*</sup>, Karamysheva T.V.<sup>1</sup>, Serdyukova N.A.<sup>2</sup>, Serov O.L.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

\* minina\_jul@mail.ru

**Key words:** CRISPR/Cas9, *Cntn6* duplication, trisomy in mice, FISH analysis

**Motivation and Aim:** The high efficiency of the CRISPR/Cas9 system used for editing human and animal genomes is achieved due to the sufficiently accurate recognition of the target DNA site by the guiding RNA of the CRISPR/Cas9 complex. However, editing errors can occur during the repair of double-stranded DNA breaks performed by the Cas9 protein. To date, there is increasing evidence of spontaneous gene mutations occurring not only away from the genome editing site, but also close to it. Previously, using the CRISPR/Cas9 technology in the zygotes, we obtained experimental mice that carry deletions with a size of 1137 kb, duplication of 2374 kb, and similar sized inversions in chromosome 6, involving only the *Cntn6* gene (encoding a protein contactin-6) [1]. Applying FISH analysis we unexpectedly found trisomy on chromosome 6 in 3.5–4.3 % of fibroblasts obtained from two adult founders (No. 1 and 20) heterozygous for the *Cntn6* duplication [2]. Taking into account that this is the first time when the phenomenon of trisomy in somatic cells is described in animals with CRISPR/Cas9 induced large-scale chromosomal rearrangements, we assumed that a possible factor provoking the onset of trisomy was a heterozygous duplication of the *Cntn6* gene in the genome.

**Methods and Algorithms:** Applying the method of fluorescent in situ hybridization (FISH) with a chromosome-specific probe for the entire mouse chromosome 6 we investigated the level of trisomy in 21 lines of mouse embryonic fibroblasts (MEF) obtained from 13-day-old embryos that are descendants of founders (No. 1 and 20).

**Results:** In 8 studied lines of MEFs heterozygous for *Cntn6* duplication the level of trisomy varied from 1 to 8 %, and in 5 homozygous lines – from 2 to 6 %. There were no statistically significant differences between these MEF lines (Mann–Whitney test, STATISTICA 10). In 3 out of 4 control MEF lines carrying the wild type allele of *Cntn6* gene, trisomy was detected in 1–2 % of cells. In order to "enhance" the heterozygosity of alleles of chromosome 6, and to see if the level of trisomy will increase, we generated and analyzed 4 MEF lines carrying heterozygous *Cntn6* duplication and *Cntn6* deletion in one genome. FISH with a chromosome-specific probe for the mice chromosome 6 showed the absence of trisomy in these 4 lines.

**Conclusion:** The obtained results demonstrate that the level of trisomy of chromosome 6 in MEF lines heterozygous and homozygous for *Cntn6* duplication is not significantly increased compared to that in control cell lines. Nevertheless, the variability of the trisomy level of chromosome 6 among heterozygous and homozygous for *Cntn6* duplication cell lines, as well as the level of total aneuploidy in these MEF lines indicates



increased genome instability. It is quite possible that trisomy on chromosome 6 in studied cell lines occurs at a high level, but is accompanied by rapid elimination of aneuploid cells in early embryo development [3]. Apparently, the presence of *Cntn6* duplication and *Cntn6* deletion on different homologues is a more stable modification for the genome than when the dose of the *Cntn6* gene in the genome is 1.5 or 2 times higher (in heterozygous and homozygous *Cntn6* MEFs).

*Acknowledgements:* The study is supported by Russian Foundation for Basic Research grant No. 20-04-00463A.

### *References*

1. Korablev A.N., Serova I.A., Serov O.L. Generation of megabase-scale deletions, inversions and duplications involving the Contactin-6 gene in mice by CRISPR/Cas9 technology. *BMC Genet.* 2017;18:112.
2. Pristyazhnyuk I.E., Minina J., Korablev A., Serova I., Fishman V. et al. Time origin and structural analysis of the induced CRISPR/cas9 megabase-sized deletions and duplications involving the *Cntn6* gene in mice. *Sci Rep.* 2019;9(1):14161.
3. Singla S., Iwamoto-Stohl L.K., Zhu M., Zernicka-Goetz M. Autophagy-mediated apoptosis eliminates aneuploid cells in a mouse model of chromosome mosaicism. *Nat Commun.* 2020;11(1):2958.

## Developmental instability in mouse conceived by *in vitro* fertilization

Moshkin M.P. \*, Gerlinskaya L.A., Moshkin Y.M.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [mmp@bionet.nsc.ru](mailto:mmp@bionet.nsc.ru)

**Key words:** *in vitro* fertilization, developmental instability, fluctuation asymmetry, gene expression, brain metabolites, spontaneous locomotion

Assisted reproductive technology (ART), including *in vitro* fertilization (IVF), is a key methodology for supporting genetic strains of laboratory animals, which number overloads the capacity of breeding facilities. Also IVF is a central technology in genomic editing. Since IVF could be reason for phenotypic shifts in next generations, study of the IVF induced epigenetic effects became actual for breeding of laboratory and domestic animals, as well as human reproduction.

In experiments on outbred strain mouse CD1 we study variability of methylation rate, gene expression, metabolic process, spontaneous behavior and fluctuating asymmetry (FA) in bilateral structure as an indicator of developmental instability. IVF leads to increase of developmental instability detected by different indicators, which were manifested at all stage of ontogenesis – from preimplantation development up to mature age.

*Acknowledgements:* Supported by RSF grant No. 20-14-00055.

## Hereditary predisposition to seizures in response to handling and its relation to life expectancy in the water vole (*Arvicola amphibius* L.)

Nazarova G.\*, Proskurnyak L.

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

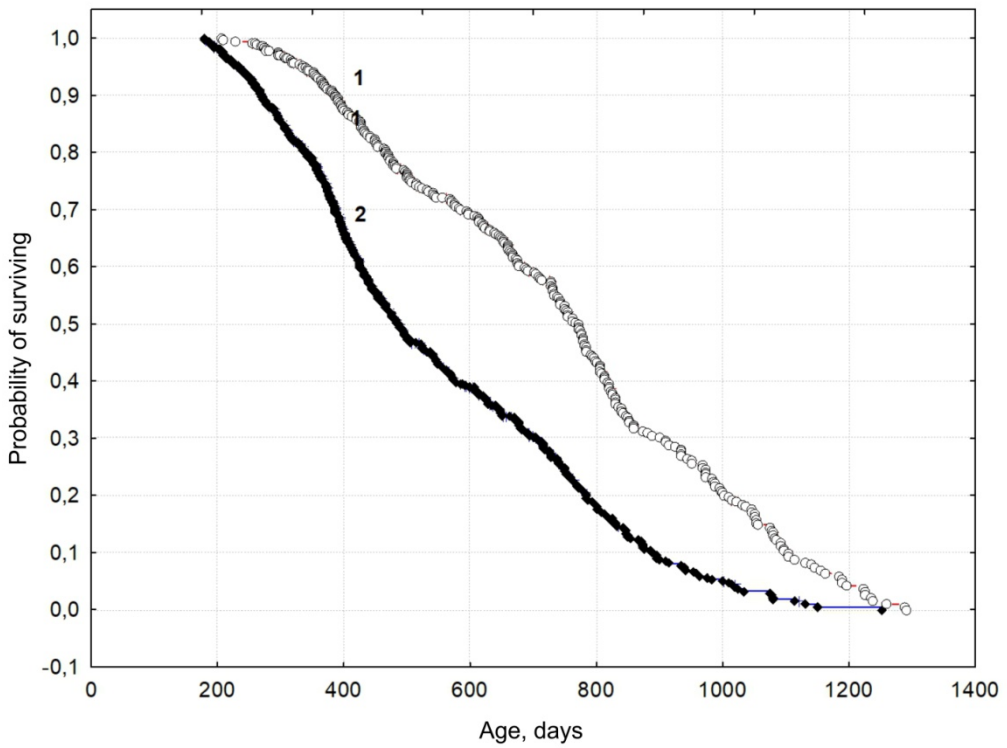
\* [galinanazarova@mail.ru](mailto:galinanazarova@mail.ru)

**Key words:** water vole, seizures, age, sex, population cycle, hereditary predisposition, maternal influence, life expectancy

*Motivation and Aim:* Individuals predisposed to seizures were found in natural and captive-bred populations of water voles. Finding out the genetic factors of predisposition to seizures in response to specific external stimuli (such as flashes of light, auditory, olfactory, or tactile stimuli) is important for understanding the causes of seizures and developing new methods for prevention and therapy of convulsive states.

*Methods and Algorithms:* The data of long-term maintenance and breeding of water voles in vivarium were analyzed in order to find out a hereditary predisposition to convulsive seizures in response to handling, and the influence of sex and age on their development.

*Results:* Under vivarium conditions, seizures in water voles were provoked by handling. A seizure typically commenced with jerks with the vibrissae, gradually increasing in magnitude. The next stages were a tonic tension of the muscles of the anterior part of the body, rhythmic movement of limbs, “wild” running. Seizures were observed in 2.4 % of voles taken from the natural population with cyclic fluctuations in abundance and in 29.8 % of descendants of captive-born water voles. The minimum age of the first observation of seizures was 39 days, the maximum age – 1105 days, and the median age – 274 days. Predisposition to seizures was not related to sex. We hypothesize that the seizure is a behavioral anti-predator adaptation that evolved through evolution. Seizure-prone individuals may have a selective advantage at the peak phase of population cycle, when predation increases. Propensity to seizures is a heritable trait. The proportion of individuals prone to seizures increased significantly if one or both parents had seizures. In parent–offspring pairs, a positive correlation was found between the average age of the onset of the first seizures in parents and their offspring,  $r = 0.42$ ,  $p < 0.01$ . Genes, that control the occurrence of seizures, have a pleiotropic effect on life expectancy: seizure-prone individuals live longer, than individuals with a normal phenotype (Fig. 1). The presence of an interconnectedness of the effects of genes that control the occurrence of seizures and life expectancy confirms the data on the effect of maternal convulsive status on the life expectancy of offspring. The offspring of seizure-prone mothers live longer than the offspring of mothers with a normal phenotype. The life expectancy does not depend on the convulsive status of the father. It is possible that predisposition of females to seizures is genetically correlated with physiological parameters during pregnancy, lactation, or maternal behavior, which, in turn, determine the life expectancy of the offspring. Thus, the water vole can serve as a suitable model object for studying the causes of convulsive states and the evolution of longevity.



**Fig. 1.** Survival curves for two groups of voles: seizure-prone (1) and seizure-free (2)

*Conclusion:* The data obtained indicate the possibility of selection for the predisposition to seizures in response to handling in water voles, in order to study the mechanisms of epileptogenesis and longevity.

*Acknowledgements:* The study was supported by the budget project No. 122011800268-1.

## Genetic diversity of bryozoan relict *Hislopia placoides* (Gymnolaemata: Hislopiidae) in Lake Baikal

Peretolchina T.<sup>1\*</sup>, Bukin Yu.<sup>1</sup>, Adelshin R.<sup>2</sup>, Sherbakov D.<sup>1</sup>

<sup>1</sup>Limnological Institute, SB RAS, Irkutsk, Russia

<sup>2</sup>Irkutsk Antiplague Research Institute of Siberia and the Far East, Irkutsk, Russia

\*[tatiana.peretolchina@gmail.com](mailto:tatiana.peretolchina@gmail.com)

**Key words:** *H. placoides*, COI, 16S rDNA, mitochondrial DNA, sequencing, MinION

**Motivation and Aim:** Bryozoans (phylum Bryozoa) are simple aquatic invertebrates living in sedentary colonies. The bryozoans are classified as the Stenolaemata, Phylactolemata and Gymnolemata. Members of Gymnolemata usually occur in seas and brackish water, but there are families, inhabiting fresh water (Paludicellidae, Hislopiidae). The only representative of Gymnolemata described from Lake Baikal is endemic species *Hislopia placoides* (Korotneff, 1901). In Lake Baikal, this species is regarded as a relic of past, warmer, geological periods. *H. placoides* is characterized by ecological plasticity. In the lake Baikal, several morphs of *H. placoides* have been described that differ from each other in the form of colonies and zooids, confined to certain habitat conditions. At the same time, the genetic diversity of *H. placoides* morphs has not yet been studied. Thus, the aim of this investigation was to study the genetic polymorphism and evolutionary history of different morphs of *H. placoides*.

**Results:** Representatives of two morphs of *H. placoides* were studied. *H. placoides* m. *sabulosa* were collected in the open littoral of Lake Baikal (near Babushkin, depth 18–60 m). *H. placoides* m. *ripariensis* were collected from Posolsky Sor (depth 1–3 m) and Angara River (depth 3–5 m). A total of 18 bryozoan colonies were analyzed. For molecular genetic studies, the gene fragments of the COI and 16S rDNA of mitochondrial DNA were used. The four COI haplotypes and two 16S rDNA haplotypes were revealed among the *H. placoides* specimens studied. Moreover, the shared haplotypes were found in the different morphs. Thus, morphological variability and geographic distances are not related to genetic polymorphism and, possibly, due to the ecological features of local habitats, inducing the modification variability of the colonies.

In addition, during the DNA sequencing, we encountered ambiguous peak profile on the some electrophoregrams, which did not allow us to build a nucleotide sequence. To solve this problem, we performed nanopore sequencing of two colonies for the COI marker and one colony for the 16S marker using the portable DNA-sequencer MinION (Oxford Nanopore Technologies). A total of 146,117 and 89,504 reads were obtained for COI amplicon, and 346,016 reads were revealed for 16S rDNA amplicon. Analysis of the nucleotide sequences showed that within each colony, both for the COI amplicon and for the 16S rDNA amplicon, reads were reliably divided into two clusters in a ratio of 1:4. This indicates the presence of mitochondrial DNA polymorphism within each of the clones studied. It is possible that the obtained results are related to the fact that mitochondria in *H. placoides* are inherited both maternally and paternally during sexual reproduction. The method of sequencing of the single molecules through the MinION has shown its effectiveness for detecting mitochondrial DNA polymorphism within a single colonial organism.

**Acknowledgements:** We are grateful to Igor Khanaev and Ivan Nebesnykh (LIN SB RAS) for the *H. placoides* sampling.

## **ISIAH rat genotype-specific SNPs located at microRNA-binding sites in the 3'UTRs of genes associated with behavior and arterial hypertension**

Redina O.E.\*, Babenko R.O., Ershov N.I., Smolenskaya S.E., Markel A.L.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *oredina@bionet.nsc.ru*

A model of a stress-sensitive arterial hypertension – the ISIAH rat strain is characterized by an increased reactivity of the hypothalamic-pituitary-adrenocortical and sympathetic adrenal systems compared to the control normotensive rats. In addition, ISIAH rats demonstrate enhanced reactivity of locomotion responses and exploratory behavior in various behavioral tests, as well as an increased level of anxiety (emotionality) in an unfamiliar environment. Numerous studies have confirmed the role of miRNAs in various biological processes, including cardiovascular complications. Binding of a mature microRNA integrated into the RNA-induced silencing complex (RISC) to the 3'-untranslated region of the mRNA leads to mRNA degradation or translation inhibition. The aim of this study was to identify SNPs located at microRNA binding sites in the 3'UTR of genes that may contribute to the manifestation of specific phenotypic features of ISIAH rats. Transcriptome sequencing of several organs/tissues of hypertensive ISIAH and normotensive WAG rats revealed 2773 SNPs that are specific for the ISIAH genotype and are located in the 3'UTRs of 1384 genes. According to the Rat Genome Database (<https://rgd.mcw.edu/>), among them, 177 SNPs were assigned to 71 genes associated with hypertension, and 645 SNPs were assigned to 306 genes associated with behavior/neurological phenotype. The miRDB database (<http://mirdb.org/>) was used to identify SNPs that disrupt microRNA binding sites. 11 SNPs were identified at microRNA binding sites in the 3'UTRs of 11 genes associated with hypertension. 59 SNPs were identified at microRNA binding sites in the 3'UTRs of 50 genes associated with behavior/neurological phenotype. According to the Gene Ontology functional annotation, 21 of these genes are associated with neurogenesis (FDR 1.44E-04). Conclusion: In the genotype of hypertensive ISIAH rats, 61 SNPs were identified that can disrupt microRNA binding to 3'UTRs of genes associated with behavior/neurological phenotype and/or arterial hypertension, which may contribute to the regulation of the expression of a number of genes and the manifestation of specific phenotypic features of hypertensive ISIAH rats.

*Acknowledgements:* The work was supported by RFBR grant 20-04-00119a and Budget project FWNr-2022-0019.

## Effect of divalent cations on the uptake of Mn<sub>3</sub>O<sub>4</sub> nanoparticles by olfactory epithelial cells

Romashchenko A.V.<sup>1,2\*</sup>, Sharapova M.B.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Computational Technologies, SB RAS, Novosibirsk, Russia

\* yuter2006@yandex.ru

**Key words:** nanoparticles, olfactory uptake, nose-to-brain transport

*Motivation and Aim:* Inorganic nanoparticles (NPs) are promising tools for the treatment of brain diseases. Given their high surface-to-volume ratio, NPs provide efficient loading of therapeutic molecules, while their material composition and structure properties offer alternative therapeutic possibilities. One of the most promising methods of drug delivery to the brain, bypassing the blood-brain barrier (BBB), is intranasal administration [1–3]. The success of the implementation of the therapeutic effect of NPs is largely determined by the intensity of their capture by target structures, which depends both on the physicochemical properties of the NPs themselves and on the functional state of the tissue. It is known that the transport system of divalent cations plays a key role in the transport of heavy metals from the nasal cavity to the brain. Whether this system is involved in the transport of insoluble NPs of the same metals is not known. The aim of this work was to evaluate the effect of various di-, tri-, and multivalent cations on the uptake of Mn<sub>3</sub>O<sub>4</sub>-NPs by nasal epithelial cells.

*Methods and Algorithm:* The work was performed on Balb/c of SPF-status mice at the age of 8 weeks. The accumulation of NPs in the olfactory bulbs was assessed using T1-weighted images obtained on a BioSpec 17/16USR ultrahigh-field magnetic resonance tomograph (Bruker, Germany) with a magnetic field strength of 11.7 T. Tomography was performed 1 hour after the introduction of NPs. Inhibitors were introduced 5 min before the introduction of Mn<sub>3</sub>O<sub>4</sub>-NPs. The Mn<sub>3</sub>O<sub>4</sub>-NPs used in the work had a particle size (20–80 nm), hydrodynamic diameter (118 ± 15 nm), surface charge (−18 ± 5.1 mV) and r1: 0.395 mM<sup>−1</sup>s<sup>−1</sup>.

*Results:* Based on the data of T1-weighted MRI, the uptake of Mn<sub>3</sub>O<sub>4</sub>-NPs by nasal epithelium cells was affected by the preliminary administration of 1 mM solutions of FeCl<sub>2</sub>, CoCl<sub>2</sub>, ZnCl<sub>2</sub>, but not FeCl<sub>3</sub>, GdCl<sub>3</sub>, CaCl<sub>2</sub>, MgCl<sub>2</sub>. The results obtained correspond to the data on the selectivity of the transport system of divalent cations (Cd<sup>2+</sup> > Fe<sup>2+</sup> > Co<sup>2+</sup>, Mn<sup>2+</sup> >> Ni<sup>2+</sup>, but not Ca<sup>2+</sup>, Mg<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>3+</sup>, Gd<sup>3+</sup>) [4]. In addition, administration of a polyanion (0.1 mM heparin), a polycation (0.1 mM protamine), or inhibitors of L- and N-type voltage-gated calcium channels that regulate the excitability of the olfactory neuron also did not affect the nasal uptake of Mn<sub>3</sub>O<sub>4</sub>-NPs. At the same time, salinomycin, a known inhibitor of the divalent cation transporter (DMT 1) [5], significantly suppressed the uptake of Mn<sub>3</sub>O<sub>4</sub>-NPs.

*Conclusion:* For the first time, we were able to demonstrate the role of the non-transferin transport system of divalent cations in the uptake of insoluble metal oxide NPs by nasal epithelial cells. The results obtained will be useful both in the development of means for protecting the CNS from the penetration of xenobiotics and in solving the problems of

targeted delivery of NPs acting as independent therapeutic agents or "containers" for the drugs.

*Acknowledgements:* The work was carried out within the framework of the Russian Science Foundation grant No. 20-14-00055.

### *References*

1. Khan A.R. et al. Progress in brain targeting drug delivery system by nasal route. *J Control Release*. 2017;268:364-389.
2. Pardo M. et al. Size-dependent intranasal administration of magnetoelectric nanoparticles for targeted brain localization. *Nanomed Nanotechnol Biol Med*. 2021;32:102337.
3. Yokel R.A. Nanoparticle brain delivery: a guide to verification methods. *Nanomedicine*. 2020;15(04):409-432.
4. Torabi M. et al. Time dependent difference effects of MgO and ZnO nanoparticles on the serum and hippocampus Mg<sup>2+</sup>, Zn<sup>2+</sup>, Fe<sup>2+/3+</sup> and Ca<sup>2+</sup> levels in the stressed rats. *Nanomed Res J*. 2018;3(4):197-205.
5. Turcu A.L. et al. DMT1 inhibitors kill cancer stem cells by blocking lysosomal iron translocation. *Chem Eur J*. 2020;26(33):7369-7373.



## The role of manganese oxide nanoparticles in the formation of stress granules in the mouse olfactory system

Sharapova M.B.<sup>1\*</sup>, Morozova K.N.<sup>1,2</sup>, Kiseleva E.V.<sup>1,2</sup>, Romashchenko A.V.<sup>1,3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Federal Research Center for Information and Computational Technologies, SB RAS, Novosibirsk, Russia

\* barberry1505@gmail.com

**Key words:** nanoparticles, stress granules, nose2brain transport

*Motivation and Aim:* In recent years, air pollution has increased significantly, and humans, like other terrestrial animals, are constantly exposed to solid aerosols, which may contain various xenobiotics, including nanoparticles (NPs) that can enter the brain from the nasal cavity [1]. Moreover, it was shown that the addition of nanoparticles to the cellular environment induces the formation of stress granules (SG), which are the cellular response to stress [2]. However, whether nanoparticles cause the formation of SG in brain structures is unknown. The aim of this study was to investigate the ability of intranasally administered Mn<sub>3</sub>O<sub>4</sub>-NPs to induce SG formation in neurons of the mouse olfactory system.

*Methods and Algorithms:* The work was performed on mice of SPF-status strain C57Bl/6 at the age of 8 weeks. The accumulation of NPs in the olfactory bulbs was assessed using T1-weighted images obtained on a BioSpec 17/16USR ultrahigh-field magnetic resonance tomograph (Bruker, Germany) with a magnetic field strength of 11.7 T. The amount of SG was determined on paraffin sections stained with a mixture of antibodies anti-eIF3 $\eta$ +anti-G3BP1. The intracellular localization of nanoparticles was studied using transmission electron microscopy (TEM). Paramagnetic Mn<sub>3</sub>O<sub>4</sub>-NPs were used as nanoparticles.

*Results:* Using T1-weighted MRI, we were able to demonstrate that the accumulation of Mn<sub>3</sub>O<sub>4</sub>-NPs in the main olfactory bulb (MOB) of the mouse reaches a maximum 24 hours after their intranasal application. To estimate the role of manganese ions in the formation of the observed MRI signal, we determined the concentration of Mn in the MOB tissue before and after dialysis in mice that were injected with either Mn<sub>3</sub>O<sub>4</sub>-NPs or MnCl<sub>2</sub> into the nasal cavity 24 hours before the isolation of MOB. As the analysis showed, in animals that were injected with Mn<sub>3</sub>O<sub>4</sub>-NPs, as a result of dialysis, there is practically no decrease in the concentration of manganese in the tissue. Whereas in mice injected with MnCl<sub>2</sub>, dialysis effectively reduces Mn concentrations in the MOB homogenate. This suggests that one day after the intranasal application of Mn<sub>3</sub>O<sub>4</sub>-NPs, manganese in the MOB is mainly in an insoluble form. Using TEM, we demonstrated that nanoparticles can be transported along the axons of olfactory neurons both intra- and extra-vesicularly, which indicates the possibility of contact between Mn<sub>3</sub>O<sub>4</sub>-NPs and the contents of the neuron cytoplasm. Finally, using immunohistochemistry, we were able to show that the accumulation of Mn<sub>3</sub>O<sub>4</sub>-NPs in the MOB can induce the formation of SG in mitral neurons.

*Conclusion:* Intranasal application of NPs leads to the formation of SG in the olfactory bulbs. Apparently, the role of the dissolution of  $Mn_3O_4$ -NPs in the formation of SG in the mitral cells of the MOB is not significant.

*Acknowledgements:* The study was supported by Russian Science Foundation grant No. 20-14-00055.

### *References*

1. Romashchenko A.V. et al. The role of olfactory transport in the penetration of manganese oxide nanoparticles from blood into the brain. *Vavilov J Genet Breeding*. 2019;23(4):482-488. doi: 10.18699/VJ19.517.
2. Illarionova N.B. et al. 'Trojan-Horse' stress-granule formation mediated by manganese oxide nanoparticles. *Nanotoxicology*. 2020;14(10):1432-1444. doi: 10.1080/17435390.2020.1856433.

## Short day exposition affects the brain serotonin system in zebrafish *Danio rerio*

Sorokin I.E.\*, Evsiukova V.S., Kulikov A.V.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [sorokin.iv.ev@gmail.com](mailto:sorokin.iv.ev@gmail.com)

**Key words:** short day exposition, zebrafish, serotonin, brain

**Motivation and Aim:** Reduction of natural illumination in fall/winter months causes in vulnerable individuals seasonal affective disorders (SAD) – a significant medical, economic and social problem for countries located in high and temperate latitudes [1, 2]. Although some authors show the involvement of brain neurotransmitter, serotonin (5-HT), in the mechanism of SAD, the role of brain 5-HT system in the pathogenesis of SAD is still obscure. Existing models of SAD include either nocturnal animals (laboratory mice) or wild diurnal rodents with a poorly studied genome. This greatly complicates the study of the mechanisms of SAD. Diurnal animal zebrafish (*Danio rerio*) seems to be more suitable model organism for studying the mechanisms of SAD compared to nocturnal mice. The aim of the study is to compare the level of 5-HT, 5-HIAA, the activity of tryptophan hydroxylases, as well as the concentration of mRNA of the genes *Tph1a*, *Tph1b*, *Tph2*, *Mao*, *Slc6a4a*, *Htr1aa*, *Htr2aa* in the brain of male and female *D. rerio* kept at short (4:20) and normal (12:12) daylight hours.

**Methods and Algorithms:** Here we compare the level of 5-HT, 5-HIAA, the activity of tryptophan hydroxylases, as well as the concentration of mRNA of the genes *Tph1a*, *Tph1b*, *Tph2*, *Mao*, *Slc6a4a*, *Htr1aa*, *Htr2aa* in the brain of male and female *D. rerio* kept at short (4:20) and normal (12:12) daylight hours.

**Results:** The exposition to short day did not alter 5-HT level, but increased concentration its main metabolite, 5-hydroxyindoleacetic acid in the brain. The short day exposition decreased expression of *Mao* gene, encoding the main enzyme of 5-HT degradation, but did not alter the expression of *Tph1a*, *Tph1b*, *Tph2*, *Slc6a4a*, *Htr1aa*, *Htr2aa* genes, encoding the key enzymes of 5-HT synthesis, tryptophan hydroxylases, 5-HT transporter, 5-HT<sub>1A</sub> and 5-HT<sub>2A</sub> receptors, in zebrafish brain compared to those reared at normal day.

**Conclusion:** Thus, zebrafish are sensitive to short daytime and can be used as a promising model organism for study the involvement of 5-HT in SAD mechanism.

**Acknowledgements:** The study was supported by Russian Foundation for Basic Research (grant No. 20-34-90063).

### References

1. Levitan R.D. The chronobiology and neurobiology of winter seasonal affective disorder. *Dialogues Clin Neurosci.* 2007;9(3):315-324.
2. Miller A.L. Epidemiology, etiology, and natural treatment of seasonal affective disorder. *Altern Med Rev.* 2005;10(1):5-13.

## The impact of temperature conditions of incubation on mouse embryonic development during *in vitro* fertilization

Stanova A.K.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [aliya.stanova@mail.ru](mailto:aliya.stanova@mail.ru)

**Key words:** *in vitro* fertilization, mouse embryo, temperature influence, global DNA methylation

*Motivation and Aim:* Nowadays, *in vitro* fertilization (IVF) is used not only in IVF clinics, but also when replicating farm animals, and creating and preservation of genetic collections of laboratory animals. The effects of incubation conditions on the destabilization of individual development of embryo at the stage from fertilization to transplantation to surrogate mothers remain unclear. The aim of the work was to investigate the degree of destabilization of the pre-implantation development of mouse embryos depending on the incubation conditions.

*Methods and Algorithms:* The experiments were carried out on 12-14-week-old CD-1 mice of SPF-status on the basis of Centre for Collective Use “SPF-vivarium” of the Institute of Cytology and Genetics, SB RAS. After superovulation procedure, IVF was performed at 37 °C. Zygotes were collected and transferred to incubators with certain temperature: 35, 37, and 39 °C for 24 hours. A day later, only 2-cell embryo were selected and transferred to an incubator with a temperature of 37 °C, where further embryonic development took place up to the morula stage. The state of the embryos was fixed every 2 hours using an automatic microscope Lionheart FX. To analyze the global level of DNA methylation, embryos were stained with antibodies to 5mC (anti-5-methylcytosine) and propidium iodide.

*Results:* The passage of the first division under different temperature conditions is reflected on certain developmental parameters. Lowering the temperature affects the fusion of pronuclei and the first cell division, so they occur more slowly. 24-hours incubation at different temperatures also affects the rate of cell division. Embryos of 35 °C group divide more slowly up to the 4-cell stage. By the morula stage, the rate of cell divisions aligns in all groups. Blastomere size is affected by factors such as incubation temperature, cell numbers, and interplay of factors. Sizes increase from minimal at 35 °C to larger at 37 and 39 °C. The proportion of synchronously dividing embryos increases with the rise in the temperature of the first division. Synchronously dividing embryos are more likely to develop to a morula compared to asynchronously dividing embryos. Incubation temperature during the first 24 hours of development affects both methylation parameters: mean and fluctuation. At 39 °C, both mean methylation and inter-individual fluctuation in methylation are comparable to control. At 35 and 37 °C both of these parameters increase.

*Conclusion:* The results obtained in the experiments suggest that modification of the temperature conditions of the embryonic development during IVF as in *in vivo* fertilization, namely, incubation in the first 24 hours of development at 39 °C, followed by a decrease to 37 °C, shows that these embryos have a higher percentage of synchronous divisions, which provides them with a greater likelihood of successful development and a significantly lower level of methylation.

*Acknowledgements:* The study is supported by Russian Science Foundation grant No. 20-14-00055.

# Analysis of robust pattern formation in the developing *Drosophila* eye using two mathematical models

Surkova S.<sup>1\*</sup>, Kurova A.<sup>1</sup>, Gursky V.<sup>2</sup>, Samsonova M.<sup>1</sup>

<sup>1</sup> Peter the Great Saint Petersburg Polytechnic University, St. Petersburg, Russia

<sup>2</sup> Ioffe Institute, St. Petersburg, Russia

\* [surkova\\_syu@spbstu.ru](mailto:surkova_syu@spbstu.ru)

**Key words:** *Drosophila* eye disc, mathematical modeling, pattern formation, Scabrous, Notch-Delta

**Motivation and Aim:** *Drosophila* larval development is characterized by the formation of an ordered pattern of light sensitive units (ommatidia) in the presumptive eye structure, called eye imaginal disc. This pattern is initially determined by the precise spatial arrangement of R8 cells, future R8 photoreceptors. The mechanisms of formation and maintenance of the correct R8 cell arrangement, despite variability in expression of main regulators, remain elusive. Here, we use an *in silico* approach to study the regulatory mechanisms underlying the robust spatial pattern of R8 cells.

**Methods and Algorithms:** We implemented two mathematical models using the approach first introduced in [1, 2]. Previously developed mathematical models mostly include ‘generalized’ regulators that summarize the activating and repressive inputs of several gene products. Our approach provides incorporation of individual regulators into the model. The numerical simulations of cell differentiation process were visualized on the two-dimensional grid consisting of 2600 cells. To analyze the robustness of pattern formation, we randomly introduced noise into the model parameters.

**Results:** The spatial patterning of R8 cells comprises several steps, which are manifested in the expression of major proneural gene *atonal*. At the first step, the proneural clusters, including several R8 cells, are formed. Next, a single R8 cell is selected within each cluster to become the R8 photoreceptor. There are several concepts regarding the R8 cell selection, however, all of them involve a classic mechanism of lateral inhibition by the Notch-Delta pathway at a certain stage of pattern formation [3]. We studied the behavior of two models that differently implement the mechanism of local activation of *atonal* in the R8 cells and inhibition of its activity in the surrounding cells [1, 2]. The first model used the mechanism of local *atonal* activation by *scabrous* gene. In the second model, *scabrous* played a role of an indirect activator, acting through the inhibition of repression by the Notch-Delta pathway. We analyzed the response of the pattern structure to parameter perturbations within the two models. As a result, we characterized the features of stable pattern formation associated with each activation mechanism.

**Acknowledgements:** This study is supported by Russian Foundation for Basic Research (RFBR) grant 20-04-01047-a.

## References

1. Lubensky D.K. et al. A dynamical model of ommatidial crystal formation. *Proc Natl Acad Sci USA*. 2011;108(27):11145-11150.
2. Gavish A. et al. Periodic patterning of the *Drosophila* eye is stabilized by the diffusible activator Scabrous. *Nat Commun*. 2016;7:10461.
3. Frankfort B.J., Mardon G. R8 development in the *Drosophila* eye: a paradigm for neural selection and differentiation. *Development*. 2002;129(6):1295-1306.

## MRI neurophenotyping of the consequences of experimental hyperglycemia in different genotypes of mice

Tur D.<sup>1</sup>, Shevelev O.<sup>1,2</sup>, Khotskina A.<sup>1</sup>, Sharapova M.<sup>1</sup>, Akulov A.<sup>1,2\*</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *International Tomography Center, SB RAS, Novosibirsk, Russia*

\* [akulov@bionet.nsc.ru](mailto:akulov@bionet.nsc.ru)

**Key words:** MRI, hyperglycemia, neurophenotyping, mouse, genotype

*Motivation and Aim:* Neurophenotyping of laboratory animals in models of human pathology has biological and further medical significance. There is a question: how the choice of the genetic line of mice can affect the results of the study? We have focused on the pharmacological model of hyperglycemia, knowing that prolonged hyperglycemia affects brain function and this effect is quite strong. Thus, we have obtained a comparative study of the effect of prolonged hyperglycemia on the neurophenotype of different genetic lines of mice.

*Methods and Algorithms:* The neurophenotypes of 4 genetic lines of mice were studied in the work: CD1, C57BL/6, Cntn6, NODSCID. In animals of these genotypes, prolonged hyperglycemia was modeled by intraperitoneal administration of streptozotocin. The neurophenotype was assessed using MRI methods: classical MRI methods such as T2-weighted images and proton-weighted images, MRI angiography and MRI spectroscopy methods.

*Results:* When studying the morphological, circulatory and metabolic characteristics of the brain, the data indicate their significant interlinear variability. The greatest deviation in the neurophenotype when exposed to prolonged hyperglycemia was found in NODSCID mice, a smaller deviation in the CD1 line, and insignificant in the C57BL/6 and Cntn6 lines. At the same time, similar changes were found in all 4 lines: a decrease in brain size and an increase in the level of taurine in the brain. Also, neurophenotyping of the Cntn6 transgenic line created at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences using the CRISPR/Cas9 genetic technology based on the C57BL/6 line was performed for the first time.

*Conclusion:* In the course of the study, we found that the genotype of an animal can affect the level of neurophenotype changes during prolonged hyperglycemia. They also performed neurophenotyping using MRI methods in 4 lines of mice, in the Cntn6 line, for the first time.

*Acknowledgements:* The work was carried out within the framework of the project of the state task of the ICG No. FWNR-2022-0004.

## 7.2 Section “Population and evolutionary genomics/genetics of wild and domestic animals”



# Genome-wide association study of carcass traits in an Angus beef cattle Russian herd using imputed Whole-Genome Sequence data

Baneh H.<sup>1\*</sup>, Grigoreva E.<sup>1</sup>, Nizamutdinov I.<sup>2</sup>, Elatkin N.<sup>2</sup>, Lebedev I.<sup>2</sup>, Gentzbittel L.<sup>1</sup>

<sup>1</sup> *Project Center for Agro Technologies, Skolkovo Institute of Science and Technology, Moscow, Russia*

<sup>2</sup> *Miratorg-genetics LLC, Moscow, Russia*

\* *H.Baneh@skoltech.ru*

**Key words:** Genotype imputation, GWAS, carcass traits, beef cattle

The new genotyping technologies has provided a great opportunity for genomic studies and in particular identifying the genome regions associated with complex traits. Genotype imputation is a cheap and efficient tool to expand genotypes of low-density array to a much higher dense panel, by making use of whole genome sequence information. It appears that denser SNP panels obtained after imputation are more powerful and reliable in deciphering genetic background of the traits, in particular for Genome-wide association studies (GWAS). This study was carried out to optimize a corporate SNP chip array for Angus beef cattle via incorporating new significant SNPs affecting carcass weight (CW), rib-eye area (REA), marbling score (MS) and back fat thickness (FAT). As a first step, the 50K SNP chip genotypes of 338,824 Angus cows belonging to Miratorg Company were imputed to 9,351,538 variants (~173X increase in information), using whole genome sequences of 128 and 47 Angus bulls from the 1,000 genome project and Miratorg Company respectively, as reference population. Results indicated that population-based imputation method is more efficient compared to pedigree-based method and the Beagle software outperforms other programs. As a second step, the high accurately ( $DR2 > 0.8$ ) imputed SNPs ( $n = 6,544,904$ ) of 13,241 bulls having phenotype information were used for association studies. Genome-wide association studies were performed by applying Mixed Linear Model analysis, considering polygenic effect fitted as random part of the model and using a genomic relationship matrix. According to Bonferroni correction adjusted for the number of effective SNPs (threshold =  $3e-7$ ), 1,697 SNPs were found to be associated with CW (842 SNPs), REA (745), MS (340) and FAT (101). Among the signals, those located on BTA6 (for CW) and BTA19 (for MS) were not detected by 50K SNP genotypes. The identified signals were located within either the coding region or regulatory regions (upstream/downstream) of 47 coding protein, miRNA and siRNA genes. In addition, several significant SNPs were found in common between the traits, suggesting the genes with pleiotropic effects or functional polymorphic markers that affect more than one biological pathway. The potential candidate genes for the studied traits play a critical role in some biological process like apoptosis, lipid metabolism, immunity, ion transportations and regulation of translation and transcription. Our findings highlight the power of using imputation to perform GWAS, and provide some valuable information for better understanding of the underlying genetic background of carcass traits in the beef cattle.



## **Domestication explains two-thirds of differential gene expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation**

Chadaeva I., Ponomarenko P., Kozhemyakina R., Suslov V., Bogomolov A., Klimova N., Shikhevich S., Savinkova L., Oshchepkov D., Kolchanov N.A., Markel A., Ponomarenko M.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [ichadaeva@bionet.nsc.ru](mailto:ichadaeva@bionet.nsc.ru)

Before genomes were sequenced, zoologists had discovered destabilizing selection as a general pattern of animal domestication that in foxes and minks had yielded fur colors never seen in the wild. Today, known genomes of humans and domestic and wild animals arouse interest in a common whole-genome pattern of animal domestication that may at least explain differences between the effects of natural and artificial environments on organisms. Here we sequenced hypothalamic transcriptomes of tame and aggressive rats, identified their differentially expressed genes (DEGs), and, for the first time, applied principal component analysis to compare them with all the known DEGs of domestic versus wild animals that we could find. Two principal components, PC1 and PC2 respectively explained 67 and 33 % of differential-gene-expression variance (hereinafter: log<sub>2</sub> value) between domestic and wild animals. PC1 corresponded to multiple orthologous DEGs supported by homologs; these DEGs kept the log<sub>2</sub> value sign from species to species and from tissue to tissue (i.e., a common domestication pattern). PC2 represented stand-alone homologous DEG pairs reversing the log<sub>2</sub> value sign from one species to another and from tissue to tissue (i.e., representing intraspecific and interspecific variation). Accordingly, here we found that the artificial environment of domestic animals alters activities of genes in the same direction as that seen in corresponding human genes during human diseases, whereas the natural environment maintains a normal gene expression pattern in wild animals (matching human health).

*Acknowledgements:* This work was supported by Russian government project FWNR-2022-0020.

## Social behavior and spatial memory in tame and aggressive mammals

Herbeck Yu.E.<sup>1\*</sup>, Shepeleva D.V.<sup>1</sup>, Kozhemyakina R.V.<sup>1</sup>, Makovka Yu.V.<sup>1,2</sup>, Malyavko A.A.<sup>1</sup>, Klyuchereva A.A.<sup>1,2</sup>, Shikhevich S.G.<sup>1</sup>, Gulevich R.G.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [herbek@bionet.nsc.ru](mailto:herbek@bionet.nsc.ru)

**Key words:** domestication, behaviour, tame, aggressive, fox, rat

*Motivation and Aim:* Social behavior in mammals has a wide polymorphism from strictly social animals through facultative social to non-social animals. It is believed that a high sociality of their wild ancestors is one of the prerequisites for domestication. However, it should be noted that not all domesticated animals are subject to this rule; in addition, a destruction of the social structure of the group is observed in a number of cases (in particular, as seen in dogs) [1, 2]. The same applies to monogamous relationships of wild foxes that turned into promiscuity when bred in the captivity, and the selection of rats for tame behavior led to a deterioration in maternal behavior. However, it is likely that domestic and tame animals acquired the ability to socialize easier with a human showing an affiliative behavior towards him, which was the main hereditary change. However, the mechanisms of these changes are still poorly understood. Apparently, neurogenesis in the hippocampus in the juvenile period is one of the mechanisms of the formation of affiliative behavior [3]. It is known that neurogenesis in the hippocampus is higher in tame foxes as compared with foxes unselected for behavior (and, perhaps, in tame rats compared with aggressive ones) [4]. A high level of neurogenesis can be also associated with a high activity of oxytocin system, which, in addition, is associated with affiliative behavior in a number of cases. A change in the level of neurogenesis in the hippocampus can be also associated with spatial learning.

*Methods and Algorithms:* In this work, the expression of genes related to neurogenesis (particularly, to retinoic acid (RA) system playing an important role in initiating stress) was studied using qPCR in foxes and rats (models of domestication) in the juvenile and adult state. To change the activity of RA system and, accordingly, neurogenesis in the hippocampus (and possibly social behavior), the experiments were carried out with keeping rats and foxes on a diet with different amount of vitamin A in the juvenile period. In adulthood, tests for behavior towards conspecifics (three-chamber social test) and human (in the new enclosure) and a test for spatial learning (Barns test) were conducted. In addition, the behavior was tested after a nasal administration of oxytocin.

*Results:* It was demonstrated that tame rats and foxes exhibit the affiliative behavior to a greater extent than aggressive ones. At the same time, the expression of genes associated with neurogenesis is increased in tame animals. The administration of vitamin A in the juvenile period or oxytocin before the tests changes the behavior, reinforcing or, on the contrary, weakening the affiliative behavior. In addition, the administration of vitamin A improved the spatial learning in aggressive rats, apparently, through an increase in the levels of neurogenesis.

*Conclusion:* Vitamin A and oxytocin systems, associated with neurogenesis, stress, and other processes in CNS, seem to be involved in the formation of differences in social and cognitive behavior in the aggressive and tame animals (models of domestication). However, these effects are non-linear, a high dosage of vitamin A appears to have the opposite effect (increasing the aggression), while oxytocin can increase the anxiety and aggression under stressful conditions of the test.

*Acknowledgements:* The study is supported by the Russian Science Foundation (21-44-04405).

### *References*

1. Coppinger R., Coppinger L. *Dogs: A startling new understanding of canine origin, behavior, evolution.* Simon and Schuster, 2001.
2. Boitani L. et al. Behaviour and social ecology of free-ranging dogs. *Behav Biol Dogs.* 2007;9:147-165.
3. Wei L. et al. Affiliative behavior requires juvenile, but not adult neurogenesis. *J Neurosci.* 2011;31(40):14335-14345.
4. Huang S. et al. Selection for tameness, a key behavioral trait of domestication, increases adult hippocampal neurogenesis in foxes. *Hippocampus.* 2015;25(8):963-975.

# Imputation-based whole-genome analysis identifies candidate variants affecting body temperature under the cold stress in Siberian cattle

Igoshin A.V.<sup>1\*</sup>, Yudin N.S.<sup>1,2</sup>, Larkin D.M.<sup>3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Royal Veterinary College, University of London, London, UK*

\* [igoshin@bionet.nsc.ru](mailto:igoshin@bionet.nsc.ru)

**Key words:** genome-wide association study, imputation, cold adaptation

**Motivation and Aim:** Adaptability of cattle to environmental stresses has gained much interest over the last decades. Cold stress resilience, among others, represents a promising phenotype amenable to genetic improvement by means of selection or gene editing techniques. Previously, using SNP array genotyping, we identified *GRIA4* as a candidate gene associated with body temperature maintenance under cold stress in Siberian populations of Hereford and Kazakh Whiteheaded breeds [1]. Now, we performed imputation of the SNP array data to whole genome sequence level to identify candidate variants underlying this phenotype.

**Methods and Algorithms:** For our target SNP array set comprising 183 individuals, genomic coordinates of 119,841 SNPs were lifted from Btau8 to Btau9 assembly using the UCSC LiftOver tool. Then, we used autosomal variants of 141 Hereford animals from the 1000 Bull Genomes Project as an imputation reference panel. Low quality (GQ < 15) genotypes were removed and only biallelic SNPs with at least 80 % non-missing genotypes were retained. The target and reference sets were checked for an allele mismatch using *check\_strands.py* script and then pre-phased with Eagle v.2.4.1. The imputation was performed using Minimac4 software. After removing variants with low frequency and quality (maf < 0.025, rsq < 0.3) from the imputed dataset, we conducted genome-wide association analysis using EMMAX software. The area under the curve of body temperature over the period of five coldest days in February 2017 was taken as a phenotype. Sex and breed were taken as covariates.

**Results:** As a result, 769 SNPs, of which 597 found in *GRIA4* gene, reached a significance level (q-value < 0.05). The highest level of association ( $p = 1.03E-07$ ,  $q = 0.0045$ ) was observed for the variant BTA15:1,862,815 in the *MSANTD4-GRIA4* intergenic region. As the candidate functional variation, we consider three consecutive SNPs (BTA15:1,905,535-1,905,537;  $p = 1.10E-06$ ,  $q = 0.017$ ) located in the poly-T tract near the splice acceptor site of the *GRIA4* exon 8. In our imputed set, these SNPs are in a strong ( $r^2 = 1$ ) linkage disequilibrium and thus represent a trinucleotide substitution. Another promising (though reaching a suggestive level of association,  $q < 0.1$ ) candidate is BTA15:1,866,804 ( $p = 1.57e-05$ ,  $q = 0.074$ ) whose alternative allele frequency in the Russian Herefords is 4-fold higher than that of this breed's worldwide population (0.265 vs. 0.066) according to the 1000 Bull Genomes Project. Also, this variant can be found at moderate (0.136–0.238) frequencies in cold-tolerant Turano-Mongolian breeds (Yakut, Buryat, Yanbian) and at moderate-to-high frequencies (0.142–0.5) in several

heat-adapted indicine breeds (e.g. Red Sindhi, NDama, Ankole, Senepol, Nelore) implying its possible role in thermoregulation efficiency.

*Conclusion:* After an experimental genotype confirmation in carrier animals and validation of association in an independent sample set, the *GRIA4* variants highlighted in this study may be used in breeding practice.

*Acknowledgements:* This study was supported by the grant from the Russian Science Foundation (project No. 19-76-20026).

### *References*

1. Igoshin A.V., Yurchenko A.A., Belonogova N.M., Petrovsky D.V., Aitnazarov R.B., Soloshenko V.A., Yudin N.S., Larkin D.M. Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC Genetics*. 2019;20(Suppl 1):26.

## Phylogeography and local endemism of *Artemia salina* (Crustacea, Anostraca) in the hypersaline Lake Sasyk-Sivash (Crimea)

Meger Y.<sup>1\*</sup>, Shadrin N.<sup>2</sup>, Anufriieva E.<sup>2</sup>, Gadzhi A.<sup>1</sup>, Lantushenko A.<sup>1</sup>

<sup>1</sup> Sevastopol State University, Sevastopol, Russia

<sup>2</sup> A.O. Kovalevsky Institute of Biology of the Southern Seas, RAS, Sevastopol, Russia

\*meger\_yakov@mail.ru

**Key words:** *Artemia salina*, phylogeography, COI marker, population structure, mtDNA

**Abstract:** For the first time, the molecular genetics study and phylogenetic analysis of the bisexual *Artemia* population were carried out in the hypersaline lake Sasyk-Sivash (Crimea, Russia). A comparative analysis of the lake population and 20 different populations of *A. salina* studied earlier was carried out. The *A. salina* population from Lake Sasyk Sivash holds substantial genetic diversity, have a pronounced phylogeographic structure and high regional endemism.

**Motivation:** The brine shrimp genus *Artemia* (Crustacea, Anostraca) are widely used in aquaculture and biotechnology as a live feed for fish and shrimps larvae (cysts) and its biomass as source of fatty acids, antioxidants and other biologically active substances. The amount and type of secondary metabolites synthesized by *Artemia* strongly depends on the species genetic peculiarities. Previously, genetic screening of the genus *Artemia* in the hypersaline Sasyk-Sivash lake (Crimea) was not carried out. In this work, the *Artemia* specimens collected in the summer of 2021 has been studied.

**Methods:** 7 bisexual (4 male and 3 female) individuals were used. Total DNA was extracted from individual specimens using the DNA extraction kit (Syntol, Russia) after dissection for gut removal. Quantitative determination of the obtained genomic DNA and evaluation of its purity was carried out on an Inplen nanophotometer (Germany) and using gel electrophoresis in 1 % agarose gel. The PCR reaction was performed using pairs of primers jgLCOI490 and jgHCO2198 for the mitochondrial *COI* gene [1]. The PCR reaction was carried out in a volume of 25 µl using ScreenMix reagents (Eurogen, Russia). Sequencing of the obtained fragments was carried out on the NANOPHOR-05 sequencer (Syntol, Russia) at the CCU "Molecular Structure of Matter" of Sevastopol State University. The obtained *COI* gene sequences were compared with those available in the National Center for Biotechnological Information (NCBI) database. All previously studied populations of *A. salina* were used in the work [2, 3]. Phylogenetic reconstruction was performed using a Bayesian Inference approach implemented in MrBayes version 3.2.6 [4]. MrBayes analyses consisted of two simultaneous runs of 100 million generations each, sampling trees every 10,000 generations. Mixing and convergence among runs were evaluated by checking the average standard deviation of split frequencies, the EES values and the Potential Scale Reduction Factor (PSRF) for each parameter. A majority consensus tree was reconstructed after discarding the first 2,000 sampled trees as burn-in. The *Daphnia Tenebrosa* sequence (GenBank accession number HQ972028) was used as an external group. Relationships among mitochondrial haplotypes were reconstructed based on the TCS method implemented in the PopART [5].

**Results:** Haplotype diversity (HD), nucleotide diversity ( $\pi$ ), number of polymorphic sites (V), and number of nucleotide substitution (M) were calculated for each species using DnaSP 6.0 [6]. The analysis was carried out on 21 different populations of *Artemia Salina*, the total number of sequences used in the analysis was 115, the length of the sequences was 528 bp. No insertions, deletions or stop codons were present. The *COI* alignment consisted of 197 variable sites and 117 parsimony informative sites. The studied *A. salina* populations formed 72 haplotypes, the Sasyk-Sivash population forms 4 separate haplotypes. A comparative analysis of *A. salina* populations studied earlier and isolated in this work, based on estimates of evolutionary divergence over sequence pairs between groups, is carried out in MEGA X [7]. So, phylogeography and local endemism of *A. salina* from Sasyk Sivash Lake were compared with the results for *A. salina* in the Mediterranean basin [3], and its populations in Algeria and Tunisia, obtained by Eimanifar [2]. The Sasyk-Sivash population is most genetically close to CYP (Cyprus: Larnaca salt lake) and EBR (Spain: Tarragona) populations.

**Conclusion:** Thus, the *A. salina* population from Sasyk Sivash Lake holds substantial genetic diversity and, like the previously studied *A. salina* populations, have a strong phylogeographic structure and high regional endemism. It is important to note that the world's most widespread population of *Artemia* bisexual species *A. monica* (= *franciscana*) does not have a clearly defined phylogeographic structure.

**Acknowledgements:** This work is funded by the program 'Prioritet-2030' of Sevastopol State University (strategic project No. 3, No. 121121700318-1).

### References

1. Geller J., Meyer C., Parker M., Hawk H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol Ecol Resour*, 2013;13(5):851-861.
2. Eimanifar A. et al. *Artemia* biodiversity in Asia with the focus on the phylogeography of the introduced American species *Artemia franciscana* Kellogg, 1906. *Mol Phylogenet Evol*. 2014;79:392-403.
3. Munoz J. et al. Phylogeography and local endemism of the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda: Anostraca). *Mol Ecol*. 2008;17:3160-3177.
4. Ronquist F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539-542.
5. Leigh J.W., Bryant D. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110-1116.
6. Rozas J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299-3302.
7. Kumar S. et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547.

## The role of the *rhino* gene in the transcriptional regulation of different piRNA clusters

Milyaeva P.A.<sup>1</sup>, Kuzmin I.V.<sup>1</sup>, Lavrenov A.R.<sup>1,2</sup>, Dyachenko A.I.<sup>1</sup>, Nefedova L.N.<sup>1</sup>

<sup>1</sup> *Moscow State University, Faculty of Biology, Moscow, Russia*

<sup>2</sup> *A.N. Severtsov Institute of Ecology and Evolution, RAS, Moscow, Russia*

\* *atemed@mail.ru*

**Key words:** retrotransposon, piRNA silencing, mobile element

*Motivation and Aim:* piRNA silencing is one of major ways of retrotransposon activity suppression in eukaryotes. Nowadays this system is thoroughly explored in *Drosophila melanogaster* and in other model objects, including nematode and mouse. But regardless of description of this mechanism in details, many questions are still unsolved.

piRNAs can be processed from satellite retrotransposon transcripts or from extended accumulations of transposons known as piRNA clusters. In *D. melanogaster* these clusters are classified into two types: uni-strand (*20A* and *flamenco*) and dual-strand (*42AB* and *38C*). Most of them are expressed in the germ line, and transcripts are processed in the special area Nuage on the out surface of nucleus and interact with transposable elements' mRNAs. It is strongly believed that proteins coded by the *del* (*deadlock*), *rhino* (*rhi*) and *cutoff* (*cuff*) genes help RNA-polymerase to start transcription of a piRNA cluster from non-canonical promoters. Moreover, these proteins prevent such transcripts from splicing and it leads to balance of spliced and unspliced forms in cells.

In opposite to the most piRNA clusters, *flamenco* is considered to be expressed only in somatic ovary tissues *rhi*-independently. Its transcripts are generally alternatively spliced and are processed in Yb-body on the outer mitochondrial membrane. Subsequently such piRNA are used as guide-RNA by PIWI to find corresponding expressed transposons in the nucleus and start their repression. Earlier we have showed that *D. melanogaster* strain, demonstrating *flamenco* phenotype, also shows lower expression of *rhi* and increased amount of spliced transcripts in comparison with wild type strains.

*Methods and Algorithms:* We utilized *rhino* gene knockdown system (hybrids of *rhino* strain and tubulin driver strain) and hybrids of w1118 strain and tubulin driver strain as a control group. We evaluated relative expression of different retrotransposons, spliced and unspliced forms of *flamenco* and *42AB* clusters and various transcripts of *20A* and *38C* clusters in somatic and germline tissues of flies with quantitative PCR.

*Results:* We detected enhanced expression of several retrotransposons not only in germline but in somatic tissues. Moreover, these flies demonstrated increased amount of spliced transcripts of the *flamenco* and *42AB* clusters in ovaries and in carcass tissues as well.

*Conclusion:* These facts strongly support our hypothesis about regulation of *flamenco* cluster by *rhi* and significance of *42AB* cluster in regulation of the retrotransposon activity outside the gonads.



# The architecture of the large mitochondrial genome of a pelagic amphipod *Macrohectopus branickii*, and mitochondrial genome length variability in animals

Romanova E.V.<sup>1\*</sup>, Bukin Y.S.<sup>1</sup>, Mikhailov K.V.<sup>2,3</sup>, Logacheva M.D.<sup>2,4</sup>, Aleoshin V.V.<sup>2,3</sup>, Sherbakov D.Y.<sup>1,5</sup>

<sup>1</sup>Limnological Institute, SB RAS, Irkutsk, Russia

<sup>2</sup>Belozersky Institute for Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup>A.A. Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia

<sup>4</sup>Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>5</sup>Novosibirsk State University, Novosibirsk, Russia

\* elena\_romanova@lin.irk.ru

**Key words:** long mitochondrial genomes, non-coding regions, animal phyla, amphipods, Lake Baikal

**Motivation and Aim:** Mitochondrial genomes in animals have various lengths and architecture. Investigation of the different characteristics of mitochondrial genomes may help to understand the mechanisms of mitochondrial genome functioning and maintenance and their transformation in course of evolution. Recent studies of mitochondrial genomes in amphipods (Crustacean) from ancient Lake Baikal showed relatively high variance in length and gene content and order [1]. In this study, we sequenced the mitochondrial genome of a pelagic Baikalian amphipod *Macrohectopus branickii* [2]. We additionally investigated the variability patterns of mitochondrial genome length in amphipods and other animal taxa and assumed common mechanisms of their lengthening.

**Methods and Algorithms:** The mitochondrial genome of *M. branickii* was assembled from 41 million paired reads generated by Illumina HiSeq 4000 system from a DNA sample of a single specimen. The regions of long repeated sequences were additionally verified by PCR and Sanger sequencing. The statistical analyses of mitochondrial genome length characteristics were made using appropriate sequence data from the RefSeq database (released before 1 January 2020). We assessed the contribution of the coding and non-coding regions to the entire mitochondrial genome length using regression analysis. ANOVA was used to test the dependence of mitochondrial genome lengths on belonging to animal phyla. Additionally, we examined the repeat pattern in the lengthy mitochondrial genomes of different animal phyla.

**Results:** The complete mitochondrial genome of *M. branickii* has 42,256 bp in total. It has a unique gene order within amphipods, duplications of the four tRNA genes and a part of Cox2 gene, and a long non-coding region that makes up about two-thirds of the genome's size. Most of the non-coding region consists of inverted repeat sequences of different lengths. Statistical analysis showed no dependence of genome length characteristics on the number of sequences in animal phyla (sample size), but there is a significant dependence of genome lengths on the phylum itself.

**Conclusion:** *M. branickii* was shown to have the longest mitochondrial genome in amphipods so far and one of the longest within animals. We have shown that genome length distributions and the ratio of the longest mitochondrial genomes (upward outliers)

depend on the phylum and are not affected by sampling bias. The diversity in mitochondrial genome length in animal mitochondrial genomes is defined mainly by the lengths of their non-coding regions. A rare pattern of inverted repeats in the mitochondrial genome of *M. branickii* implies a more complex mechanism of the sequence lengthening than mere duplication and loss. The mitochondrial genome of *M. branickii* is a rare intermediate stage of mitochondrial genome evolution bearing signs of multiple duplications, that has not yet been deleted by selective pressure forces. *Acknowledgements*: This work was supported by the Ministry of Science and Higher Education of Russian Federation; project 0279-2021-0010 (121032300196-8).

### *References*

1. Romanova E.V. et al. Hidden cases of tRNA gene duplication and remodeling in mitochondrial genomes of amphipods. *Mol Phylogenet Evol.* 2020;144:106710.
2. Romanova E.V. et al. The mitochondrial genome of a freshwater pelagic amphipod *Macrohectopus branickii* is among the longest in Metazoa. *Genes.* 2021;12(12):2030.

## Search for convergently evolving and selected in parallel amino acids in domestic and wild animals

Romashov G.A.<sup>1,2\*</sup>, Yudin N.S.<sup>1,2</sup>, Yurchenko A.A.<sup>1</sup>, Larkin D.M.<sup>3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Royal Veterinary College, University of London, London, UK*

\* [romashov@bionet.nsc.ru](mailto:romashov@bionet.nsc.ru)

**Key words:** cattle, convergent evolution

*Motivation and Aim:* The study of convergent evolutionary events is a developing area of modern genetics and is of a great interest from both the theoretical and practical perspectives. The fundamental value of such a work is in obtaining a new knowledge about common genetic mechanisms of phenotype formation independently manifested in different species. Breeds of domesticated animals are a great model for studying convergent evolution of adaptation to environments because these populations have migrated with humans to new habitats and express adaptations often absent from the ancestral wild forms.

*Methods and Algorithms:* In the present work we looked for the nucleotide convergent changes utilizing the whole genome assembly and *de novo* annotated proteins of the cold-adapted Yakut cattle, proteins of 12 hibernating/cold-adapted/deep-diving mammals, and five species not expressing these phenotypes (outgroups). The species and phenotypic groups have been used in the analysis are: deep diving (*Odobenus divergens*, *Physeter catodon*, *Eumetopias jubatus*, *Mirounga leonine*), hibernating (*Myotis brandtii*, *Marmota marmota marmota*, *Microcebus murinus*, *Chrysochloris asiatica*), cold-resistant (*Vulpes lagopus*, *Rangifer tarandus*, *Ovis nivicola*, *Ursus maritimus*, *Odobenus divergens*) and outgroups (*Homo sapiens*, *Loxodonta Africana*, *Canis familiaris*, *Pteropus alecto*, *Bos taurus* (Hereford)). Coding sequences (CDSs) for all species have been downloaded from NCBI RefSeq or predicted with GeMoMa (v. 1.7). For each protein the longest isoform has been chosen. The protein sequences have been obtained with PRANK (<http://wasabiapp.org/software/prank/>). Orthogroups were identified with OrthoFinder (v. 2.5.2) and filtered to ensure each species had maximum one sequence in each orthogroup and each orthogroup contained  $\geq 3$  orthologous sequences for phenotypic group, all gaps in alignment of proteins sequences orthogroups were removed. Nucleotide sequences within each orthogroup were aligned with PRANK (v. 170703). PCOC method looking both at exact same amino acids changes and shifts in amino acid type preferences has been applied. For detection of the putative convergent changes between phenotypic groups the detection threshold has been set to 0.8. To identify convergent changes between the Yakut cattle and the ingroup species, all amino acids shared between the Yakut cattle and Hereford reference genome were filtered out.

*Results:* Eighteen putative convergent events were found in 18 genes (16 shared with the deep-diving and 2 with hibernating animals). This analysis confirmed the previously reported NRAP convergent mutation and identified putative events in 17 additional genes. The  $F_{ST}$  analysis for all 18 positions between the Yakut cattle and the dataset of

1000 Bull Genome Project suggests that all but one (*NRAP*) alleles were selected in parallel rather than occurred convergently in the Yakut cattle and the wild species because the Yakut cattle alleles were also present in other cattle breeds at a lower frequency and in *Bovinae* species. The functional analysis has revealed two genes previously found under selection in the Siberian human populations (*PLA2G4F* ( $F_{ST}=0.98$ ) and *PAFAH1B2* ( $F_{ST}=0.01$ )), a gene involved in resistance to environmental toxicants (*FNBP4* ( $F_{ST}=0.63$ )), genes involved in controlling the heart function (*NRAP* ( $F_{ST}=1.0$ ) and *POPDC2* ( $F_{ST}=0.3$ )), in regulation of blood pressure and insulin secretion (*SNX19* ( $F_{ST}=0.53$ )), and shortening of long bones (*WDR60* ( $F_{ST}=0.95$ )).

*Conclusion:* Our study demonstrates that while *de novo* convergent evolution events in breeds of domestic and wild species is a rare phenomenon (limited to *NRAP* in our study), in line with rare convergent changes between different species, a parallel selection for alternative ancestral alleles could be more widespread (17 genes). On the other hand, more ingroup and outgroup genomes should be used for a more powerful analysis. Both the convergently evolving and selected in parallel alleles should be considered when studying the biology of wild animals and livestock breeds and designing new generations of locally adapted livestock.

*Acknowledgements:* The study was supported by the Institute of Cytology and Genetics, SB RAS Budget Project No. FWNR-2022-0023.

## Correction of the aggressive behavior by immune cells modulated *ex vivo* by psychoactive drug

Serenko E.V.\*, Markova E.V.

Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia

\* [serenko.evgeniy@mail.ru](mailto:serenko.evgeniy@mail.ru)

**Key words:** aggressive behavior, immune cells, chlorpromazine, brain, cytokines

**Motivation and Aim:** Increasing evidence shows that there is an important relationship between aggression traits and the immune system, including immune cells function activity. Psychological stress can trigger cytokine release, and growing evidence has shown an important role for the immune system in regulating negative emotional states as well as personality. To produce deleterious behavioral effects in response to stress, peripheral cytokines must enter and act upon brain circuitry controlling mood and emotion. Within the CNS, activated microglial cells, astrocytes, neurons, and endothelial cells have all been shown to produce several cytokines and express many cytokine receptors. Thus, brain cytokines have important roles beyond inflammatory processes and can act as neuromodulators to regulate neuronal transmission and plasticity [5]. Earlier we demonstrated the influence of immune cells with certain functional phenotype on behavioral parameters and cytokine gene expression in the brain of experimental animals [4]. Findings from studies using animal models of social stress have uncovered numerous mechanisms of the immune cells functional impairment and immune-brain interactions that have opened up new avenues in novel drug targets for the treatment of psychiatric conditions with high aggression using syngeneic immune cells with modulated with psychoactive drugs functional activity. Objective of this study was to investigate the effect of the *in vitro* chlorpromazine-modulated immune cells on the behavioral patterns and brain cytokines in aggressive mice.

**Methods and Methods:** Adult male (CBA×C57Bl/6)F1 were tested in the open field; active animals were used for aggressive behavior generate with consecutive experience of 20- fold victories in daily agonistic confrontations with passive mice of approximately the same weight under the sensory contact model [2]. Aggressive mice immune cells were obtained in sterile conditions from splenocytes suspension, cultured *in vitro* with chlorpromazinee ( $15 \times 10^6$  cells/150  $\mu$ g drug) for 25 minutes, and intravenously administered to syngeneic aggressive recipients [3]. In the control group of recipients, the immune cells preparation and transplantation was carried out under similar experimental conditions, except that the cells were cultivated without the presence of chlorpromazine. The behavioral parameters (level of aggressive motivation, attacking behavior using Partition Test and behavior in the Open Field Test) were registered in the experimental and control recipient's group. In all aggressive recipients cytokines content in the pathogenetically significant for stress-induced forms of behavior brain structures (frontal cortex, striatum, hypothalamus, hippocampus) were determined by ELISA. Statistical data analysis was performed using an analytics software portfolio Statistica 10.0 for Windows (StatSoft, Tulsa, OK, USA). Mann–Whitney U test was applied;  $p \leq 0.05$  was considered statistically significant.

*Results:* The features of the aggressive mice immune cells functional activity after the treatment *in vitro* with chlorpromazine were described earlier [3]. It was found that transplantation of chlorpromazine-modulated immune cells change the recipient's behavior. In particular, total time (seconds) spent near the partition, when mice touch the partition with their fore parts (nose, paws) reacting to the control partner in the neighboring compartment was less ( $p < 0.05$ ); the number of approaches did not differ significantly in experimental and control recipient's group. After removal of the partition total time of attacks (seconds) and number of attacks registered in recipients after the transplantation of chlorpromazine-treated immune cells were significantly decrease ( $p < 0.01$ ). Transplantation of chlorpromazine-modulated splenocytes in aggressive mice-recipient was accompanied also by changed behavior parameters in the Open Field Test, manifested as decreased horizontal motor activity parameters ( $p < 0.01$ ), reflecting the behavioral motor component and vertical motor activity ( $p < 0.01$ ), reflecting the exploratory component of the behavior. As mentioned above, cytokines are involved in the central mechanisms of aggression. The aggressive behavior strategy formation correlates with changes in cytokine profile in some brain structures, such as the hypothalamus, hippocampus, striatum and frontal cortex; particularly increased levels of IL-1 $\beta$ , IL-2, and IL-6 has been shown in mice dominating in intermale collisions [1]. It was shown that the above-mentioned behavioral changes in aggressive recipients after the transplantation of chlorpromazine -modulated immune cells were accompanied by cytokines changes in the brain: in the hippocampus IL-1 $\beta$ , IL-2, IL-6, IFN- $\gamma$  levels were decreased ( $p < 0.01$ ); in the hypothalamus IL-4 level was increased ( $p < 0.01$ ) and IFN- $\gamma$  – decreased ( $p < 0.01$ ); in the frontal cortex IL-1 $\beta$  level was decreased ( $p < 0.01$ ). So, opposite changes in the brain cytokines during formation of behavioral aggressive strategy and its correction by chlorpromazine-modulated immune cells suppose its involved in the mechanisms underlying the development of aggressive reactions.

*Conclusion:* Thus, the present study demonstrated that immune cells with *in vitro* chlorpromazine-modulated function activity have a cytokine-mediated aggressive behavior corrective effect.

*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", theme No. 122011800324-4 (2021–2023).

## References

1. Idova G.V., Gevorgyan M.M., Alperina E.L., Zhukova E.N., Markova E.V. Changes in production of cytokines by C57BL/6J mouse spleen during aggression provoked by social stress. *Bull Exp Biol Med.* 2016;160(5):679-682.
2. Kudryavtseva N.N. The sensory contact model for the study of aggressive and submissive behaviors in male mice. *Aggr Behav.* 1991;17:285-291.
3. Markova E., Knyazheva M., Shushpanova T. Neuroleptic effect in aggressive mice after the transplantation of immune cells treated *in vitro* with chlorpromazine. *Eur Psychiatry.* 2016;33(1):263.
4. Markova E.V., Abramov V.V., Korotkova N.A., Kozlov V.A. Effect of transplantation of immunocompetent cell on orientation and exploratory behavior and cytokine gene expression in the brain of experimental animals. *Bull Exp Biol Med.* 2006;142(3):338-340.
5. Takahashi A., Flanigan M.E., Bruce S., McEwen B.S., Russo S.J. Aggression, Social Stress, and the Immune System in Humans and Animal Models. *Front Behav Neurosci.* 2018;12:56.

## Freezing-induced metabolic changes in the liver and hindlimb muscle of the moor frog *Rana arvalis* and the Siberian salamander *Salamandrella keyserlingii*

Shekhovtsov S.V.<sup>1,2\*</sup>, Bulakhova N.A.<sup>1</sup>, Tsentalovich Y.P.<sup>3</sup>, Zelentsova E.A.<sup>3</sup>, Meshcheryakova E.N.<sup>1</sup>, Poluboyarova T.V.<sup>2</sup>, Berman D.I.<sup>1</sup>

<sup>1</sup> *Institute of the Biological Problems of the North, FEB RAS, Magadan, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *International Tomography Center, SB RAS, Novosibirsk, Russia*

\* *shekhovtsov@bionet.nsc.ru*

**Key words:** freezing, metabolomics, *Rana arvalis*, *Salamandrella keyserlingii*

**Motivation and Aim:** Among amphibians, only a few species can survive prolonged (several weeks) freezing. Berman et al. [1] demonstrated that adult Siberian salamanders *Salamandrella keyserlingii* may survive prolonged freezing at up to 50 °C, while the moor frog *Rana arvalis* is capable of surviving freezing at –16 °C [2]. Little is known on the freezing response of these Eurasian species. We performed a metabolomics analysis of the liver and hindlimb muscles of the frozen and control Siberian salamanders and moor frogs in order (1) to quantify the major metabolites present in the tissues; (2) to determine the low molecular cryoprotectants used by these species; (3) to elucidate the changes in energy-related molecules; and (4) estimate the degree of cellular stress.

**Methods and Algorithms:** For each species, five adult animals were exposed to 7 days of freezing (–5 °C for *R. arvalis*, –8 °C for *S. keyserlingii*), and the control sample of the same size was kept at +1 °C. Liver and hindlimb muscles of the studied animals were extracted as quickly as possible, frozen in liquid nitrogen and used for metabolite extraction. Identification and quantification of metabolites was performed using 1H NMR and mass spectrometry analysis.

**Results:** We determined concentrations of approximately 60 substances in the studied organs. The only low molecular weight cryoprotectant in *S. keyserlingii* was glycerol (earlier found in [3]); unlike all other studied amphibians, the Siberian salamander does not use glucose as a cryoprotectant [4]. *R. arvalis* used both glycerol and glucose, which was unexpected, since the representatives of the genus *Rana* were believed to employ glucose only. We found high levels glycolysis products in both species, mostly lactate and alanine. Low amounts of ethanol and 2,3-butanediol were also detected; the latter is an interesting substance also detected in significant quantities *Rana amurensis* in response to anoxia [5]. The moor frog accumulated succinate, typical for ischemic conditions in vertebrates. However, this was not the case for the Siberian salamander, which suggests alternative strategies for ischemia adaptation. Freezing resulted in profound DNA damage, indicated by high concentration of nucleotide degradation products in both organs.

**Conclusion:** Our findings indicate that freeze-tolerant amphibians differ in the changes of many important metabolites, suggesting species-specific adaptation pathways.

**Acknowledgements:** Supported by the Russian Science Foundation (RSF) grant No. 21-74-20050.

## References

1. Berman D.I. et al. Extreme negative temperatures and body mass loss in the Siberian salamander (*Salamandrella keyserlingii*, Amphibia, Hynobiidae). *Dokl Biol Sci.* 2016;468:137-141.
2. Berman D.I. et al. Overwintering and cold tolerance in the moor frog (*Rana arvalis*) across its range. *Can J Zool.* 2020;98:705-714.
3. Berman D.I. et al. Winter hibernation of the Siberian salamander *Hynobius keyserlingi*. *J Evol Biochem Physiol.* 1984;3:323-327.
4. Shekhovtsov S.V. et al. Biochemical response to freezing in the Siberian salamander *Salamandrella keyserlingii*. *Biology.* 2021;10(11):1172.
5. Shekhovtsov S.V. et al. Metabolic response of the Siberian wood frog *Rana amurensis* to extreme hypoxia. *Sci Rep.* 2020;10:14604.



# Genetic structure analysis of 157 transboundary and local populations of cattle (*Bos taurus*, *Bos indicus* and *Bos grunniens*) based on STR markers

Solodneva E.V.<sup>1\*</sup>, Svishcheva G.R.<sup>1,2</sup>, Bazhenov S.A.<sup>1</sup>, Stolpovsky Y.A.<sup>1</sup>

<sup>1</sup> Vavilov Institute of General Genetics, RAS, Moscow, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* eugenia.575.2012@yandex.ru

**Key words:** STR analysis, cattle, local breeds, phylogenetic relationship

*Motivation and Aim:* According to the International Food Organization of the United Nations (FAO), biodiversity conservation is one of the most pressing problems of modern agriculture [1]. To develop breeding programs and programs for the conservation of rare breeds, methods of molecular genetic analysis, including those using STR markers, are widely used. In order to determine the population genetic characteristics, as well as to clarify the phylogenetic relationships of 157 cattle populations from different regions of the world, the analysis of the genetic variability of 12 highly variable microsatellite loci was carried out.

*Methods and Algorithms:* The study object was the original STR analysis data of 30 cattle populations provided by the laboratory team in the course of previous studies [2], as well as data obtained using microsatellite analysis of 127 populations, from open sources using the Mendeley Data database. Before combining this data into one array the data were reduced to a general form by eliminating nucleotide shifts calculated using heat maps. The genetic distance matrices were calculated using the Hierfstat and PopGenReport packages of the R environment. To reconstruct the phylogenetic tree using the poppr R-package and the Trex-online program the Neighbor-Joining approach was implemented. The visualization of the tree by the type of a circular dendrogram followed by the color identification of clusters was performed using the tools of the ITOL application. Allelic diversity parameters, observed and expected heterozygosity, fixation index, and Hardy-Weinberg equilibrium were calculated using the PopGenReport R-package.

*Results:* Using STR analysis data, 157 cattle populations from different geoclimatic zones were analyzed in order to study their genetic diversity and phylogenetic relationships. The main attention was paid to the analysis of poorly studied local breeds. Cluster analysis based on genetic distances, as well as plots of data display using principal components analysis, made it possible to distinguish three large groups: 1) yaks (*Bos grunniens*), 2) zebu (*Bos indicus*) and African cattle, and 3) domestic cattle (*Bos taurus*). The greatest phylogenetic distances from a common ancestor were noted for yaks, which correlates with the modern idea of the phylogeny of cattle. The introgression of zebu genes into the gene pool of African cattle has been noted in a number of works [3, 4], which explains their joint clustering. A high level of allelic diversity was identified among cattle breeds from Spain, Italy, Russia, Kyrgyzstan, Brazil, Colombia, Mexico, Mozambique, Egypt, Suriname and Kenya, as well as among Creole cattle. Analyzing the list of private alleles, we noted that all the identified unique allelic variants are

characteristic of local breeds of *Bos taurus* from different countries, populations of domestic Yak and Zebu.

*Conclusion:* The results obtained during the study can be used to improve the efficiency of breeding and conservation programs of local breeds in different countries.

*Acknowledgements:* The study is supported by the state task. State registration of the topic 122020800034-4.

### *References*

1. Bélanger J., Pilling D. (Eds.). The State of the World's Biodiversity for Food and Agriculture. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. 2019.
2. Svishcheva G. et al. Microsatellite diversity and phylogenetic relationships among East Eurasian *Bos taurus* breeds with an emphasis on rare and ancient local cattle. *Animals*. 2020;10(9):1493.
3. MacHugh D.E. et al. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics*. 1997;146(3):1071-1086.
4. Decker J.E. et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics*. 2014;10(3):e1004254.

## Regulatory effects of genes controlling behavior and painting shaping control american mink (*Neovison Vison*) as a model

Trapezov O.<sup>1,2\*</sup>, Nekrasova M.<sup>1,3</sup>, Stepanova M.<sup>1,3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State Agrarian University, Novosibirsk, Russia

\* [trapezov@bionet.nsc.ru](mailto:trapezov@bionet.nsc.ru)

**Key words:** Academician D.K. Belyaev, coat color patterns, domestication

*Motivation and Aim:* Evolutionary geneticist D.K. Belyaev considered the evolution of domestic animals as a process of great changes in the process of selection of regulatory genes, or transcription factor genes involved in the regulation of genetic activity at different stages of morphogenetic processes. In an analysis of the possible genetic nature of these regulatory changes, D.K. Belyaev assigned a key role to genes that control behavior. Genes that determine behavior are, as a rule, genes with multiple regulatory functions that contribute to the adaptation of animals to humans and to the anthropogenic environment as a whole. Selection that draws these genes into its sphere of action, D.K. Belyaev defined the term "destabilizing" [1–6].

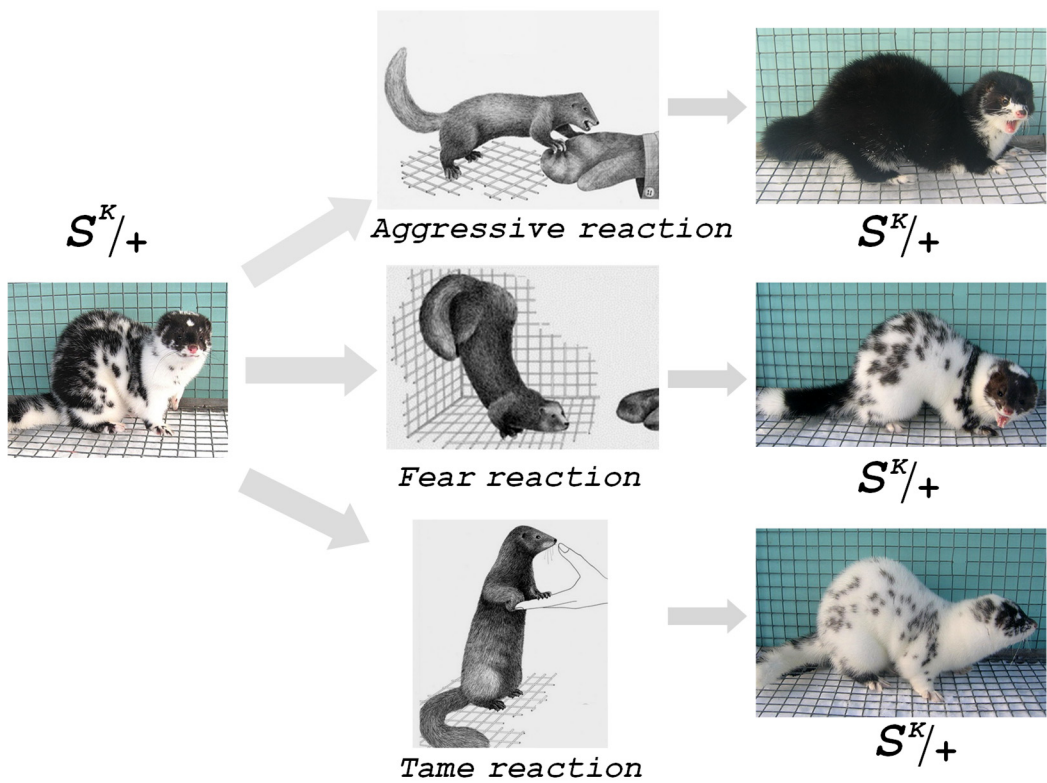
*Methods and Algorithms:* We used minks carrying the color mutation *Karel'skaja spotted* ( $S^k/+$ ) and *standard dark brown* mink ( $+/+$ ) from the aggressive and tame behavioral lines bred at the experimental fur farm of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Fig. 1).

*Conclusion:* One and the same object can be a carrier of one or the other information. Depending on which genotypic environment the signal enters, it can have either one or a completely different meaning.

*Acknowledgements:* The study is supported by Institute of Cytology and Genetics, SB RAS No. FWNR-2022-0023.

### References

1. Belyaev D.K. Domestication of animals. *Science*. 1969;5:47-52.
2. Belyaev D.K., Khvostova V.V. Domestication, Plant and Animal. In: Encyclopaedia Britannica. 1974;936-942.
3. Belyaev D.K. Destabilizing selection as a factor in domestication. *J Heredity*. 1979;70:301-308.
4. Belyaev D.K., Ruvinsky A.O., Trut L.N. Inherited activation-inactivation of the star in foxes. *J Heredity*. 1981;4:267-274.
5. Belyaev D.K., Trut L.N. Accelerating evolution. *Science USSR*. 1982;5:24-64.
6. Copping R., Copping L. Dogs. The University of Chicago Press, 2002.



**Fig. 1.** Changes in the pigmentation pattern in minks carrying the  $S^K/+$  mutation, depending on the genotypic environment created by the behavior genes

# Estimating telomere lengths of the local and imported to Russia cattle breeds from the whole-genome resequencing data

Yudin N.<sup>1\*</sup>, Romashov G.<sup>1</sup>, Larkin D.<sup>2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Royal Veterinary College, University of London, London, UK*

\* [yudin@bionet.nsc.ru](mailto:yudin@bionet.nsc.ru)

**Key words:** cattle, local breed, whole genome sequence, telomere length

*Motivation and Aim:* Health and survival are key elements in cattle selection. However, selection success is limited by the low heritability of these traits in addition to phenotype measurement difficulties. Telomere length (TL), measured in leukocytes, could be a good biomarker for breeding purposes since TL has been associated with health, ageing, and stress in humans and other species. However, we still do not know to what extent such associations in cattle would be determined by the genetics or/and the environmental factors. The TL is known to differ between human ethnic groups [1], inbred mouse strains [2], *C. elegans* strains [3] etc. The aim of the study was to investigate the differences of TL between the local and imported to Russia cattle breeds.

*Methods and Algorithms:* The average leukocyte telomere length (TL) from whole genome resequencing data of 59 animals of the Wagyu ( $n = 20$ ), Yaroslavl ( $n = 12$ ), Hereford ( $n = 12$ ), Yakut ( $n = 6$ ), Kholmogory ( $n = 6$ ), and Kalmyk ( $n = 3$ ) cattle breeds was estimated computationally using the TelSeq program (<https://github.com/zd1/telseq>).

*Results:* The TelSeq TL estimates in our study, ranging from 4.88 Kb to a maximum of 22.87 Kb with a median TL of 8.99 Kb, were similar to the TelSeq TL estimates for human [4] and *C. elegans* [3]. Measurements of TLs in younger animals were significantly higher in the whole dataset ( $r = -0.30$ ,  $p < 0.05$ ), and a significant difference was observed between breeds for the TLs. For example, TLs of the Wagyu bulls at the age of two years was significantly lower than in bulls of the Kholmogory ( $p = 0.0005$ ) and Yakut ( $p = 0.0007$ ) breeds of the same age.

*Conclusion:* Thus, a difference in the TLs between breeds might indicate that the genetic background affects this trait in cattle.

*Acknowledgements:* This work was funded by the RSF grant 22-26-00143.

## References

1. Ly K. et al. Telomere length in early childhood is associated with sex and ethnicity. *Sci Rep.* 2019;9(1):10359.
2. Hemann M.T., Greider C.W. Wild-derived inbred mouse strains have short telomeres. *Nucleic Acids Res.* 2000;28(22):4474-4478.
3. Cook D.E. et al. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics.* 2016;204(1):371-383.
4. Ding Z. et al. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* 2014;42(9):e75.

## Loci associated with negative heterosis for viability and meat productivity in interspecies sheep hybrids

Zlobin A.S.<sup>1\*</sup>, Volkova N.A.<sup>2</sup>, Zinovieva N.A.<sup>2</sup>, Iolchiev B.S.<sup>2</sup>, Bagirov V.A.<sup>2</sup>, Borodin P.M.<sup>3</sup>, Axenovich T.I.<sup>3</sup>, Tsepilov Y.A.<sup>1</sup>

<sup>1</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> L.K. Ernst Federal Science Center for Animal Husbandry, Dubrovitsy, Moscow Region, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* zlobin@bionet.nsc.ru

**Key words:** sheep, hybrids, heterosis, viability, meat productivity traits

*Motivation and Aim:* The modern market demands the animal breeds or breed types characterized by good viability, high growth rate and ability to effectively use feed. One of the way to improve the existing breeds is obtaining of breed hybrids. Hybrids are expected to have the increasing of the economically important traits because of mixing the genetic contributions of its parents. However, earlier for hybrids of romanovskaya sheep and argali (*Ovis ammon*) negative heterosis was shown for meat productivity traits and viability. In this work, we aimed to find genetic loci associated with negative heterosis in these hybrids.

*Methods and Algorithms:* In total we have studied 87 backcross of romanovskaya x argali. Viability information was available for four time-points after animal birth (6 days, 42 days, 3 months, 6 month). We used four indexes related to meat productivity (compactness, shortlegged, length and cumulative) measured at 42 days after animal birth. We used Cox regression for viability GWAS analysis and linear model for GWAS analysis of meat productivity traits. For each of the trait we used additive and three non-additive models (dominant, recessive, overdominant). As covariates, we used sex, number of fetuses, last time interval with available phenotype information, two first genetical principal components of kinship matrix and PRS (Polygenic risk score) for SNPs shown as associated with mortality.

*Results:* As a result, we discovered 3 loci associated with viability (rs422463993, rs415115903, rs417748866) and 19 loci associated with two indexes (compactness and shortlegged). For each locus, we manually checked the boxplots to infer the model of heritability. For four loci, we have demonstrated the effect of negative heterosis. Two of four loci (rs419262145 and rs399332377) explained 4.2 and 7.8 % variance of negative heterosis respectively.

*Conclusion:* The results can be useful for further association studies and preliminary estimation of genetic variability for economically important traits in different sheep breeds.

*Acknowledgements:* The study was supported by the Kurchatov Genomic Center of ICG SB RAS (075-15-2019-1662).

## 7.3 Section “Neurogenomics and genetics of behavior”



## Transcriptome analysis of *Trl* mutants reveals the role of mitochondria in the mass death of germ cells

Fedorova S.<sup>1,2\*</sup>, Dorogova N.<sup>1</sup>, Karagodin D.<sup>1</sup>, Baricheva E.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\**fsveta@bionet.nsc.ru*

*Drosophila* protein GAGA encoded by the *Trithorax-like (Trl)* gene is a factor of epigenetic transcription regulation of a large group of genes with a wide variety of cellular functions. In our previous works, we showed that this protein is necessary for the development of the reproductive system of *Drosophila* males. Decreased expression of the *Trl* gene in mutants induces autophagic death of spermatocytes. Previously we performed a detailed cytological analysis of *Trl* mutant's testes and revealed, in addition to autophagosomes and lysosomes, many abnormal mitochondria. We hypothesized that cell death in spermatogenesis may be associated with the mitochondrial pathologies. To test this hypothesis, we analyzed the RNA-seq data of testes from flies carrying a combination of the null allele *Trl<sup>R85</sup>* and the hypomorphic mutation *Trl<sup>362</sup>*. We revealed that *Trl* knockdown affects the expression of a large number of genes in the testes. The expression of about 2.5 thousand genes changes more than twofold. However, we did not identify anyone particular pathway whose activation could lead to germ cells death in the *Trl<sup>R85</sup> / Trl<sup>362</sup>* mutants. Whereas many genes related to mitochondrial structure and function change expression in GAGA deficient testes, 19 of them also affect cell death. Of the 778 genes belonging to the Flybase GO "mitochondrion" term (GO:0005739), 267 genes changed expression significantly: 123 increased expression, with 35 being upregulated twofold or more, and 144 downregulated, 13 of them downregulated 2 or more times. Of these, *Buffy*, *Mdh2*, *Rab32*, *park*, *CG14118*, and *Hsp60C* genes changed expression the most.

Thus, as a result of RNA-Seq data analysis, we concluded that the mass germ cells death is a consequence of mitochondrial abnormalities that may arise due to misregulation of GAGA target genes responsible for the structure and function of these organelles.

*Acknowledgements:* This study was supported by Projects No. 20-04-00496-a and No. 0259-2021-0011.



# Changes in the expression of lncRNAs RGD1564482 and RT1-CE6 in the adrenal glands may be associated with the development of a stress-sensitive form of hypertension in ISIAH rats

Fedoseeva L.A.\*, Markel A.L., Redina O.E.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\**fedoseeva@bionet.nsc.ru*

**Key words:** ISIAH rats, adrenal glands, lncRNA, DEGs

The functional role of the adrenal glands in the regulation of blood pressure and the development of arterial hypertension is well known. The ISIAH rat strain is a model of stress-sensitive arterial hypertension. Previously, we have shown that the manifestation of increased activity of the neuroendocrine hypothalamic-pituitary-adrenal (HPA) and sympathetic-adrenal systems (SAS) in these rats is accompanied by differences in the level of transcription of numerous genes, both in brain structures (brain stem, hypothalamus) and in peripheral target organs (adrenals, kidneys). The aim of this work was to identify lncRNAs that may be involved in the manifestation of the hypertensive phenotype in ISIAH rats. The study was performed using adrenal transcriptome sequencing data from 3-month-old hypertensive ISIAH and normotensive WAG rats. Three animals were analyzed in each group. The Cufflinks/Cuffdiff software packages were used to identify interstrain differences in transcription levels. Genes were considered differentially expressed at a false discovery rate (q value)  $< 5\%$ . Pearson's correlation coefficients were calculated to identify correlations between the level of lncRNA transcription and the level of transcription of genes differentially expressed in the adrenal glands of ISIAH and WAG rats. A 99 and 99.9% (df = 4;  $P < 0.01$  (0.917);  $P < 0.001$  (0.974)) two-tailed confidence intervals were recognized as significant for the identification of the lncRNA-DEGs coexpression. Functional analysis of DEGs whose expression correlates with lncRNA levels was performed in DAVID (Database for Annotation, Visualization and Integrated Discovery) gene annotation tool (<https://david.ncifcrf.gov/tools.jsp>). A total of 25 expressed lncRNAs were identified, five of which demonstrated differential levels of transcription in the adrenal glands of ISIAH and WAG rats. Correlation analysis showed that the expression of three of them correlated with the expression of a large number of DEGs: Aoc2-ps1 (244 DEGs), RGD1564482 (439 DEGs), RT1-CE6 (406 DEGs). Expression of these lncRNAs has not previously been studied in rat adrenal glands, and they have not been functionally characterized to date. Correlation analysis showed that the expression of lncRNAs RGD1564482 and RT1-CE6 correlates with each other ( $r = 0.985$ ) and with the expression of 10 DEGs associated with hypertension and making the most significant contribution to interstrain differences, which were previously identified when comparing gene transcription profiles in the adrenal glands of ISIAH and WAG rats [1]. The expression of lncRNAs RGD1564482 and RT1-CE6 correlates with the expression of DEGs that, according to the functional annotation, are associated with various biological processes, including neurogenesis, cell adhesion, inflammatory response, response to

oxidative stress, response to insulin, regulation of signaling. In addition, many terms associated with processes that contribute to the regulation of blood pressure levels and the development of hypertension have been identified – vasculature development, regulation of smooth muscle cell proliferation, regulation of calcium ion transport, regulation of catecholamine secretion, as well as blood circulation, blood coagulation, regulation of blood pressure. The results of our study suggest that the downregulation of lncRNAs RGD1564482 and RT1-CE6 in the adrenal glands of ISIAH rats may be associated with the development of a stress-sensitive form of hypertension in this strain of rats.

*Acknowledgements:* The work was supported by the Russian Science Foundation (project No. 22-14-00082).

### *References*

1. Fedoseeva L.A. et al. Molecular determinants of the adrenal gland functioning related to stress-sensitive hypertension in ISIAH rats. *BMC Genomics*. 2016;17(Suppl 14):989.

## Different immune challenges alter hypothalamic microglia and astrocytes activity in the adult BTBR and C57Bl/6 mice

Mutovina A.<sup>1,2\*</sup>, Ayriyants K.<sup>1</sup>, Ryabushkina Y.<sup>1</sup>, Reshetnikov V.<sup>1</sup>, Khantakova J.N.<sup>1</sup>, Bondar N.<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* [a.mutovina@g.nsu.ru](mailto:a.mutovina@g.nsu.ru)

**Key words:** BTBR T+Itr3tf/J, acute inflammation, LPS, Poly I:C, haematological parameters, lymphocyte immunophenotyping

*Aim:* Autism spectrum disorder (ASD) is a neurodevelopmental disease characterized by aberrant social interactions, deficits in social communication, and stereotyped repetitive behavior. In addition, an impaired immune response including a trend toward high cytokine production and atypical immune cell function is often observed in people with ASD. Therefore, the immune system may play a role in the development of this disorder. The BTBR T+Itr3tf/J (BTBR) mice model not only demonstrates an autism-like phenotype, but also has impairments in the immune system: they exhibit increased levels of serum antibody and enhanced cytokine expression, that can disrupt the development of the immune response in inflammation. However, the immune response to bacterial and viral infections is still poorly understood. Therefore, we explored differences in immune function in adult (60 days old) male BTBR and C57Bl/6 (B6) mice (used as control) after acute viral and bacterial inflammation.

*Methods:* BTBR and B6 mice received intraperitoneal injections of either lipopolysaccharide (LPS *Escherichia coli* serotype O55:B5, 50mcg/kg) in pyrogen-free saline or polyinosinic:polycytidylic acid (poly I:C, 10 mcg/kg) in pyrogen-free saline or a combination of them. Equivalent volumes of pyrogen-free saline were administered as control. Blood lymphocyte immunophenotyping and hematological response were conducted at 2- and 16-hours post-injection. In addition, levels of hypothalamic *c-Fos*, *Gfap* and *Aif1* transcripts were examined at 16-hours post-injection.

*Results:* In normal state BTBR mice had the same level of WBC and differential, 2-fold increase in T-cells, particularly CD4+ T-helper cells, and 3-fold decrease in B-cells and CD8+ T-killer cells in the peripheral blood compared to B6 mice. LPS, poly I:C and their combination induced physiological responses in both strains.

In BTBR mice LPS injection did not alter the number of cells in the complete blood count (CBC) with differential. Perhaps this is due to an insufficient dose of LPS. But, in B6 mice after LPS injection an increased number of granulocytes and decreased number of lymphocytes was indicated in CBC. Such changes are associated with a bacterial infection. The strongest response was shown after poly I:C administration in both B6 and BTBR mice. The CBC analysis showed that poly I:C administration induced leukopenia associated with sepsis in B6 mice and didn't change cell count in BTBR. However, in both strains CD8+ cytotoxic T-cells decrease (1.6-fold in BTBR vs 2.9-fold in B6) and CD4+ T-helper cells increase (1.15-fold in BTBR vs 1.3-fold in B6) indicates an acute phase of inflammation. Therefore, BTBR mice demonstrated impaired immune

response to bacterial and viral mimetic compared to B6 mice. Combination of LPS and poly I:C does not lead to increased immune response.

It has been shown that peripheral inflammation can change the activity of neurons, microglia and astrocytes in the hypothalamus, as it is the brain area that has been proposed to underlie the different behavioral and metabolic components of cytokine-induced sickness behavior. To determine hypothalamus neurons activation to different mimetics *c-Fos* was measured. Glial fibrillary acidic protein (GFAP) is a cell-specific intermediate filament protein in astrocytes and is widely used as an astrocyte marker. *Aif1* expression indicates microglial activation. In normal state in BTBR mice hypothalamic *c-Fos* is no different from B6. But *Gfap* expression was higher and *Aif1* expression was lower in the hypothalamus of BTBR mice compared to B6 mice. That indicates different base cell activity in the brain of BTBR mice.

Analysis of *c-Fos* expression 16-h after different mimetics administration showed no significant changes in both strains. But, *c-Fos* expression in the hypothalamus was significantly lower in BTBR mice compared to B6 mice. Peripheral inflammation did not change *Gfap* expression in astrocytes in both strains. *Aif1* expression was increased 16-hour after injection of a combination of mimetics in BTBR mice and after injection of LPS or poly I:C in B6 mice. Besides, BTBR mice, even after administration of LPS or poly I:C, retained a reduced level of microglia activation compared to B6.

*Conclusion:* Thus, poly I:C has been shown to have a more pronounced effect on the immune response and activity of hypothalamic cells than LPS. Co-administration of mimetics does not enhance the induced changes. In the BTBR strain, the injection of poly I:C causes a development of sickness behavior, with prevalence of CD4<sup>+</sup> T-helper cells in blood, reduced neuronal and microglial activity. Ultimately, BTBR mice have shown an impairment in immune response compared to B6 mice.

*Acknowledgements:* This study was supported by Russian Science Foundation 21-15-00142.

## Effects of amisulpride on the Tau-protein hyperphosphorilation in OXYS rats – a model of sporadic Alzheimer’s disease

Naumenko V.S.\*, Kondaurova E.M., Molobekova K.A., Rodnyy A.Ya., Stefanova N.A., Kolosova N.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [naumenko2002@mail.ru](mailto:naumenko2002@mail.ru)

**Key words:** Alzheimer’s disease, Tau pathology, brain serotonin system, 5-HT<sub>7</sub> receptor, amisulpride, rats

*Motivation and Aim:* Alzheimer’s disease (AD) is the most common cause of dementia worldwide with an increasing impact on the aging society. It is noteworthy that the percentage of people with Alzheimer’s dementia increases dramatically with age: 3 % of people age 65–74, 17 % of people age 75–84 and 32 % of people age 85 or older have Alzheimer’s dementia [1]. On the other hand, increasing life expectancy has been declared as one of the priorities of Russian domestic policy. Hence, development of effective therapeutic strategies against AD and AD-related disorders is among extremely topical problems of modern neuroscience and fundamental medicine.

Brain serotonin (5-HT) is a principal player in the mechanisms of brain and behavioral plasticity. Polyfunctionality of the brain 5-HT is due to a large variety of the receptors mediating 5-HT action on neurons. One of the most recently described 5-HT<sub>7</sub> receptor is involved in memory processes that require changes in synapse morphology and synaptic transmission efficiency. Stimulation of 5-HT<sub>7</sub> receptor in hippocampal neurons resulted in an increase in neurite length, dendritic projections, and synaptic density. One of the effectors responsible for this effect is CDK5. This is supported by the fact that CDK5 inhibitors blocked or reduced the density of dendritic spines, the number of which was increased upon 5-HT<sub>7</sub> receptor activation without the use of CDK5 inhibitors. It was also shown that these proteins physically interact at the plasma membrane. The fact that CDK5 plays an important role in tau hyperphosphorylation may indicate the involvement of 5-HT<sub>7</sub> receptor in tau phosphorylation status, its dissociation and further aggregation in AD. This idea was developed by the same authors who, using cellular and mouse models, showed that constitutive activation of 5-HT<sub>7</sub> receptor led, firstly, to increased CDK5 activity and, secondly, to excessive phosphorylation of the tau protein with further formation of insoluble aggregates. On the contrary, inhibition of 5-HT<sub>7</sub> receptor activity by antagonists led to the elimination of this effect [2].

To further confirm this idea, we investigated the effect of potent 5-HT<sub>7</sub> receptor antagonist on Tau-protein hyperphosphorilation and on the functional state of the brain 5-HT system in OXYS rats – a model of sporadic Alzheimer’s disease – an AD form that accounts for over 95 % of cases.

*Methods and Algorithms:* Experiments were carried out on 1-, 3-, and 6-month old OXYS rats. Animals were treated with 5-HT<sub>7</sub> receptor agonist amisulpride (3 mg/kg, i.p.) (Solian, Sanofi-Avensis France, France) during three weeks. This drug was selected since it is already applied in clinics for psychosis treatment. After three weeks of treatment “novel object recognition” test was performed. After behavioral testing, mRNA levels were assessed using quantitative real-time RT-PCR technique. Protein

levels were estimated by Western-Blot analysis. The data were analyzed using two-way ANOVA followed by Fisher's LSD post-hoc comparison.

*Results:* We showed that chronic amisulpride treatment significantly affected pathological Tau-protein phosphorylation in the brain of OXYS rats. It was revealed that amisulpride administration resulted in significant reduction of the pTau/Tau ratio in the hippocampus of 3-month-old rats. Moreover, in the hippocampus of 3-month-old animals amisulpride reduced expression of CDK5 – the main kinase involved in the pathological Tau phosphorylation. Nevertheless, observed amisulpride effect on the pathological phosphorylation of the Tau-protein and CDK5 failed to affect disrupted learning behavior assessed in the “novel object recognition” test.

Additionally, we revealed anxiolytic effect of amisulpride for 3-month-old rats that is not surprising since the investigated drug is currently utilized in clinics as antipsychotic. Moreover, we showed age-dependent decrease in 5-HT<sub>7</sub> receptor level in the frontal cortex and hippocampus of OXYS rats. However, taking together with the published data this reduction in 5-HT<sub>7</sub> receptor level seems to be related to age rather than Tau-pathology. Also we found the age-related increase in Tau and pTau levels in frontal cortex and hippocampus of OXYS rats that is undoubtedly linked with Tau-pathology in investigated animals.

*Conclusion:* Thus, amisulpride significantly reduced pTau/Tau ratio in the hippocampus of 3-months-old OXYS rats that was accompanied by reduced expression of CDK5 – the main kinase involved in the pathological Tau phosphorylation. However, investigated drug failed to restore disrupted learning in “novel object recognition” test. Obtained results allow to conclude that 3 month is the most sensitive age for Tau-pathology treatment with 5-HT<sub>7</sub> receptor antagonists in OXYS rats. Additionally, we showed the age-related increase in Tau and pTau levels in frontal cortex and hippocampus of OXYS rats.

*Acknowledgements:* The cost of animal maintenance was supported by the basic research project No. FWNR-2022-0023; the study was supported by the Russian Science Foundation (grant No. 22-15-00011).

### *References*

1. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement.* 2020. doi: 10.1002/alz.12068.
2. Labus J., Rohrs K.F., Ackmann J. et al. Amelioration of Tau pathology and memory deficits by targeting 5-HT<sub>7</sub> receptor. *Prog Neurobiol.* 2021;197:101900.

## Effects of the original synthetic ligand of benzodiazepine receptors in long-term alcoholized mice

Savkin I.V. \*, Markova E.V.

*Research Institute of Fundamental and Clinical Immunology, Laboratory of Neuroimmunology, Novosibirsk, Russia*

\* [i.v.savkin@ngs.ru](mailto:i.v.savkin@ngs.ru)

**Key words:** ortho-fluorobenzonal, experimental alcohol dependence, behavior, cytokines, immune response

*Motivation and Aim:* Currently, the role of the immune system in the pathogenesis of various mental disorders, in particular, chemical and behavioral addictions, is being actively studied. One of the molecular mechanisms underlying neuroimmune interactions, both in normal and pathological conditions, is the commonality of the receptor system of immunocytes and neurons with respect to such signaling molecules as cytokines and neurotransmitters. Of particular interest is the presence of gamma-aminobutyric acid (GABA) receptors type A, which are the main target of ethanol in the CNS, also on the surface of T-lymphocytes, which suggests the possibility of direct regulation of the immune response by inhibitory neurotransmitters. According to the modern paradigm, an essential pathogenetic mechanism of chronic alcoholism is neuroinflammation. Moreover, the cellular components of both non-adoptive and adoptive immunity are involved in the pathological process. Molecules of the GABA-receptor (GABA-R) complex, molecules of the Toll-like receptor family (TLRs), and cytokines play a key role in maintaining the pathological state. Ethanol consumption induces stable long-term activation of microglia predominantly in the hippocampus, mediated by TLRs [2]. Recently, an alternative pathway of TLR4 activation mediated by GABA-R has been identified, leading to a change in the balance of pro- and anti-inflammatory cytokine ligands on neurons, which is involved in the mechanisms of disturbance of higher nervous activity processes in alcoholism [1]. *ortho*-Fluorobenzonal is a synthetic ligand of the BDR-GABA-R complex, which, as we have previously established, have immunomodulatory properties mediated by lymphocytic GABA-R [3, 4]. The purpose of this work was to study the effect of the synthetic ligands of the GABA-R *ortho*-fluorobenzonal on the functional activity of the nervous and immune systems in experimental alcoholism.

*Materials and Methods:* Male (CBAxC57Bl/6)F1 mice in the state of alcohol dependence owing to 6-month 10 % ethanol exposure were undergoing intragastric administration of *ortho*-fluorobenzonal (100 mg/kg for 10 days). *ortho*-Fluorobenzonal was designed and synthesized in the Department of Biotechnology and Organic Chemistry, National Research Tomsk Polytechnic University. 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”. Animal's alcohol consumption, behavior and immune parameters before and after receiving the *ortho*-fluorobenzonal were estimated.

*Results:* Daily consumption of ethanol in mice with alcohol dependence decreased starting from 2 days of *ortho*-fluorobenzonal administration and stopping the

consumption of ethanol under free choice with water conditions on day 4, which indicated the alcohol motivation decrease. Chronic exposure to ethanol resulted also in significant decrease in animals behavioral activity in the Open Field Test. The pronounced changes in motor (horizontal locomotor) and exploratory (vertical locomotor) activities were registered after the 10 days substance administration: mice treated with ortho-fluorobenzonal showed increase in behavioral activity. Neuroimmune regulatory factors, including cytokines, contributed to changes in brain function and behavior associated with alcohol consumption. In particular, interaction of the key regulatory cytokine IL-1 $\beta$  with GABA $\alpha$ -R might account for the locomotory-depressive effects of this cytokine. ortho-Fluorobenzonal treatment modulated the level of IL-1 $\beta$ , as well as other cytokines (IL-6, TNF $\alpha$ , IFN- $\gamma$ , IL-10) in some brain regions (hippocampus, frontal cortex, striatum, hypothalamus) in mice with alcohol dependence, what might serve as one of the mechanisms of the demonstrated animal exploratory behavior stimulation. We have previously shown the ortho-fluorobenzonal ability of GABA-R – mediated modulation of spontaneous and mitogen-induced immune cells proliferative activity in mice with experimental alcohol dependence [3–5]. We also revealed the stimulation of the humoral and cellular immune response, estimated respectively by the relative number of antibody-producing spleen cells and the intensity of the delayed type hypersensitivity reaction. The immune response intensity in mice with experimental alcohol dependence after the ortho-fluorobenzonal treatment were comparable with that in the control group of equal age healthy animals.

*Conclusion:* The original synthetic ligand of benzodiazepine receptors ortho-fluorobenzonal, acting on molecular targets of ethanol in the nervous and immune systems, has a positive neuroimmunomodulatory effect, reduces alcohol motivation, so it has a promise in the treatment of alcoholism.

*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”, theme No. 122011800324-4 (2021–2023). The authors thank Tamara Shushpanova, PhD, for providing ortho-fluorobenzonal substance for the research.

### References

1. Aurelian L., Balan I. GABA A R  $\alpha$ 2-activated neuroimmune signal controls binge drinking and impulsivity through regulation of the CCL2/CX3CL1 balance. *Psychopharmacology*. 2019;236(10):3023-3043. doi: 10.1007/s00213-019-05220-4.
2. Lawrimore C.J., Coleman L.G., Zou J., Crews F.T. Ethanol Induction of Innate Immune Signals Across BV2 Microglia and SH-SY5Y Neuroblastoma Involves Induction of IL-4 and IL-13. *Brain Sci*. 2019;9(9):228. doi: 10.3390/brainsci9090228.
3. Markova E.V. Immunocompetent cells and regulation of behavioral reactions in normal and pathology conditions. Krasnoyarsk, 2021. doi: 10.12731/978-5-907208-67-4. (in Russian)
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., Knyazheva M., Shushpanova T. Original compound with anticonvulsant activity in the treatment of alcoholism. *Europ Psychiatry*. 2021;64(S1):S821-S822.



## Distribution of cell junction proteins in the descending colon of *Muc2* mice

Saydakova S.S.<sup>1,2,3\*</sup>, Ogienko A.A.<sup>1</sup>, Boldyreva L.V.<sup>1,2</sup>, Kozhevnikova E.N.<sup>1,2</sup>

<sup>1</sup> *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *custodian.of.midnight@gmail.com*

*Muc2* mouse strain is an experimental model of inflammatory bowel disease (IBD) lacking Mucin2, which is the prevalent component of intestinal mucus layer. *Muc2* mice exhibit the “leaky gut” syndrome characterized by intestinal barrier disruption, chronic inflammation, diarrhea, bloody stool and increased colorectal cancer risk. Using transmission electron microscopy, we have previously shown that tight junctions (TJ) and adherens junctions (AJ) are impaired in this IBD model. TJ represent the most apical selectively impermeable connections between enterocytes. AJ localize following TJ and serve for cell adhesion and tissue connectivity. Here, we describe the localization of several cell junction proteins in the colon of *Muc2* mice using fluorescent confocal microscopy. As for TJ marker proteins, we used claudin 3, claudin 7, zonula occludens 1 and JAM-A. E-cadherin and  $\beta$ -catenin were used to investigate AJ. Our findings demonstrate the dissociation of claudin 3 from the plasma membrane into cytoplasmic vesicles in the descending colon of mutant mice. Claudin 7 appeared to be less abundant at the apical part of lateral membrane of *Muc2* enterocytes compared to control ( $p$ -value  $< 0.05$ ,  $N = 6$ ). Zonula occludens 1, E-cadherin,  $\beta$ -catenin and JAM-A have showed no significant difference between *Muc2*<sup>+/+</sup> and *Muc2*<sup>-/-</sup> mice. Previous findings demonstrate that F-actin has impaired dynamics in *Muc2* mice leading to AJ’ and TJ’ disassembly. We conclude that the increase of intestinal permeability can be at least partially explained by the redistribution of the claudin family proteins possibly due to actin cytoskeleton instability.

*Acknowledgements:* This study was supported by the Russian Science Foundation (RSF) Grant No. 20-74-10022.

## **Risk factors for the development of post-stress neuroinflammation: studies on rat strains with contrast excitability of the nervous system**

Shalaginova I.G.<sup>1</sup>, Dyuzhikova N.A.<sup>2</sup>

<sup>1</sup> *Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

<sup>2</sup> *Pavlov Institute of Physiology, RAS, St. Petersburg, Russia*

\* *shalaginova\_i@mail.ru*

**Key words:** neuroinflammation, proinflammatory cytokines, stress, nervous system excitability, post-stress disorders, microglia

*Motivation and Aim:* The prevalence of disorders such as depression and anxiety is increasing worldwide and leads to disability of up to 18 % of people [1]. At the same time, the mechanisms of pathogenesis of these diseases remain unclear. Stress is considered one of the factors that, in combination with the biological vulnerability of the organism, can lead to the manifestation of psychopathologies. It is known that chronic stress leads to the development of neuroinflammation, which is now considered as a possible cause of cellular plasticity disturbances, which leads to dysfunctions in the nervous tissue.

The aim of this work was to evaluate the dynamics of post-stress peripheral and neuroinflammation in rat strains with contrast excitability of the nervous system.

*Methods and Algorithms:* Experiments were carried out on adult male rats of two strains with high (LT – low threshold) and low (HT – low threshold) levels of the peripheral and central nervous system excitability. The severity and dynamics of inflammatory response was evaluated by assessing the number of microglial cells in the brain (prefrontal cortex, hippocampus and amygdala), the shift of the leukocyte formula and the expression of proinflammatory cytokines in the blood and brain.

Experimental animals were subjected to psychogenic prolonged emotional and pain stress according to the protocol of K. Hecht: every day for 15 days the animals were exposed to 6 unsupported and 6 current-reinforced (2.5 mA, 2 ms) light signals. According to the scheme, combinations of conditional and unconditional stimuli were not repeated, but alternated with a probability of 0.5, which did not allow the animals to develop a conditioned reflex. Stressed animals and the corresponding control groups were decapitated at four time points: 1) 24 hours after exposure to stress; 2) 7 days after exposure to stress; 3) 24 days after exposure to stress and 4) 60 days after.

rtPCR was used to evaluate changes in the expression levels of IL1  $\beta$  and TNF $\alpha$  genes. The number of microglial cells was evaluated by immunohistochemical staining (Iba1+). The cells calculation was carried out using a fluorescent microscope Axio imager A2 (Carl Zeiss, Germany). The neutrophils /lymphocytes ratio (N:L ratio) was evaluated using a blood smear stained by Giemsa, and the phenotype of blood cells was determined in accordance with the recommendations for hematological evaluation in rats.

Distribution of data was evaluated by Shapiro-Wilk test using the GraphPad Prism software (v. 6, GraphPad Software Inc., La Jolla, CA). Non-parametric tests (Mann–

Whitney U tests, Fisher's exact test, Wilcoxon) were also performed with the GraphPad Prism software.

*Results:* In response to stress, a significant change in N:L ratio compared with the control animals was observed in stressed HT rats 24 hours after stress exposure. No significant differences were found between the LT control and stressed groups. We also found significant interstrain differences in the cytokine levels in the animal's blood already before exposure of these animals to any stress. A higher level of IL1  $\beta$  mRNA was found in the blood of highly excitable LT strain rats compared to low-excitable HT strain animals. However, we did not find any effect of stress on the expression of the studied cytokines in the blood of highly excitable LT animals at any time points. But we observed a significant increase in the expression of IL1  $\beta$  in the blood of HT strain compared to the control after 24 hours. Perhaps in this case we observe a normal pro-inflammatory reaction in the early stages of stress [2], which disappears by day 7.

Highly excitable intact LT rats showed significantly fewer Iba1<sup>+</sup> cells in the hippocampus compared to low excitable HT rats. In HT stressed rats we found a significant increase in microglial cells number only in the CA1 region 7 days after stress (compared to the HT control). In LT rats the number of microglial cells in all studied areas of the hippocampus was significantly increased 7 days after long-term stress and 24 days after stress exposure, the number of microglial cells did not differ from the control.

Analysis of the mRNA level of inflammatory cytokines in the brains of animals with varying levels of excitability showed regional heterogeneity of poststress changes. At 24 days after the stress stimulus, the level of IL1  $\beta$  mRNA increased in the stressed animals of both the LT and HT strains compared to the control groups of these strains. In the PFC, 24 hours after stress exposure there was a significant decrease in the level of IL1  $\beta$  mRNA in the experimental group of low-excitable HT rats compared to control animals. At 24 days after the stress, there was a significant increase in the level of IL1  $\beta$  mRNA in highly excitable animals (LT) compared to the control. In the amygdala of highly excitable animals, stress led to increased IL1  $\beta$  expression 24 days after the stress. A significant decrease in TNF $\alpha$  mRNA expression levels compared to the control observed 24 hours and 7 days after stress in animals of the HT strain (low-excitable).

*Conclusion:* We found that animals with high and low excitability have specific features in the severity and dynamics of post-stress peripheral and neuroinflammation. Thus, genetically determined excitability of the nervous system is a promising marker of individual vulnerability to stress, since it affects the severity and dynamics of post-stress neuroinflammation.

### *References*

1. Simpson H.B., Neria Y., Lewis-Fernández R., Schneier F. (Eds.). Anxiety Disorders: Theory, Research and Clinical Perspectives. Cambridge University Press, 2010.
2. Sorrells S.F., Sapolsky R.M. An inflammatory review of glucocorticoid actions in the CNS. *Brain Behav Immun.* 2007;21(3):259-72. doi: 10.1016/j.bbi.2006.11.006.

# Nanoparticle axonal transport assessment in neurodegeneration susceptible mice strain

Zuev D.<sup>1\*</sup>, Romashchenko A.<sup>2</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* zuevdaniil.zuevdaniil@gmail.com

**Key words:** nanoparticles, magnetic resonance imaging, axonal transport, neurodegeneration

**Abstract:** Neurodegenerative diseases have no treatment yet, due to their stealthy nature: to the moment when pathology can be diagnosed, the brain damage is already unrecoverable. The tight link between neuronal transport impairments and various forms of neurodegeneration has long been established. Noninvasive assessment of axonal transport can be valuable in studying and diagnosing such diseases. Paramagnetic nanoparticles are a new promising group of magnetic resonance contrasts, which may be a solution to this problem.

**Motivation and Aim:** It was proposed that neurodegenerative diseases, such as Parkinsons, are often caused by axonal transport disruption. Such disruption occurs long before the neurodegeneration starts [1, 2]. Therefore a noninvasive method to assess axonal transport rate could potentially allow for early diagnosis of such diseases. Earlier it was shown that Mn<sub>3</sub>O<sub>4</sub> nanoparticles (Mn<sub>3</sub>O<sub>4</sub>-NP) can act as an magnetic resonance imaging (MRI) contrast, that upon intranasal injection can be engulfed by olfactory neurons, and subsequently get transported into the deeper brain areas [3]. In this study we have assessed Mn<sub>3</sub>O<sub>4</sub>-NP spreading in brain of two congenic mice strains: *C57Bl* and *B6.Cg-Tg(Prnp-SNCA<sup>A53T</sup>)23Mkle* (with high  $\alpha$ -synucleinopathy risk, motor impairments become evident at the age of 12 months [4]).

**Methods and Algorithms:** 4 groups of 2 mice strains (*C57Bl* and *B6.Cg-Tg*), and 2 ages (8 and 12 week old) were used. Each group consisted of 6 males. Mn<sub>3</sub>O<sub>4</sub>-NP were prepared as previously described [3]. They were suspended in saline and intranasally instilled. Then 3, 6, 9, 12, 24, 48, 96 hours post injection mice head was MRI-scanned using Rapid Acquisition Relaxation Enhancement (RARE) pulse-sequence to obtain T1 weighted images. Acquired images were automatically parcellated based on manually marked keypoints. For each parcel maximum T1 signal value and maximum T1 signal change rate were calculated. Comparison between the groups was carried out using analysis of variance (ANOVA) with Fisher least significant difference (LSD) post-hoc analysis.

**Results:** Elevated T1 signal (proportional to Mn<sub>3</sub>O<sub>4</sub>-NP concentration) spreads primarily into olfaction-related brain structures: main olfactory bulbs (MOB), lateral olfactory tract, anterior olfactory nuclei, olfactory tubercle, piriform and piriform-amygdalar (AmPir) cortex, cortical-amygdalar area, entorhinal cortex. T1 signal peaks in these structures coincide with the structure position along the olfactory tract.

ANOVA with Fisher LSD showed significant differences for T1 maximum signal was found in AmPir: 12 week old *B6.Cg-Tg* mice had elevated T1 signal compared to other groups. T1 maximum change rate was significantly elevated in the MOB of 12 week old *C57Bl* mice compared to other groups.

**Conclusion:** It was shown that intranasally instilled Mn<sub>3</sub>O<sub>4</sub>-NP have a characteristic spreading pattern in mice brain, as they accumulate primarily in the olfactory structures. Differences observed in between 12 week old mice groups may be related to the  $\alpha$ -synuclein accumulation [4], that is believed to upregulate axonal transport in normal physiological quantities and impair transport in pathologically high quantities [5].

**Acknowledgements:** The study is supported by the Ministry of Education and Science of Russia (RFMEFI62117X0015).

### References

1. Angot E. et al. Alpha-synuclein cell-to-cell transfer and seeding in grafted dopaminergic neurons *in vivo*. *PLoS One*. 2012;7(6):e39465.
2. Braak H. et al. Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *J Neural Transm*. 2003;110(5):517-536.
3. Romashchenko A.V. et al. The role of olfactory transport in the penetration of manganese oxide nanoparticles from blood into the brain. *Vavilov J Genet Breeding*. 2019;23(4):482-488.
4. Lee M.K. et al. Human  $\alpha$ -synuclein-harboring familial Parkinson's disease-linked Ala-53  $\rightarrow$  Thr mutation causes neurodegenerative disease with  $\alpha$ -synuclein aggregation in transgenic mice. *PNAS USA*. 2002;99(13):8968-8973.
5. Toba S. et al. Alpha-synuclein facilitates to form short unconventional microtubules that have a unique function in the axonal transport. *Sci Rep*. 2017;7(1):16386.

## 7.4 Section “Neurogenomics of behavior”



## Therapy of Parkinson's disease-like deficits with rapamycin and trehalose in murine models with attenuated neuroinflammation

Akopyan A.A.\* , Pupyshev A.B., Dubrovina N.I., Tikhonova M.A.

*Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

\* [akopyanaa@physiol.ru](mailto:akopyanaa@physiol.ru)

**Key words:** neurodegeneration, cognitive deficits, neuroinflammation, microglia, brain, autophagy

*Motivation and Aim:* Weakening of the mechanism for cleaning neurons from damaged proteins and organelles is considered as one of the factors inducing neurodegeneration at Parkinson's disease (PD). Being the main catabolic mechanism of a cell, autophagy appears to be promising therapeutic target for PD treatment. In pharmacological modeling of PD, combined damage to nigrostriatal neurons accompanied by acute neuroinflammation leads to behavioral impairment. The inflammation largely subsides in a week after neurointoxication, however, the damage to nigrostriatal neurons persists [1]. On the other hand, the chronic course of PD in humans is not accompanied by severe neuroinflammation [2]. The present study was aimed at the possibilities of PD-like pathology in murine models with attenuated neuroinflammation by activating autophagy along the mTOR-dependent (rapamycin) and mTOR-independent (trehalose) pathways. *Methods and Algorithms:* Possible inhibition of neurodegeneration in a pharmacological PD model induced by MPTP (20 mg/kg i.p. daily, 4x) in C57Bl6/J mice and in a transgenic model of early PD stage (five-month-old B6.Cg-Tg(Prnp-SNCA\*A53T)23Mkle/J mice) through activation of autophagy by mTOR-dependent (rapamycin; 10 mg/kg i.p. every other day, 7x) and mTOR-independent (trehalose; 2 % solution in drinking water, 2 weeks) pathways was studied. We applied the autophagy-stimulating therapy to the MPTP-induced model of PD in a "postponed" mode (7 days after intoxication) and to a transgenic model at early stage of the disease progression (at the age of 5 months) when neuroinflammatory markers are even decreased in the brain [3]. Using immunohistochemical analysis of the brain cryosections, we evaluated damaged dopaminergic neurons by the expression of tyrosine hydroxylase (TH) and the expression of a marker of autophagy (LC3-II) in the striatum, substantia nigra, frontal cortex, and hippocampus. Recovery of cognitive function was assessed in the passive avoidance test.

*Results:* A significant reduction (more than two times less) of the autophagosome marker LC3-II indicating the decrease in autophagy activity was observed after 31 days in the substantia nigra in mice of MPTP-treated group vs. controls. Under the influence of the autophagy inducers, the immunofluorescence of the marker significantly augmented after all therapies. Similar pattern was revealed in the transgenic PD model. In the striatum, the decrease in autophagy was less pronounced, as was the activating effect of rapamycin and trehalose in both models. Therapy with autophagy inducers restored the expression of the dopaminergic neuronal marker TH both in the striatum and substantia nigra and significantly improved cognitive function of learning and memory in mice with PD-like pathology. An enhanced therapeutic effect of the combined treatment with the autophagy inducers was revealed on the expression of TH but not in the passive avoidance test.

*Conclusion:* Thus, in PD models with diminished neuroinflammation, activation of both pathways of autophagy regulation can restore the dopaminergic function of neurons and the cognitive activity in mice.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (theme No. 122042700001-9 (2021-2025)). The studies were partially implemented using the Unique scientific installation “Biological collection – Genetic biomodels of neuropsychiatric disorders” (No. 493387) at the Scientific Research Institute of Neurosciences and Medicine.

### *References*

1. Machado V. et al. Microglia-Mediated Neuroinflammation and Neurotrophic Factor-Induced Protection in the MPTP Mouse Model of Parkinson’s Disease-Lessons from Transgenic Mice. *Int J Mol Sci.* 2016;17(2):151. doi: 10.3390/ijms17020151.
2. Streit W.J. et al. Microglial pathology. *Acta Neuropathol Commun.* 2014;2:142. doi: 10.1186/s40478-014-0142-6.
3. Korolenko T.A. et al. Early Parkinson’s Disease-Like Pathology in a Transgenic Mouse Model Involves a Decreased Cst3 mRNA Expression But Not Neuroinflammatory Response in the Brain. *Med Univ.* 2020;3(2):66-78. doi: 10.2478/medu-2020-0008.



## Modeling of neuropathology on zebrafish: traumatic brain injury

Amstislavskaya T.G.<sup>1,2\*</sup>, Tikhonova M.A.<sup>1,2</sup>, Kalueff A.V.<sup>3</sup>, Chizhova N.D.<sup>1,2</sup>,  
Maslov N.A.<sup>4</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup>St. Petersburg University, St. Petersburg, Russia

<sup>4</sup>Khristianovich Institute of Theoretical and Applied Mechanics, SB RAS, Novosibirsk, Russia

\* amstislavskayatg@physiol.ru

**Key words:** neurodegeneration, neuroregeneration, animal models, motor activity, anxiety, behaviour, BDNF, HIF-1 $\alpha$ , brain

*Motivation and Aim:* Traumatic brain injury (TBI) consists of a combination of pathophysiological functional and morphological changes in the brain caused by external impacts. The pathogenesis of neurotrauma includes neuroinflammation, neurodegeneration, cerebral edema, and neuroregeneration. Research of the pathogenetic mechanisms of this neurological disorder because of the high level of hospitalization of patients with TBI on the one hand and a significant percentage of mortality due to the imperfect methods of TBI curation and treatment on the other. Modern trends in translational biomedicine prioritize the task of correct choice of a model organism. Its anatomical and physiological features, the correct translation of the results obtained from a model animal to humans, the molecular genetic features of the species as well as the bioethical aspects of *in vivo* experiments should be taking into account. In this regard, the zebrafish occurs as a promising model for studying the pathogenesis of neuropathologies including neurotrauma [1]. A large number of developed behavioral tests and techniques for working with its genome, high regenerative capacity of zebrafish, high translational potential of the results of molecular genetic and even behavioral studies along with relatively low maintenance expenses make this organism extremely relevant and promising for neuroscience research and create a competitive advantage in comparison with other model animals. In this study, we propose a novel model of non-penetrating brain injury in adult zebrafish induced by a laser irradiation.

*Methods and Algorithms:* An important advantage of laser radiation is the distance of exposure. Transparent tissues are not affected, while opaque, deeper ones would be sensitive to laser radiation. Zebrafish strain used in experiments had transparent skin on top of the head. Thus, visible laser radiation could penetrate it without beam distortion. It allowed to selectively heat brain surface to temperatures high enough to cause hyperthermia or protein denaturation. The experimental setup included a laser diode (power – 500 mW, wavelength – 405 nm) and aiming system consisted of video camera and 3D translation stage. The radiation of laser diode was focused into a spot of 0.1 mm on the surface of the telencephalon of anesthetized zebrafish; the duration of exposure was 2 or 4 s. Neuromorphological and molecular changes were assessed using standard histological staining and immunohistochemical analysis of the telencephalon cryosections. Behavioral phenotyping was performed in the Novel tank test to evaluate

motor disturbances and anxiety. Locomotor activity of the zebrafish was measured by the total distance traveled, freezing frequency and duration, mean velocity. Time spent in the top zone and latency to enter the top zone were regarded as indices of anxiety.

*Results:* Damage in the brain tissue was confirmed by neuromorphological examination of the samples using histological staining. As early as 30 min after the injury, protein expression of the transcription factor HIF-1 $\alpha$ , which is related to TBI-induced neurogenesis [2], increased in the fish telencephalon and persisted further until the Day 7 after the lesion. The expression of neurotrophic factor BDNF decreased 30 min after the injury with a subsequent elevation (the maximal increase was observed on the 3rd day after TBI. Both parameters reached control values on the Day 7. Locomotor activity in zebrafish was substantially reduced by laser-induced TBI in 30 min and on the next day after the injury while it had recovered since the Days 3–5. Laser-induced lesion of the dorsomedial telencephalon increased the anxiety of fish according to the decreased number of visits to the upper part of a tank on the day of injury (30 min after the trauma) and in one day after the injury. Moreover, TBI group significantly differed from Sham group in the latency to the first entry into the top zone on the Day 7.

*Conclusion:* Thus, the method of laser-induced injury is suggested as an adequate and useful tool for TBI research including the study of pathogenetic mechanisms related of TBI related to regeneration of an injury as well as preclinical screening of treatment approaches that will open new possibilities for treating patients with TBI.

*Acknowledgements:* The study was supported by Russian Science Foundation (grant No. 20-65-46006).

### *References*

1. Babchenko V.Ya. et al. Models of Traumatic Brain Injury in Zebrafish (*Danio rerio*). *Russ Fiziol Zh.* 2021;107(9):1059-1076. doi: 10.31857/S0869813921090028. (In Russian).
2. Lu K.T. et al. NKCC1 mediates traumatic brain injury-induced hippocampal neurogenesis through CREB phosphorylation and HIF-1 $\alpha$  expression. *Pflugers Arch.* 2015;467(8):1651-1661. doi: 10.1007/s00424-014-1588-x.

## Combined treatment with autophagy inducers rapamycin and trehalose as an experimental therapy for Alzheimer's disease-like pathology in mice

Bashirzade A.<sup>1,2\*</sup>, Pupyshev A.<sup>1</sup>, Belichenko V.<sup>1</sup>, Tenditnik M.<sup>1</sup>, Dubrovina N.<sup>1</sup>,  
Ovsyukova M.<sup>1</sup>, Akopyan A.<sup>1</sup>, Amstislavskaya T.<sup>1,2</sup>, Tikhonova M.<sup>1</sup>

<sup>1</sup> *Scientific Research Institute of Neuroscience and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *alim.bashirzade95@mail.ru*

**Key words:** rapamycin, trehalose, Alzheimer's disease, autophagy

Autophagy is considered to impede amyloid- $\beta$  (A $\beta$ ) accumulation in Alzheimer's disease (AD). Autophagy can be induced through mTOR-dependent and mTOR-independent pathways (here, by means of rapamycin and trehalose respectively). In order to evaluate the contribution of these pathways to AD treatment, we studied the effects of autophagy inducers rapamycin (10 mg/kg, intraperitoneally, every other day, 14 days), trehalose (2 % solution in drinking water, 14 days), and their combination in a murine model of AD. A $\beta$  fragment (25–35) was administered intracerebroventricularly in 2 days prior the start of treatment with autophagy inducers. The open-field, plus-maze, and passive avoidance tests were used for behavioral phenotyping. Neuronal density, amyloid accumulation, the expression of autophagy marker LC3-II and neuroinflammatory marker IBA1 were measured in the frontal cortex, hippocampus, and amygdala. mRNA levels of autophagy genes (Atg8, Becn1, Park2) were assessed in the hippocampus. Trehalose increased the autophagy gene expression and hence produced substantial and long-term autophagy induction. Both drugs efficiently prohibited microglia activation and A $\beta$  deposition. Rapamycin and trehalose significantly reversed behavioral and neuronal deficits in A $\beta$ -injected mice. Some parameters were better restored with the combined treatment. The results suggest that the trehalose alone or combined with rapamycin appeared to be a promising approach to therapy for AD.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (AAAA-A21-121011990039-2).

## Effects of two types of the central (intracerebroventricular vs. intrahippocampal) administration of A $\beta$ <sub>25-35</sub> on the induction of Alzheimer's disease-like pathology in mice

Belichenko V.M.<sup>1\*</sup>, Bashirzade A.A.<sup>1</sup>, Tenditnik M.V.<sup>1</sup>, Dubrovina N.I.<sup>1</sup>, Akopyan A.A.<sup>1</sup>, Ovsyukova M.V.<sup>1</sup>, Fedoseeva L.A.<sup>1,2</sup>, Amstislavskaya T.G.<sup>1</sup>, Tikhonova M.A.<sup>1</sup>

<sup>1</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *belichenkovm@neuronm.ru*

**Key words:** models of Alzheimer's disease, A $\beta$ <sub>25-35</sub>, mice, neuroinflammation, microglia, autophagy, neurodegeneration

*Motivation and Aim:* Identification of key pathogenetic mechanisms of Alzheimer's disease (AD) using adequate animal models that reproduce various signs of the pathology seems to be a priority area of research. Animal models of AD induced by intracerebroventricular (ICV) or intrahippocampal (IH) administration of oligomeric forms of amyloid-beta (A $\beta$ ) to laboratory rodents are widely used in current research [1]. It remains unclear whether these models provide similar outcomes or mimic core pathological mechanisms of AD equally. The aim of the work was to compare two models of AD induced by ICV or IH administration of A $\beta$ <sub>25-35</sub> oligomers (A $\beta$ <sub>25-35</sub>) to C57BL6 mice (3 m.o.). *Methods and Algorithms:* The analysis was performed using parameters characterizing behavior (passive avoidance test, open field test), protein content (IBA1, A $\beta$ , LC3-II), and expression of genes for neuroinflammation (*Aif1*, *Lcn2*, *Nrf2*), autophagy (*Atg8*, *Becn1*, *Park2*), and markers of neurodegeneration (*Cst3*, *Insr*, *Vegfa*). *Results:* Similar severity of cognitive deficits was found in both models of A $\beta$ <sub>25-35</sub> administration. The detected deposits of A $\beta$  were associated with the activation of microglia and neuroinflammatory response in the frontal cortex and hippocampus. Autophagy activity was increased in the hippocampus (LC3-II, ICV, IH) and frontal cortex (LC3-II, IH). The mRNA levels of genes related to neuroinflammation were elevated in the hippocampus (*Lcn2*, ICV) and amygdala (*Aif1*, ICV). Mitophagy activity (*Park2*) was suppressed in the cortex (IH) and increased in the amygdala (ICV). Neurodegenerative alterations were revealed by the reduced mRNA levels in the hippocampus (*Vegfa*, ICV, IH), frontal cortex (*Cst3*, ICV; *Insr*, IH; *Vegfa*, ICV), and amygdala (*Vegfa*, ICV). *Conclusion:* Despite similar effects on behavior, there were certain differences in the course of AD-like pathological processes induced by ICV or IH, which should be taken into account while modeling various aspects of AD.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (theme No. 122042700001-9 (2021-2025)). The studies were partially implemented using the Unique scientific installation "Biological collection – Genetic biomodels of neuropsychiatric disorders" (No. 493387) at the Scientific Research Institute of Neurosciences and Medicine.

### References

1. Wang L.S. et al. *Front Pharmacol.* 2019;10:1084. doi: 10.3389/fphar.2019.01084.

## Antipsychotic activity of a new PTPN5 blocker (TC-2051) in mice with the Disc1-L100P mutation

Chizhova N.<sup>1,2\*</sup>, Khomenko T.<sup>3</sup>, Volcho K.<sup>3</sup>, Amstislavskaya T.<sup>2,4</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup> N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

\* chnadezhda1995@gmail.com

**Key words:** schizophrenia, antipsychotics, benzopentathiepines, PTPN5, STEP, behavioural tests, Disc1-L100P mice

**Motivation and Aim:** Schizophrenia is a serious disabling mental disorder that, according to WHO, affects more than 20 million people worldwide [1]. The heterogeneity of the disease and frequent resistance to the therapy used often leads to ineffective or side-effect therapy and sometimes a combination of drugs may be the best choice – this also increases the number of adverse reactions. This presents a challenge for the search for new drugs in the treatment of schizophrenia [2, 3]. The main effect of AP is associated with a decrease in dopamine, however, the modern pharmacology trend is a point effect on the key links in the pathological process. Of interest is the STEP enzyme (striatal-enriched protein tyrosine phosphatase) encoded by the *Ptpn5* gene (*protein tyrosine phosphatase non-receptor type 5*). STEP has two isoforms, cytosolic STEP46 and membrane STEP61, which are highly expressed in the striatum, while only STEP61 is present in the hippocampus and cortex. This phosphatase, being expressed in dopaminergic neurons, is involved in signaling along the ERK pathway mediated by dopamine type 2 receptors (D2R), in the regulation of the development of dopaminergic neurons and, apparently, plays a role in the development of cognitive impairment [4,5]. The study of STEP blockers in models of schizophrenia as a putative AP seems to be relevant. We are screening new compounds capable of blocking STEP in Disc1-L100P mice (a genetic model of schizophrenia). The aim of this study was to evaluate the antipsychotic effects of the STEP blocker TC-2051 in a number of behavioral tests for manifestations of schizophrenia-like behavior in mice of the Disc1-L100P line.

**Methods and Algorithms:** The work was performed on 4-month-old mutant male mice of the Disc1-L100P line, and wild-type males (inbred line C57BL/6NCrI; WT). Mice were used from the National University "Biological Collection – Genetic Biomodels of Neuropsychiatric Diseases" (No. 493387) of the Research Institute of Neuroscience and Medicine [6]. Mice of the Disc1-L100P line carry a point mutation in the 2nd exon of the Disc1 gene. These mice are a model of schizophrenia and meet the criteria for models of mental disorder [7]. Testing was carried out between 9:00 and 16:00 in accordance with the PPI test protocols (using sound signals with loudness: prestimuli – 72, 78, 82, 86 dB; stimuli that cause startle response – 110 dB; background noise – 65 dB) [8], LI (sound signal loudness 80 dB, current strength 0.4 mA) [9] – the most commonly used to determine the schizophrenia-like phenotype. The effect of the drug on motor activity was assessed in the open field test [10]. The working conditions with the animals complied with international standards (Council of the European Communities Directive

86/609/EES). The preparation TC-2051, synthesized at the Novosibirsk Institute of Organic Chemistry, was intraperitoneally administered to mice at a dose of 10 mg/kg (which showed the most pronounced effect for the TC-2053 compound) 30 minutes before the start of testing. There were from 7 to 10 males in each group.

**Results:** The introduction of TC-2051 reduced locomotor activity in the open field in mice of both genotypes in comparison with the WT, which brought the value of Disc1-L100P closer to that of the wild type without the drug. Disc1-L100P mice showed a reduced PPI relative to WT and the introduction of TC-2051 slightly increased this indicator in them at prestimulus volumes of 82 and 86 dB. It is noteworthy that after administration of the drug, PPI in Disc1-L100P mice at 86 dB became indistinguishable from the values in the wild type. In LI test were found differences between the groups of Disc1-L100P mice with and without pre-exposure against the background of TC-2051 administration, which is typical for wild type mice.

**Conclusion:** The results obtained indicate the prospects for the use of STEP-blocking compounds of a number of benzopentathiepinines as new antipsychotics.

**Acknowledgements:** The study is supported by the integration project "An integrated approach to the creation of a new generation of antipsychotics" within the confines of the research work "Evaluation of the antipsychotic effects of STEP blockers in an experimental model of schizophrenia."

### References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1789-1858.
2. McCutcheon R.A. et al. The efficacy and heterogeneity of antipsychotic response in schizophrenia: A meta-analysis. *Mol Psychiatry*. 2019;26(4):1310-1320.
3. Potkin S.G. et al. The neurobiology of treatment-resistant schizophrenia: Paths to antipsychotic resistance and a roadmap for future research. *NPJ Schizophr*. 2020;6:1.
4. Kim S.Y. et al. Striatal-enriched protein tyrosine phosphatase regulates dopaminergic neuronal development via extracellular signal-regulated kinase signaling. *Exp Neurol*. 2008;214(1):69-77.
5. Fitzpatrick C.J., Lombroso P.J. The Role of Striatal-Enriched Protein Tyrosine Phosphatase (STEP) in Cognition. *Front Neuroanat*. 2011;5:47.
6. Petrova E. et al. Maintenance of genetically modified mouse lines: input to the development of bio-collections in Russia. *Lab Anim Sci*. 2018;2:2-15.
7. Lipina T.V., Roder J.C. Disrupted-In-Schizophrenia-1 (DISC1) interactome and mental disorders: impact of mouse models. *Neurosci Biobehav Rev*. 2014;45:271-294.
8. Braff D.L. et al. Sensorimotor gating and schizophrenia: human and animal model studies. *Arch General Psychiatry*. 2001;47:181-188.
9. Weiner I. The "two-headed" latent inhibition model of schizophrenia: modeling positive and negative symptoms and their treatment. *Psychopharmacology*. 2003;169(3-4):257-297.
10. Seibenhener M.L., Wooten M.C. Use of the Open Field Maze to Measure Locomotor and Anxiety-like Behavior in Mice. *JoVE*. 2015;96:e52434.

## The effect of diol on behavior in a MPTP-induced model of Parkinson's disease in mice

Gild E.-Y.V.<sup>1,2\*</sup>, Tenditnik M.V.<sup>1</sup>, Dubrovina N.I.<sup>1</sup>, Volcho K.P.<sup>3</sup>,  
Salakhutdinov N.F.<sup>3</sup>, Amstislavskaya T.G.<sup>1</sup>, Tikhonova M.A.<sup>1</sup>

<sup>1</sup>Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup>N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia

\*emmagild@gmail.com

**Key words:** neurodegeneration, animal model, monoterpenoid diol, locomotor activity, cognitive deficits

*Motivation and Aim:* According to recently published data, with the progression of the disease almost all patients with Parkinson's disease (PD) develop cognitive deficits. Moreover, cognitive impairment can occur earlier than motor symptoms of PD and includes impaired attention and working memory. Thus, experimental therapy for cognitive deficits in PD is area of high priority in current research. Recently synthesized monoterpenoid (1R,2R,6S)-3-methyl-6-(prop-1-en-2-yl)cyclohex-3-ene-1,2-diol 1 (Diol) alleviates motor manifestations of PD in animal models. This study was aimed at evaluating cognitive function using behavioral tests in mice with MPTP-induced PD-like disturbances treated with Diol.

*Methods and Algorithms:* C57BL6 mice were injected with a single dose of MPTP (30 mg/kg) or saline intraperitoneally; Diol was given orally one time (20 mg/kg) on the next day of the experiment; L-DOPA was given orally one time (100 mg/kg) on the same day as a control to Diol. Behavioral testing started 5 days after. Open field test and Rotarod were performed to assess general locomotor and exploratory activity. To evaluate cognitive function, the mice were tested using the passive avoidance learning paradigm; additional tests were made for 12 days to assess memory extinction and reconsolidation.

*Results:* In MPTP-treated mice, an increase in motor activity was observed, as well as a weakening of learning and accelerated extinction of the conditioned response of passive avoidance. Mice treated with Diol did not differ in locomotor activity from MPTP-treated mice, nor did they demonstrate signs of motor disturbances. Compared to mice treated with MPTP, those treated with MPTP and Diol showed a recovery in learning and enhanced memory reconsolidation. L-DOPA had similar to Diol positive effect on learning while the dynamics of memory extinction was similar to that in control group in mice treated with MPTP and L-DOPA. Thus, the beneficial effect of Diol on MPTP-induced cognitive deficits in mice has been demonstrated.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (theme No. 122042700001-9 (2021-2025)). The studies were partially implemented using the Unique scientific installation "Biological collection – Genetic biomodels of neuropsychiatric disorders" (No. 493387) at the Scientific Research Institute of Neurosciences and Medicine.

## Neurodegeneration and diabetes; positive effect of trehalose in behavior of db/db mice, model of diabetes

Korolenko T.A.<sup>1\*</sup>, Dubrovina N.I.<sup>1</sup>, Akopyan A.A.<sup>1</sup>, Pupyshev A.B.<sup>1</sup>, Zavyalov E.L.<sup>2</sup>, Tikhonova M.A.<sup>1</sup>, Amstislavskaya T.G.<sup>1</sup>

<sup>1</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [t.a.korolenko@physiol.ru](mailto:t.a.korolenko@physiol.ru)

Diabetes development is associated with neurodegeneration development, and new approaches for prevention and treatment of neurodegeneration development are important [1, 2]. Autophagy suppression was shown in aging, diabetes. Db/db mice is a model of diabetes, related to leptin receptor deficiency, was used in experiments [3, 4]. Our topic was to study the effect of an autophagy inducer, trehalose, on brain autophagy and behavior in a genetic model of diabetes with signs of neuronal damage (db/db mice). In experimental models of neurodegenerative diseases, the correction of autophagy in the brain reverses neuronal and behavioral deficits and hence seems to be a promising therapy for neuropathologies. Our aim – to evaluate effect of autophagy inducer trehalose on autophagic activity in brain of db/db mice, and behavioral changes. The study was carried out on a specific pathogen-free mice of db/db strain, 3 months old (SPF-vivarium of the Institute of Cytology and Genetics, SB RAS, Novosibirsk). Mice were subdivided into following groups: 1) WT mice + H<sub>2</sub>O; 2) WT mice drinking 2 % trehalose (24 days); 3) db/db mice + H<sub>2</sub>O; 4) db/db mice drinking 2 % trehalose (24 days). Expressions of markers of autophagy (LC3-II), neuroinflammation (IBA1), redox state (NOS), and neuronal density (NeuN) in the brain were assessed by immunohistochemical analysis. For behavioral phenotyping, the open field, elevated plus-maze, tail suspension, pre-pulse inhibition, and passive avoidance tests were used. It was shown that treatment by trehalose reduced weight of db/db mice, improved glycemic profile and behavioral characteristics of db/db mice. Immunohistochemical analysis revealed significant decrease of LC3-II expression in hippocampal areas in db/db mice compared to the control WT mice, while trehalose treatment significantly increased this index. We have shown positive effect of trehalose in liver and heart of db/db mice, revealing activation of lipophagy by trehalose. Trehalose can restore the suppressed autophagy induced by high glucose *in vivo* in db/db mice. Autophagy activation may be useful for treatment of diabetes, regulating the balance between apoptosis and autophagy by inhibiting mTOR signaling. Mechanism of protective effect of trehalose includes activation of hepatic transcription factor EB (TFEB), increasing TFEB nuclear translocation, elevating level of LC-3-II (marker of autophagosome) in mouse liver. These findings provide the basis for trehalose usage in pathology. Trehalose exerted some beneficial peripheral and systemic effects and partially reversed behavioral alterations in db/db mice. Thus, trehalose as an inducer of mTOR-independent autophagy is effective at alleviating neuronal and behavioral disturbances accompanying experimental diabetes.



## References

1. Burillo J., Marqués P., Jiménez B. et al. Insulin Resistance and Diabetes Mellitus in Alzheimer's Disease. *Cells*. 2021;10(5):1236. doi: 10.3390/cells10051236.
2. Labandeira C.M., Fraga-Bau A., Arias Ron D. et al. Parkinson's disease and diabetes mellitus: common mechanisms and treatment repurposing. *Neural Regen Res*. 2022;17(8):1652-1658. doi: 10.4103/1673-5374.332122.
3. Korolenko T.A., Dubrovina N.I., Ovsyukova M.V., Bgatova N.P., Tenditnik M.V., Pupyshev A.B., Akopyan A.A., Goncharova N.V., Lin C.L., Zavjalov E.L., Tikhonova M.A., Amstislavskaya T.G. Treatment with Autophagy Inducer Trehalose Alleviates Memory and Behavioral Impairments and Neuroinflammatory Brain Processes in db/db Mice. *Cells*. 2021;10(10):2557. doi: 10.3390/cells10102557.
4. Korolenko T.A., Ovsyukova M.V., Bgatova N.P., Ivanov I.D., Makarova S.I., Vavilin V.A., Popov A.V., Yuzhik E.I., Koldysheva E.V., Korolenko E.C., Zavjalov E.L., Amstislavskaya T.G. Trehalose Activates Hepatic and Myocardial Autophagy and Has Anti-Inflammatory Effects in db/db Diabetic Mice. *Life (Basel)*. 2022;12(3):442. doi: 10.3390/life12030442.

## Effects of chronic social stress on the expression of carcinogenesis-associated and apoptosis-associated genes in the hypothalamus of mice

Kovalenko I.L., Galyamina A.G., Smagin D.A., Kudryavtseva N.N.\*

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [matnik@bionet.nsc.ru](mailto:matnik@bionet.nsc.ru)

**Key words:** hypothalamus, transcriptome, carcinogenesis, apoptosis, chronic social conflicts

*Motivation and Aim:* Chronic social stress leads to a mixed anxiety/depression disorder [1] that is accompanied by psychogenic immunodeficiency and – as shown previously [2] – by the stimulation of oncogenic processes in mice. Undoubtedly, the hypothalamus is involved in the pathophysiology of psychoemotional disorders and the related immune pathologies and cancers. In this brain region of male mice with severe depressive symptoms, we aimed to identify differentially expressed genes (DEGs) encoding proteins related to mechanisms of carcinogenesis and apoptosis.

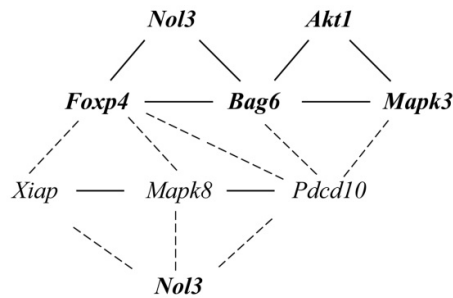
*Materials and Methods:* To induce a mixed anxiety/depression disorder in male mice, the model of chronic social conflict [1] was used, which causes a pronounced depressive state as a result of the chronic social defeat stress in animals under the influence of 20-day agonistic interactions. For comparison, male mice with positive fighting experience in the same model were analyzed. Hypothalamus transcriptomes of the animals with the opposite social experiences (aggression and defeat) were sequenced at the multi-access center Genoanalytica (<http://genoanalytica.ru/>, Moscow, Russia). For the analysis, the genes were selected that participate in the mechanisms underlying apoptosis and carcinogenesis according to databases (<http://deathbase.org> and <https://webs.iitd.edu.in/raghava/apocand/browse.php>).

*Results:* In the comparison of the aggressive with depressed animals, analysis of the RNA-Seq data revealed similar expression changes for numerous DEGs (related to carcinogenesis and apoptosis) as compared with the control group (no agonistic interactions). Almost all genes whose expression changed under the influence of positive fighting experience showed expression changes in the depressed mice (under the influence of chronic social stress). The number of DEGs and the magnitude of expression changes were lower in the aggressive than in the depressed mice. For instance, in the aggressive mice, there were 40 DEGs; among them, 27 genes (68 %) were upregulated and 13 genes (32 %) downregulated. In the depressed animals, there were 62 DEGs, of which 39 genes (63 %) were upregulated and 23 genes (37 %) downregulated.

The expression of genes *Casp2*, *Foxp2*, *Opal*, and *Tnfrsf22* changed in the aggressive mice but not in the depressed ones (encoded proteins are Caspase2, Forkhead Box p2, Mitochondrial Dynamin-like GTPase Tumor Necrosis Factor Receptor Superfamily Member 22 respectively). Compared to the aggressive animals, in the depressed mice, genes *Arc*, *Bax*, and *Nosip* (coding for regulatory molecules), *Alkbh7* and *Mapk3* (encoding protein enzymes), and *Foxp2* (coding for a transcription factor) showed higher expression, whereas genes *Casp2*, *Casp8*, and *Stk17b* (involved in the regulation of

apoptotic processes) and *Anxa1* and *Pycard* (encoding regulatory proteins) manifested lower expression.

Furthermore, there were 26 genes (*Aifm1*, *Ak2*, *Akt2*, *Anxa1*, *Bcl2l13*, *Bnip3*, *Casp8*, *Ccar1*, *Cldn5*, *Cldn12*, *Cyld*, *Dpf2*, *Fastkd2*, *Mcl1*, *Nol3*, *Nol4*, *Nomo1*, *Nos1*, *Nos3*, *Pdcd5*, *Ppargc1b*, *Pycard*, *Sfrp1*, *Siva1*, *Stk17b*, and *Tnfrsf8*) whose expression changed only in the depressed mice. In a correlation analysis of gene expression in the mouse groups (control, aggressive, and depressed animals), expression of eight genes (*Akt1*, *Bag6*, *Foxp4*, *Mapk3*, *Mapk8*, *Nol3*, *Pdcd10*, and *Xiap*) most strongly correlated with the expression of other genes ( $R > 0.950$ , Fig. 1).



**Fig. 1.** Correlations of functional activity among the genes contributing to the correlation among carcinogenic and apoptotic processes. "----": negative correlation; "—": positive correlation; the regular font indicates decreased expression, and the bold font denotes enhanced expression. *Nol3* is shown twice to visualize both positive and negative correlations

**Conclusion:** In male mice with opposite social experiences (aggression and defeats), there are similar changes in the hypothalamic mRNA expression of oncogenesis and apoptosis genes. It is possible that this pattern is a consequence of long-term stress exposure, albeit differing between the defeated and aggressive males. As demonstrated earlier [1, 2], in aggressive animals, subsequent metastasis induced by tumor cell injection is significantly less pronounced than metastasis in depressed animals. Therefore, it is likely that the genes – whose expression stayed unchanged in the aggressive males under the influence of the agonistic interactions – encode hypothalamic proteins that serve as tumor suppressors. Probably, it is genes *Akt1*, *Bag6*, *Foxp4*, *Mapk3*, *Mapk8*, *Nol3*, *Pdcd10*, and *Xiap* that mostly ensure the coordination of neurotranscriptomic changes in the hypothalamus, and the leading role apparently belongs to *Mapk3*, the expression of which differs between the mice with the opposite types of social experience. Further research on the participation of these genes and their products in the consequences of chronic social stress may be useful for the development of pharmacological therapies of psychosomatic pathologies.

**Acknowledgements:** The study is supported by the Russian Science Foundation (grant No. 19-15-00026-II) and by Budget Project No. FWNR-2022-0019.

### References

1. Kudryavtseva N.N., Shurlygina A.V., Galyamina A.G., Smagin D.A., Kovalenko I.L. et al. Immunopathology of mixed anxiety/depression disorders: An experimental approach to studies of immunodeficiency states (review). *Neurosci Behav Physiol.* 2019;49(3):384-398.
2. Kudryavtseva N.N., Tenditnik M.V., Nikolin V.P., Popova N.A., Kaledin V.I. The influence of psychoemotional status on metastasis of Lewis lung carcinoma and hepatocarcinoma-29 in mice of C57BL/6J and CBA/Lac strains. *Exp Oncol.* 2007;29(1):35-38.

## Psychoneuroimmunomodulating effect of immune cells in depressive-like mice

Markova E.V. \*, Knyazheva M.A.

Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia

\* [evgeniya\\_markova@mail.ru](mailto:evgeniya_markova@mail.ru)

**Key words:** depressive like mice, modulated immune cells, brain, cytokines, immune response

*Motivation and Aim:* Immune and neuroendocrine systems play an important role in maintaining homeostasis in normal and in mental maladaptation. Depression is a prime example of a pathology characterized by impaired neuroimmune interactions. There is a sufficient amount of data on the leading role of immune cells and their biologically active products in the pathogenesis of depression [1, 5]. Unidirectional influence of the majority of psychoactive drugs on the CNS and the immune system allows to consider immune cells as model objects for influencing the intersystem functional relationship. We demonstrated the possibility of animal's behavior directed regulation by the transplantation of immune cells with definite functional characteristics [3]. It was shown also the ability of splenocytes with *in vitro* caffeine modulated functional activity to stimulate passive behavior in animals in a dose dependent manner. Based on the previous results we investigated the effect of the transplantation of *in vitro* caffeine-modulated immune cells in depressive like recipients.

*Materials and Methods:* Experiments were held on 3 month old male (CBAxC57 Bl/6)F1 mice. Given individual behavioral differences, the groups were gathered after preliminary behavioral phenotyping to collect animals with uniform behavioral responses and stress reactivity. We performed the Open Field test to select mice with passive behavior, which are the least resistant to stress. Depressive-like state was formed in passive mice as a result of repeated experience of defeat in agonistic interactions with aggressive partner during 20 days (the sensory contact model) [2]. The depressive-like phenotype was characterized in the Forced swimming test, Open Field and Plus Maze using a modern hardware and software complex EthoVision XT (Noldus, Netherlands); anhedonia was assessed by the sucrose preference test using an automated system for behavioral and cognitive phenotyping IntelliCage (TSE systems, Germany). Splenocytes from depressive-like mice were received aseptically, treated *in vitro* with caffeine for 25 minutes, then, after a 3-fold washing, cells were collected and intravenously injected to syngeneic depressive-like recipients (15 million cells per animal). Mice-recipients of control group were undergoing the transplantation of immune cells treated without caffeine. Recipient's behavior, parameters of nervous and immune systems functional activities were estimated.

*Results:* Anhedonia is considered one of the main symptoms of major depression in humans. Anhedonia is also the main sign of depression in experimental models and is estimated by the decrease in animal's consumption of sucrose solution. In assessing the mice behavior in IntelliCage we found that depressive-like mice after the transplantation of caffeine-modulated immune cells in a free choice demonstrated increased consumption and preference of 1 % sucrose solution, as compared with the control group of animals. Behavioral despair in mice is a primary screening test for antidepressants.

Porsolt swimming test revealed in depressive-like mice after the transplantation of caffeine-modulated immune cells a significant increase in the time periods of mobility with the disappearance of periods of immobility in the water. We have shown also that transplantation of *in vitro* caffeine-modulated immune cells caused in depressive like recipients stimulation of motor and exploratory activities in the Open Field test [4].

In the scientific world widely discussed phenomenon of "cytokine induced depression" [1, 5]. We have shown that behavioral changes after the transplantation of the caffeine-treated immune cells were accompanied by decreased levels of several pathogenic significance pro-inflammatory cytokines TNF- $\alpha$ , IL-1 $\beta$ , IL-6, INF- $\gamma$  in some brain regions (hippocampus striatum, hypothalamus, frontal cortex and increase of IL-4I and L -10 in hippocampus of mice- recipients. The results suggested the reduction of neuroinflammation in depressive-like recipients. In depressive disorders is decreasing the volume of the hippocampus, due to a decrease in the density of pyramidal neurons in the CA 1 and CA 3 fields. Apoptosis of neurons in these zones is one of the pathogenetic mechanisms of cognitive impairment in depression [5]. We found also that after the transplantation of caffeine-modulated immune cells in the brain of depressive-like mice the number of neurons in the CA 1 CA 3 fields of the hippocampus substantially increases on the background of increasing level of brain-derived neurotrophic factor. Cell transplantation also led to modulation of the functional activity of recipient's immune system, expressed in stimulating the humoral immune response and the proliferative activity of splenocytes, modulating the production of some cytokines by splenocytes and reducing the catabolism of tryptophan in these cells.

*Conclusion:* The presented results indicate that modulated *ex vivo* with caffeine immune cells in depressive-like animals, affecting the main pathogenetic mechanisms of depression, have a positive psychoneuroimmunomodulating effect, which determines the possibility and prospects of immunotherapy of this disease by autologous immune cells with modulated *in vitro* functional activity.

*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", theme No. 122011800324-4 (2021–2023). The authors are grateful to T.G. Amstislavskaya and M.A. Tikhonova (SRINM) for scientific cooperation.

## References

1. Idova G.V., Markova E.V., Gevorgyan M.M., Alperina E.L., Zhanaeva S.Y. Cytokine production by splenic cells in C57Bl/6J mice with depression-like behaviour depends on the duration of social stress. *Bull Exp Biol Med.* 2018;164(5):645-649. doi: 10.1007/s10517-018-4050-9.
2. Kudryavtseva N.N. The sensory contact model for the study of aggressive and submissive behaviors in male mice. *Aggr Behav.* 1991;17:285-291.
3. Markova E.V., Abramov V.V., Korotkova N.A., Kozlov V.A. Effect of transplantation of immunocompetent cell on orientation and exploratory behavior and cytokine gene expression in the brain of experimental animals. *Bull Exp Biol Med.* 2006;142(3):338-340.
4. Markova E.V., Knyazheva M.A. Immune cells as a potential therapeutic agent in the treatment of depression. *Medical Immunology.* 2021;23(4):699-704. doi: 10.15789/1563-0625-ica-2277.
5. Remes O., Mendes J.F., Templeton P. Biological, Psychological, and Social Determinants of Depression: A Review of Recent Literature. *Brain Sci.* 2021;11:1633. doi: 10.3390/brainsci11121633.

## Novel approaches for analysis of the functional activity of neural networks *in vitro*

Mishchenko T.<sup>1</sup>, Krivonosov M.<sup>2</sup>, Mitroshina E.<sup>1</sup>, Kalyakulina A.<sup>2</sup>, Ivanchenko M.<sup>2</sup>, Vedunova M.<sup>1\*</sup>

<sup>1</sup>*Institute of Biology and Biomedicine, National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*

<sup>2</sup>*Institute of Information, Technology, Mathematics and Mechanics, National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*

\* *MVedunova@yandex.ru*

**Key words:** primary neuronal cultures, neural networks, network characteristics, spontaneous calcium activity, spontaneous bioelectrical activity, data analysis

*Motivation and Aim:* According to modern concept, neural networks are the minimal functional unit of the central nervous system responsible for the processing, storage and transmitting information [1, 2]. Currently, the studies on functional neural network activity are mainly implemented using multielectrode arrays (MEAs) and functional calcium imaging. These approaches allow to evaluate the effect of certain stressors and pharmacological agents on the functional reorganization of neural networks and to identify the contribution of such essential elements as astrocytes in the modulation of neural network plasticity. On the other hand, the continuing challenges in the biological signals processing result in the loss of a large amount of information that could have great fundamental and practical relevance.

*Methods and Algorithms:* Primary hippocampal cultures were obtained from C57Bl/6 murine embryos (E18) and cultured according to the previously developed protocol [3] on multielectrode arrays or coverslips. The initial density of cells was 7000–9000 cells/mm<sup>2</sup>. The primary hippocampal cultures were grown under constant conditions of 35.5°C, 5 % CO<sub>2</sub> and a humidified atmosphere in culture incubator for 28 days. Registration of spontaneous bioelectrical activity of primary hippocampal cultures was performed using a USB-MEA120-2-InVBC-System-E-Standard (Multichannel system, Germany) according to the previously developed protocol [2]. Electrophysiological data analysis was performed using original software which allows the detection of the main network characteristics: number of small and large network burst, number of spikes in a burst, duration of the network burst. The algorithm also allows the identification of the functional network structure by the cross-correlation method and graphs and the presentation of the network activation pattern. Registration of spontaneous calcium activity of primary hippocampal cultures was performed using an LSM 800 confocal laser scanning microscope (Carl Zeiss, Germany) and a fluorescent calcium-sensitive dye Oregon Green 488 BAPTA-1 AM (Invitrogen). Analysis of calcium events was based on previously developed algorithm [4]. A pairwise correlation analysis of fluctuations in the intracellular calcium level in the individual cells was then carried out. A functional network is presented as a non-oriented graph with nodes corresponding to the cells, and the edges designating a significant correlation between the levels of calcium activity in cells exceeding the threshold value ( $p > 0.3$ ) [5, 6].

*Results:* The main regularities of reorganization of neural network activity were revealed using the novel algorithms in a model of acute hypoxia *in vitro* taken here as an example. On the one hand, the algorithms open a possibility to reveal the role of individual cellular elements in the neural network activity; on the other hand, they allow to show general regularities of formation and preservation of functionally active network centers (hubs) under conditions of the network reorganization. Our approach is also resolved the problem of detection of calcium oscillations and allows to reveal the level of correlation between both neighboring and distant cells as well as the signal speed propagation. The use of novel algorithms allows us to study in detail the features of the action of a number of substances with neuroprotective properties, such as neurotrophic factors BDNF, GDNF, HIF prolyl hydroxylase inhibitors, vitamin D3, on the functional architectonics of neural networks in normal state and hypoxic damage. The data obtained undoubtedly expand our knowledge about the CNS functioning as well as can serve as a useful tool for assessing the risk of side effects that contributes to the development of new therapeutic strategies for different pathologies of the CNS.

*Conclusion:* The use of novel algorithms of analyzing the data of spontaneous bioelectrical and calcium activity of neural networks opens the possibility of identifying the main mechanisms of action of neuroprotectors as well as conducting high-precision screening for selection the most effective methods of neurodegenerative processes correction and prediction the possible side effects of new therapeutic drugs and approaches.

*Acknowledgements:* The study was supported by Ministry of Science and Higher Education of the Russian Federation, agreement No. 075-15-2020-808.

### References

1. Yuste R. From the neuron doctrine to neural networks. *Nat Rev Neurosci.* 2015;16(8):487-497. doi: 10.1038/nrn3962.
2. Mishchenko T.A., Mitroshina E.V., Usenko A.V., Voronova N.V., Astrakhanova T.A., Shirokova O.M. et al. Features of neural network formation and their functions in primary hippocampal cultures on the background of chronic TrkB receptor system influence. *Front Physiol.* 2019;9:1925.
3. Vedunova M., Sakharnova T., Mitroshina E., Perminova M., Pimashkin A., Zakharov Yu. et al. Seizure-like activity in hyaluronidase-treated dissociated hippocampal cultures. *Front Cell Neurosci.* 2013;7:149.
4. Kustikova V., Krivonosov M., Pimashkin A., Denisov P., Zaikin A., Ivanchenko M., Meyerov I., Semyanov A. CalciumCV: Computer vision software for calcium signaling in astrocytes. In: International Conference on Analysis of Images, Social Networks and Texts. Springer, Cham. 2018. doi: 10.1007/978-3-030-11027-7\_17.
5. Savyuk M., Krivonosov M., Mishchenko T., Gazaryan I., Ivanchenko M., Khristichenko A., Poloznikov A., Hushpilian D., Nikulin S., Tonevitsky E., Abuzarova G., Mitroshina E., Vedunova M. Neuroprotective effect of HIF prolyl hydroxylase inhibition in an *in vitro* hypoxia model. *Antioxidants (Basel).* 2020;9(8):662. doi: 10.3390/antiox9080662.
6. Loginova M., Mishchenko T., Savyuk M., Guseva S., Gavrish M., Krivonosov M., Ivanchenko M., Fedotova J., Vedunova M. Double-Edged Sword of Vitamin D3 Effects on Primary Neuronal Cultures in Hypoxic States. *Int J Mol Sci.* 2021;22(11):5417. doi: 10.3390/ijms22115417.

## Oral trehalose intake as an experimental therapy for Alzheimer's disease in a mouse model induced by the central administration of amyloid- $\beta$

Pupyshev A.B.<sup>1\*</sup>, Akopyan A.A.<sup>1</sup>, Tenditnik M.V.<sup>1</sup>, Ovsyukova M.V.<sup>1</sup>,  
Dubrovina N.I.<sup>1</sup>, Belichenko V.M.<sup>1</sup>, Zozulya S.A.<sup>2</sup>, Klyushnik T.P.<sup>2</sup>, Tikhonova M.A.<sup>1</sup>

*Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Mental Health Research Center, Moscow, Russia*

\* *pupyshevab@physiol.ru*

**Key words:** Alzheimer's disease, amyloid-beta, trehalose, autophagy

**Motivation and Aim:** Disaccharide trehalose with flexible  $\alpha$ -1-1' glycoside bond (O- $\alpha$ -D-glucopyranosyl-[1 $\rightarrow$ 1]- $\alpha$ -D-glucopyranoside) has the ability to inhibit neurodegeneration via several mechanisms, the main of which are chaperone-like activity, inhibition of protein aggregation, attenuation of oxidative stress, and activation of reparative autophagy through the mTOR-independent signaling pathways [1]. The drug has a promising translational potential in therapy for Alzheimer's disease (AD). In preclinical studies using animal models, trehalose dosages vary within a range of 1–5 % (usually 2 %) solution of the drug for drinking *ad libitum*. The degrees of autophagy induction and therapeutic effects vary either [2]. For the translational transition, it is of interest to define the optimal conditions and dosage for the use of the drug. Here we studied the therapeutic effect of oral intake of trehalose solution of 2 or 4 % concentration in drinking as well as an intermittent regimen of the drug treatment using a mouse model of AD induced by the central administration of the fragment of amyloid- $\beta$  25-35 (A $\beta$ 25-35).

**Methods and Algorithms:** C57BL6 mice were bilaterally stereotaxically injected with an aggregated amyloid fragment A $\beta$ 25-35 at a dose of 5  $\mu$ g into lateral ventricles of the brain [3]. 2 days after the surgery, trehalose treatment started. Mice have been given 2 % or 4 % trehalose solution as drinking for 14 days or 4 % trehalose solution as drinking for 14 days in an intermittent regimen (every other day). In the hippocampus, the mRNA levels of the autophagy-related genes (*Atg8*, *Park2*, and *Becn1*) were studied. Using immunohistochemical analysis of the brain cryosections, we evaluated A $\beta$  deposition, the expression of markers of autophagy (LC3-II) or neuroinflammation (IBA1), and neuronal density with the expression of neuronal marker NeuN or Nissl staining in the frontal cortex and hippocampus. Inflammatory markers (leukocyte elastase and  $\alpha$ 1-proteinase inhibitor) were measured in the serum samples. Recovery of cognitive function was assessed in the passive avoidance test.

**Results:** Treatment of the AD model with trehalose increased autophagy activity in the hippocampus, as measured by an increase in the expression of LC3-II. The maximal effect was achieved by the daily treatment with 4 % trehalose solution for two weeks. Trehalose elevated mRNA levels of autophagy-related genes *Atg8*, *Becn1*, and especially *Park2*. Trehalose intake reduced the A $\beta$  load in the frontal cortex and hippocampus with a maximal effect of every day treatment with the 4 % drug solution. A therapeutic effect was also obtained for the inhibition of neuroinflammation as estimated by the expression



of a microglial marker IBA1 and the restoration of neuronal density estimated by the NeuN expression or Nissl staining. The main therapeutic effect was revealed for the restoration of cognition in the passive avoidance test with a maximal effect of daily intake of 4 % drug solution; the effect of daily intake of 2 % trehalose solution was similar to the consumption of 4 % trehalose solution in the intermittent mode.

*Conclusion:* The AD model induced by injection of the A $\beta$ 25-35 fragment into the brain of mice demonstrated a coordinated response of hippocampal neurons to autophagy activation by trehalose, which included a decrease in A $\beta$  accumulation, inhibition of neuroinflammation, restoration of neuronal density, and repair of cognitive activity related to learning and memory. In general, the most prominent therapeutic effect was produced by the daily intake of 4 % trehalose solution.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (AAAA-A21-121011990039-2). The studies were partially implemented using the Unique scientific installation “Biological collection – Genetic biomodels of neuro-psychiatric disorders” (No. 493387) at the Scientific Research Institute of Neurosciences and Medicine.

### *References*

1. Sarkar S. Regulation of autophagy by mTOR-dependent and mTOR-independent pathways: autophagy dysfunction in neurodegenerative diseases and therapeutic application of autophagy enhancers. *Biochem Soc Trans.* 2013;41(5):1103-1130.
2. He Q. et al. Treatment with Trehalose Prevents Behavioral and Neurochemical Deficits Produced in an AAV  $\alpha$ -Synuclein Rat Model of Parkinson's Disease. *Mol Neurobiol.* 2016;53(4):2258-2268.
3. Tikhonova M.A. et al. Neuroprotective Effects of Ceftriaxone Involve the Reduction of A $\beta$  Burden and Neuroinflammatory Response in a Mouse Model of Alzheimer's Disease. *Front Neurosci.* 2021;15:736786. doi: 10.3389/fnins.2021.736786.

## Modern models of brain diseases for translational studies

Salmina A.B.

Research Center of Neurology, Moscow, Russia  
allasalmina@mail.ru

**Key words:** brain, animal model, *in vitro* model, translational neurology

**Motivation and Aim:** Development of new *in vivo* and *in vitro* models of brain diseases is one of the objectives of translational studies in Neuroscience. Solving this problem will ensure progress in understanding the mechanisms of brain functioning, deciphering the pathogenesis of central nervous system diseases (neurodegeneration, neurodevelopmental disorders, neuroinflammation, etc.) [1–4].

**Methods and Algorithms:** Using modern protocols of cell biology, neurobiology, and bioengineering, we have developed several original models of the neurovascular unit, neurogenic niche, and blood-brain barrier *in vitro*. These models allow us to reproduce some important aspects of intercellular communication, metabolic control, neurogenesis, angiogenesis/barriergenesis, functional activity and damage of neuronal, glial, and endothelial cells.

**Results:** We will present an analysis of up-to-date methods and approaches to *in vitro* modeling of brain tissue, application of various multicellular ensembles, scaffolds and structures to reproduce the mechanisms of plasticity, protocols used for monitoring the viability and functional competence of cells and tissues, and key criteria applied for such models, as well as prospects for their further improvement.

**Conclusion:** The choice of an adequate model and biomarkers is a key success factor in translational research in medicine. Modern technological solutions make possible the most reliable reconstruction of mechanisms of brain plasticity in the *in vitro* systems as well as to promote the development of new pharmacological agents with precise targeting and high efficacy.

### References

1. Salmina A.B., Malinovskaya N.A., Morgun A.V., Khilazheva E.D., Uspenskaya Y.A., Illarioshkin S.N. Reproducibility of developmental neuroplasticity in *in vitro* brain tissue models. *Rev Neurosci*. 2022. doi: 10.1515/revneuro-2021-0137.
2. Salmina A.B., Kapkaeva M.R., Vetchinova A.S., Illarioshkin S.N. Novel Approaches Used to Examine and Control Neurogenesis in Parkinson's Disease. *Int J Mol Sci*. 2021;22(17):9608. doi: 10.3390/ijms22179608.
3. Salmina A.B., Kharitonova E.V., Gorina Y.V., Teplyashina E.A., Malinovskaya N.A., Khilazheva E.D., Mosyagina A.I., Morgun A.V., Shuvaev A.N., Salmin V.V., Lopatina O.L., Komleva Y.K. Blood-brain barrier and neurovascular unit *in vitro* models for studying mitochondria-driven molecular mechanisms of neurodegeneration. *Int J Mol Sci*. 2021;22(9):4661. doi: 10.3390/ijms22094661.
4. Salmina A.B., Komleva Y.K., Malinovskaya N.A., Morgun A.V., Teplyashina E.A., Lopatina O.L., Gorina Y.V., Kharitonova E.V., Khilazheva E.D., Shuvaev A.N. Blood-Brain Barrier Breakdown in Stress and Neurodegeneration: Biochemical Mechanisms and New Models for Translational Research. *Biochemistry (Mosc)*. 2021;86(6):746-760. doi: 10.1134/S0006297921060122.

## Effect of chronic unpredictable mild stress on stress resilience in Disc1-Q31L mice

Smirnova K.V.<sup>1,2\*</sup>, Berdyugina D.A.<sup>1</sup>, Lipina T.V.<sup>1</sup>, Amstislavskaya T.G.<sup>1,2</sup>

<sup>1</sup>Scientific Research Institute of Neuroscience and Medicine, Novosibirsk, Russia

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

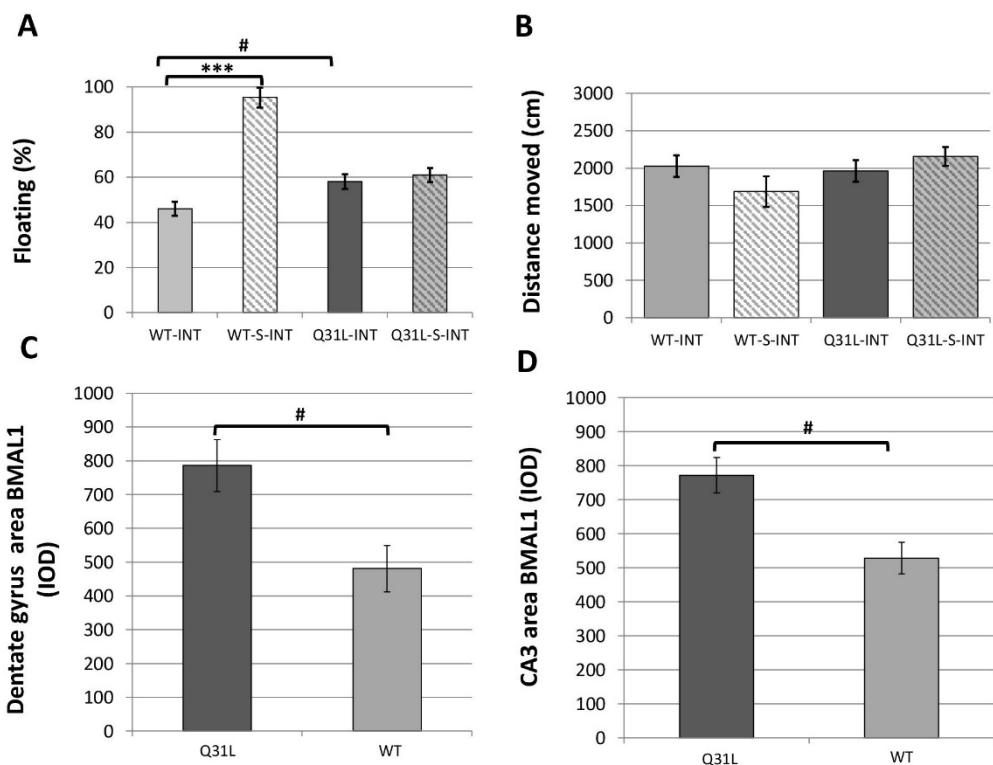
\* vedelina@mail.ru

**Key words:** *Disc1*, psychopathologies, stress, behaviour, BMAL1

**Motivation and Aim:** Mental illness in the modern world is an extremely common phenomenon, according to 2017 data, about 970 million people around the world suffer from them [1]. Adverse life events and stress are the main predictors for its development, especially in individuals genetically predisposed to these diseases [2]. Based on preclinical and clinical studies, it has been conceptualized that depression has complex etiology, including gene x stress interactions.. The chronic unpredictable mild stress (CUMS) protocol is an established method for modeling depression-like behavior in animals [3]. It is also known that CUMS increases the amount of DISC1 protein in the hippocampus [4]. At the same time, the lack of the DISC1 protein leads to the degradation of the BMAL1 protein, which has been shown to be involved in stress responses and mood disorders [5, 6]. *Disc1-Q31L<sup>-/-</sup>* mutant mice showed mood-related phenotypes [7] and, hence, study of the molecular mechanisms of *Disc1-Q31L* x CUMS interactions would facilitate understanding of the pathogenesis of depression.

**Methods and Algorithms:** In this study were used male mice of two lines C57BL/6 and *Disc1-Q31L<sup>-/-</sup>* at the age of 2–3 months. Animals were maintained under four week CUMS conditions, in accordance with previously published protocols [3]. Intact animals of both genotypes were used as comparison groups. Behavior was assessed in forced swimming (FS) and open field (OF) tests [7]. To assess the effect of a *Disc1* gene mutation on the amount of BMAL1 protein, the method of immunohistochemical analysis on frozen sections (IHC) was used. Micrographs were analyzed using the Image Pro-plus 6.0 software. Statistical analysis was performed using the STATISTICA program. For samples with a normal distribution, ANOVA analysis of variance was used, followed by post-hoc analysis by Fisher's method.

**Results:** The result of two-way ANOVA analysis showed a significant influence of the CUMS factor [ $F(1,30) = 55.105; p < 0.0001$ ], which caused pronounced depression-like behavior in C57BL/6 mice in the FS test (LSD,  $p < 0.0001$ ), in addition, the interaction of genotype and stress factors was shown [ $F(1,30) = 43.518; p < 0.0001$ ]. At the same time, no changes in motor activity were observed in the OF test. CUMS did not affect the behavior of *Disc1-Q31L<sup>-/-</sup>* mice in these tests (Fig. 1, A, B). Thus, a mutation in the *Disc1* gene prevents the occurrence of a behavioral response to CUMS. One-way ANOVA analysis of the IHC study results showed that in intact *Disc1-Q31L<sup>-/-</sup>* animals, there was a significant increase in the amount of BMAL1 protein in the areas of the dentate gyrus and CA3 areas of the hippocampus (LSD,  $p < 0.05$ ). (see Fig. 1, C, D).



**Fig. 1.** Effect of *Disc1* mutation. A – floating time in FS test. B – distance moved in OF test. C, D – IHC results, illuminance optical density (IOD). \* $p < 0.05$ , \*\*\* $p < 0.0001$

**Conclusion:** In the study was found that DISC1 plays an important role in the response to chronic stress. Disruption of its structure leads to increase in the amount of BMAL1 protein in the hippocampus, which likely prevents progression of abnormal behavior in response to CUMS.

### References

1. James S.L., Abate D. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858.
2. Uher R. Gene–environment interactions in severe mental illness. *Front Psychiatry*. 2014;5:48.
3. Antoniuk S. et al. Chronic unpredictable mild stress for modeling depression in rodents: Meta-analysis of model reliability. *Neurosci Biobehav Rev*. 2019;99:101-116.
4. Zhang X. et al. Venlafaxine increases cell proliferation and regulates DISC1, PDE4B and NMDA receptor 2B expression in the hippocampus in chronic mild stress mice. *Eur J Pharmacol*. 2015;755:58-65.
5. Lee S.B. et al. Disrupted-in-schizophrenia 1 enhances the quality of circadian rhythm by stabilizing BMAL1. *Transl Psychiatry*. 2021;11(1):1-13.
6. Leliavski A. et al. Impaired glucocorticoid production and response to stress in *Arntl*-deficient male mice. *Endocrinology*. 2014;155(1):133-142.
7. Lipina T.V. et al. Synergistic interactions between PDE4B and GSK-3: DISC1 mutant mice. *Neuropharmacology*. 2012;62(3):1252-1262.

## Rearrangements of actin cytoskeleton in atherosclerotic plaque development (experimental study)

Sudakov N.P.<sup>1,2\*</sup>, Klimenkov I.V.<sup>1,2</sup>, Popkova T.P.<sup>2</sup>, Lozovskaya E.A.<sup>2</sup>,  
Nikiforov S.B.<sup>2</sup>, Ezhikeeva S.D.<sup>3</sup>, Ten M.N.<sup>3</sup>, Konstantinov Yu.M.<sup>4</sup>

<sup>1</sup>Limnological Institute, SB RAS, Irkutsk, Russia

<sup>2</sup>Irkutsk Scientific Centre of Surgery and Traumatology, SB RAS, Irkutsk, Russia

<sup>3</sup>Regional Clinical Hospital of Irkutsk, Irkutsk, Russia

<sup>4</sup>Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia

\*npsudakov@gmail.com

**Key words:** hypercholesterolemia, atherosclerosis, foam cells, actin microfilaments, lipid droplets

*Motivation and Aim:* Atherosclerosis of the blood vessels of different localization is one of the main causes of mortality in the industrial countries [1]. The detailed conception of pathogenesis of this disease is very important for development of effective methods of prophylaxis and treatment. The cellular mechanisms of atherosclerosis development, particularly patterns of foam cells formation and behavior on the blood vessel inner surface during atherogenesis is poorly studied. The aim of study: to characterize to characterize the peculiarities of 3D-structure of actin cytoskeleton in atherosclerotic plaque morphogenesis.

*Methods and Algorithms:* Atherosclerosis was induced on "Chinchilla" rabbits with high-cholesterol diet (0.35 g of cholesterol per 1 kg of animal weight) during 3 months. Animals of control group was fed with standard diet. Rabbit peritoneal macrophages [2] was used for comparative control in foam cells cytochemical peculiarities studying. Aorta fragments and peritoneal macrophages were fixed with 4 % paraformaldehyde, permeabilized with 0.2 % Triton-X100, F-actin was stained with Alexa Fluor 488 Phalloidin, lipid droplets – Nile red, nuclei – Hoechst 33342. Slides were analyzed with laser confocal microscope LSM 710 (Zeiss Germany).

*Results:* Study of structural and functional peculiarities of peritoneal macrophages showed that in cell suspension they are spherical and they have dense submembrane actin cortex and some of them contains rare small lipid droplets in cytoplasm. There are not any blood cells adhere to endothelium on the aorta inner surface of the rabbit from control group. After 2 weeks of high-cholesterol diet were registered macrophages on the surface of aorta. These cells have spindle shape, contain multiple thin axially oriented actin microfilaments and practically do not carry lipid droplets. In the whole, macrophages on the aorta surface oriented parallel to the blood flow. Front edge of some cells trapezoidal widened and have filopodia, that is suggest about migratory activity. Basically, macrophages does not contact with one another and this cells forms clusters in the places of arteria branching from aorta. On the late stages of atherogenic diet in these locations forms atherosclerotic plaques. The analysis of rabbit aorta after 3 months of high-cholesterol diet, allowed reconstruction of the subsequent stages of atherogenesis. According to these data, non-contacting with one another macrophages begin deposit lipids, gradually transform to foam cells. Further plaque morphogenesis occurs through growth of quantity and volume of lipid droplets of existing foam cells, recruiting new macrophages, forming lateral contacts between neighbor cells with

building of monolayer, and later – multilayer structures. 3D Analysis of multilayer atherosclerotic plaques shows high morphological diversity of their cells: spheroidal, ellipsoidal, spindle, and pyramidal shaped cells. It can be assumed that cells of different form have different role and various type of activity in the tissue. The most of the cells contain thick actin cables oriented parallel to each other along the entire length of the cells. Among those cells are located the cells without thick actin cables. The cells with thick actin cables contacts with one another to form integral rigid actin framework of atherosclerotic plaque. Obviously, this is an adaptation of cell population of atherosclerotic plaque for preservation of integrity of structure in the conditions of shear stress generated with hydrodynamic forces of blood flow. It is possible that disruption structure of integral actin framework will be promote to atherosclerotic plaque destabilization and rupture, that is the main cause of thrombus formation.

*Conclusion:* Thus, in the beginning of atherogenesis the areas of monocyte/macrophage adhesion on the inner surface of aorta predetermines the shape of developing atherosclerotic plaque. Cells, composing the plaque are variable in shape, lipid droplets content and actin cytoskeleton structure. Nevertheless, the population of those cells have rigid integral actin framework, and, apparently, is a single morpho-functional complex. The studying of the patterns of morphogenesis and behavior of those cellular complexes will be promote understanding the mechanisms of the appearance of vulnerable plaques and the development of technologies for the diagnosis, prevention and treatment of complications of atherosclerosis.

### *References*

1. Dai H. et al. Global, regional, and national burden of ischemic heart disease and its attributable risk factors, 1990-2017: results from the global Burden of Disease Study 2017. *Eur Heart J Qual Care Clin Outcomes*. 2022;8(1):50-60. doi: 10.1093/ehjqcco/qcaa076.
2. Ar'Rajab A. et al. Enhancement of peritoneal macrophages reduces postoperative peritoneal adhesion formation. *J Surg Res*. 1995;58(3):307-312.

# New antiparkinsonian agent Diol restores cognitive function, enhances neurotrophic function, and diminishes neuroinflammation in a MPTP-induced model of Parkinson's disease in mice

Tikhonova M.A.<sup>1\*</sup>, Tenditnik M.V.<sup>1</sup>, Dubrovina N.I.<sup>1</sup>, Akopyan A.A.<sup>1</sup>,  
Ovsyukova M.V.<sup>1</sup>, Volcho K.P.<sup>2</sup>, Salakhutdinov N.F.<sup>2</sup>, Amstislavskaya T.G.<sup>1</sup>

<sup>1</sup>Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>2</sup>N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia

\* tikhonovama@physiol.ru

**Key words:** neurodegeneration, Parkinson's disease, cognitive deficits, BDNF, microglia, brain, mice

*Motivation and Aim:* Experimental therapy for cognitive deficits in neurodegenerative disorders including Parkinson's disease (PD) is area of high priority in current research. Recently synthesized monoterpene (1R,2R,6S)-3-methyl-6-(prop-1-en-2-yl)cyclohex-3-ene-1,2-diol 1 (Diol) alleviates motor manifestations of PD in animal models [1]. This study was aimed at evaluating the effects of Diol on cognitive deficits in mice with MPTP-induced PD-like pathology and the related mechanisms.

*Methods and Algorithms:* C57BL6 mice were injected with a single dose of MPTP (30 mg/kg) or saline intraperitoneally; Diol was given orally one time (20 mg/kg) on the next day of the experiment or chronically (20 mg/kg, daily, 31 days). Behavioral testing started 5 days after the first Diol administration. Biosamples for immunohistochemical analysis of the brain were taken 12 days after the first Diol administration.

*Results:* MPTP reduced learning and accelerated extinction of the conditioned response of passive avoidance that were associated with the increased microglia activation in the frontal cortex and hippocampus while the expression of neurotrophic factor BDNF was significantly decreased in the frontal cortex, hippocampus, and substantia nigra. Both types of Diol treatment completely reversed cognitive deficits in MPTP-treated mice. Moreover, Diol prevented neuroinflammatory response in the frontal cortex and hippocampus. Chronic Diol treatment restored the BDNF levels in the hippocampus.

*Conclusion:* The beneficial effect of Diol on MPTP-induced cognitive deficits, neuroinflammatory response, and neurotrophic function in the brain has been demonstrated in mice.

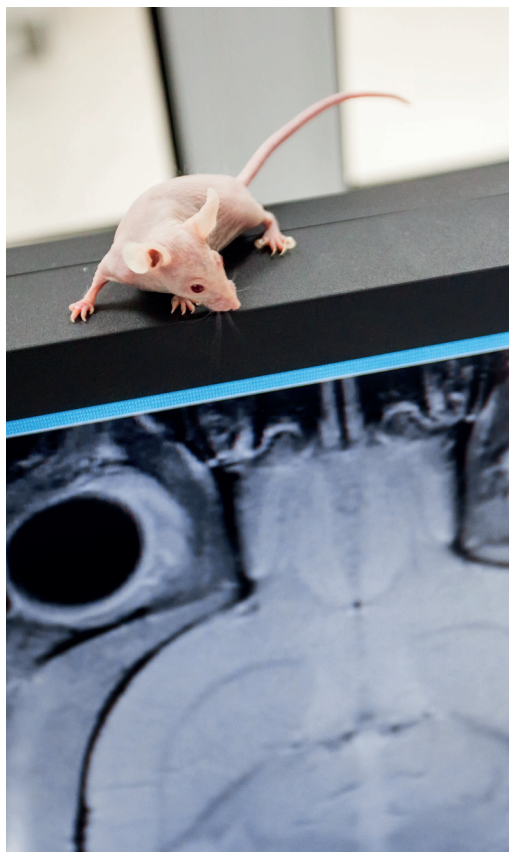
*Acknowledgements:* The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Neurosciences and Medicine (AAAA-A21-121011990039-2). The studies were partially implemented using the Unique scientific installation "Biological collection – Genetic biomodels of neuro-psychiatric disorders" (No. 493387) at the Scientific Research Institute of Neurosciences and Medicine.

## References

1. Ardashov O.V. et al. A novel small molecule supports the survival of cultured dopamine neurons and may restore the dopaminergic innervation of the brain in the MPTP mouse model of Parkinson's disease. *ACS Chem Neurosci*. 2019;10(10):4337-4349. doi: 10.1021/acchemneuro.9b00396.







# 8

Symposium “Biomedicine,  
bioinformatics and systems biology”



## piRNAs regulate expression of candidate genes in esophageal adenocarcinoma

Akimniyazova A.N.<sup>2\*</sup>, Niyazova T.K.<sup>2</sup>, Abdikarim G.M.<sup>2</sup>, Sagynbay Sh.<sup>2</sup>, Ivashchenko A.T.<sup>1</sup>

<sup>1</sup> Center for Bioinformatics and Nanomedicine, Almaty, Kazakhstan

<sup>2</sup> al-Farabi Kazakh National University, Almaty, Kazakhstan

\* a\_ivashchenko@mail.ru

**Key words:** piRNA, esophageal adenocarcinoma, candidate genes, diagnostic, treatment

*Motivation and Aim:* The available information on the biological role of piRNAs (PIWI-interacting RNA) remains at the conjecture stage. The involvement of piRNAs in the regulation of the expression of protein-coding genes is practically not studied. The purpose of this work is to show the possibility of a significant effect of piRNAs on the translation process through the interaction of these molecules with mRNA.

*Methods and Algorithms:* The nucleotide (nt) sequences of esophageal adenocarcinoma candidate genes were downloaded from NCBI. The nucleotide sequences of 200000 piRNAs were taken from Wang et al [1]. The piRNA binding sites (BSs) in the mRNAs of genes were predicted using the MirTarget program [2]. This program defines the following features of piRNA binding to mRNA: a) the initiation of piRNA binding from the first nucleotide of the mRNAs; c) the schemes of nucleotide interactions between piRNAs and mRNA; d) the free energy ( $\Delta G$ , kJ/mol) of the interaction between piRNA and the mRNA; e) the ratio  $\Delta G/\Delta G_m$  (%) is determined for each site ( $\Delta G_m$  equal to the free energy of piRNA binding with its fully complementary nucleotide sequence).

*Results:* The protein coding region of the *BTG3* gene includes from 261 to 1151 nucleotides. Binding sites of 33 piRNAs are located from 573 to 665 nt CDS mRNA with overlapping nucleotide sequences. Binding sites piR-12399, piR-37172, piR-44059, piR-126527, piR-129369 start at 626 nt, i.e. approximately in the middle of the complex of binding sites for all piRNA. The length of the CDS is 7.5 times greater than the length of the cluster of binding sites. In the CDS mRNA of the *CD55* gene, located from 295 nt to 1617 nt, 18 piRNAs binding sites were identified with the onset from 1420 nt to 1458 nt. That is, the length of the cluster of piRNA binding sites is 63 nt, which is 21 times less than the CDS length. The mRNA of the *ERBB2* gene included binding sites for only two piRNAs located in the 5'UTR with overlapping nucleotide sequences. Expression of the *ERBB3* gene may depend on 59 piRNAs whose binding sites are located in the 3'UTR from 4934 nt to 5139 nt. These piRNA binding sites formed a cluster from 4934 nt to 5171 nt, 238 nt long. The 3'UTR length was 6.1 times the cluster length. A feature of the cluster is the presence of an identical binding site for seven piRNAs: piR-5300, piR-5303, piR-5358, piR-6236, piR-7637, piR-65119, piR-96686. The mRNA of the *FKBP5* gene contained three clusters of piRNA binding sites located in the 3'UTR. In the first cluster, 27 piRNAs bound from 1366 nt to 1422 nt. The second cluster starts from 4854 nt to 4965 nt and binds 38 piRNAs. This cluster contains five binding sites with the same origin for piRNA at position 4862 nt: piR-56610, piR-62204, piR-75328, piR-92776, piR-102326. These five piRNAs have binding sites in the

third binding sites cluster at 6359 nt which includes 264 piRNAs binding sites from 6356 nt to 7329 nt. Binding sites for seven piRNAs start at position 6397 nt: piR-36976, piR-50827, piR-57845, piR-84920, piR-102738, piR-116971, piR-125671. In the 3'UTR of the mRNA of the *FOXPI* gene, there are only five piR-162173 binding sites, the beginning of which was located two nucleotides apart. This mRNA region contained 20 GU dinucleotide repeats. Nine piRNAs (piR-5300, piR-5303, piR-5358, piR-6236, piR-7637, piR-35905, piR-65119, piR-96686, piR-101606) from the third cluster bound at one position 7115 nt. These include seven piRNAs binding to the mRNA of the *ERBB3* gene. A group of five piRNAs (piR-45567, piR-93145, piR-96134, piR-123755, piR-169382) bound at position 7167 nt. Expression of the *LEP* gene may also depend on piRNAs, because 162 piRNAs can bind to its mRNA, including piR-23387, which is completely complementary due to canonical nucleotides. The binding sites of all piRNAs formed a cluster from 3084 nt to 3385 nt located in the 3'UTR from 562 nt to 3428 nt. The length of the binding sites cluster was 9.8 times less than the length of the 3'UTR. A group of nine piRNAs (piR-5300, piR-5303, piR-5358, piR-6236, piR-7637, piR-35905, piR-65119, piR-96686, piR-101606) with binding sites with 3101 nt coincides with a group of piRNAs that bind to the mRNA of the *FKBP5* gene and includes a group of seven piRNAs that interact with the mRNA of the *ERBB3* gene. The mRNA of the gene contains binding sites for three groups each of five piRNAs at positions 3087 nt, 3088 nt, and 3246 nt. 46 piRNAs bind to the mRNA of the *SEPP1* gene from 108 nt to 172 nt (65 nt) at a 5'UTR length of 188 nt, i.e., 2.9 times shorter. The length of all 46 binding piRNAs is 1367 nt, and due to the compaction of piRNAs binding sites into clusters, competition between piRNAs for binding in the cluster simultaneously arises. At position 123 nt of the 5'UTR of the mRNA of the *SEPP1* gene, a group of nine piRNAs was identified, identical to the groups of piRNAs that bind to the mRNA of the *FKBP5* and *LEP* genes. The 102 piRNAs bound to the 3'UTR of the *SMAD4* gene mRNA and their binding sites formed two clusters from 4316 nt to 4563 and from 7725 nt to 7778 nt. In the first cluster, three groups of five piRNAs were identified. The first group included piR-36976, piR-50827, piR-84920, piR-116971, piR-125671. The second group consisted of piR-19014, piR-19076, piR-82088, piR-114649, piR-125598, piR-98951. The third group was formed by piR-56309, piR-81517, piR-96803, piR-98951, piR-126140. Both clusters were only 301 nt (54 nt and 247 nt) from the 3'UTR length of 6576 nt, which is 21.8 times the length of the clusters. There were no clusters of binding sites in the mRNA of the *TP53* gene for groups with 5 piRNAs or more. The cluster of 33 piRNAs binding sites with a length of 263 nt was 5.4 times less than the 3'UTR length of 1408 nt. piR-44059 and piR-83728 bound to the mRNA of the *TP53* gene with  $\Delta G$  kJ/mol equal to  $-170$  and  $172$  kJ/mol respectively, i.e. they could strongly inhibit the translation process.

*Acknowledgements:* The study is supported by Center for Bioinformatics and Nanomedicine, Almaty, Kazakhstan.

## References

1. Wang J. et al. piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.* 2019;47:175-180. doi: 10.1093/nar/gky1043.
2. Ivashchenko A. et al. MiR-3960 Binding Sites with mRNA of Human Genes. *Bioinformation.* 2014;10:423-427. doi: 10.6026/97320630010423.

## Functional study of rs590352 (*ATXN7L3B*) associated with colorectal cancer

Antontseva E.<sup>1\*</sup>, Degtyareva A.<sup>1</sup>, Aliev T.<sup>1,2</sup>, Merkulova T.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [antontseva@bionet.nsc.ru](mailto:antontseva@bionet.nsc.ru)

**Key words:** regulatory SNP, *ATXN7L3B*, colorectal cancer, functional analysis

**Motivation and Aim:** Previously, using a bioinformatics approach based on the analysis of genome-wide data about allele asymmetry of chromatin protein and transcription factors (TF) binding and allele asymmetry of gene expression, we identified a number of potentially regulatory polymorphisms (rSNPs) [1]. Among them was rs590352 G>C, localized in exon 1 of the *ATXN7LB* gene, for which an association with the risk of developing colorectal cancer was established, the G allele is protective [2]. The aim of the work is an experimental study of the functionality of rs590352.

**Methods and Algorithms.** Electrophoretic mobility shift assay (EMSA) – method of DNA-probes retention in the gel by nuclear extract proteins was used to study the effect of single nucleotide substitution on TF binding. To search for TFs whose binding sites are affected by single nucleotide substitution, the MotifbreakR software package was used. Oligonucleotides containing experimentally confirmed binding sites for these TFs were used as competitors in EMSA experiments.

The effect of nucleotide substitution on the level of gene expression was studied in a luciferase reporter system. Vector constructs with allelic variants of the inserts were created in pGL4.23[luc2/minP] plasmid. The HepG2 cell line was used for transfection. To assess the allele-asymmetric expression of the *ATXN7LB* gene in human blood mononuclear cells, cDNA libraries of 15 healthy donors heterozygous (rs590352 G/C) were created and sequenced. Sequencing was performed on the Illumina MiSeq platform with  $\geq 1000$  reads per library.

**Results:** In EMSA experiments with a nuclear extract from the HepG2 cell line, it was shown that oligonucleotides V1 (G allele) and V2 (C allele), reproducing allelic variants of the rs590352 location region, have differences in the intensity of the bands on radioautograph, i.e., the G>C nucleotide substitution affects the binding of certain proteins to this region. In cross-competition experiments, when unlabeled V1 and V2 oligonucleotides alternately competed with the labeled DNA probe (V1 or V2), it was shown that a 5-fold and 10-fold excess of competitor V1 attenuates the intensity of DNA probe retention bands by some TFs more strongly than competitor V2, i.e. in the case of the reference allele G, the affinity for some TFs is higher than in the case of allele C. To identify this/these TF by competitive analysis, oligonucleotides were used containing binding sites for: ASCL1, ASCL2, ATOH1, TCF3, TFAP4, ZBTB, MYOG, NHLH1, OLIG2 and PTF1A, predicted using the MotifbreakR software package. It was shown that TCF3 and PTF1A oligonucleotides compete with V1 and V2 for protein binding, in upper and lower bands respectively. Thus, the G>C substitution was shown to weaken TCF3 and PTF1A binding sites. It should be noted that, according to the data of the

ENCODE project, this SNP falls within the ChIP-seq peak for TCF12 (TCF3 paralog), which indirectly confirms the correctness of our result.

It was also shown in the reporter system that luciferase gene expression increases upon transfection with a construct containing double and triple inserts with the G allele. It is of great interest to study the effect of rs590352 on *ATXN7LB* gene expression *in vivo*. In this regard, we evaluated the allele-specific expression of *ATXN7LB* in blood mononuclear cells of heterozygous individuals. It was shown that the expression of the *ATXN7LB* gene in the case of the G allele is 1.3 times higher compared to the C allele.

*Conclusion:* Thus, our results show that the *ATXN7L3B* gene rs590352, which we previously associated with colorectal cancer, is a regulatory SNP. At the same time, an analysis of the literature data showed that *ATXN7LB* reduces the activity of the SAGA histone acetyltransferase complex, which implements histone co-transcriptional modifications, leading to an increase in the level of ubiquitinated histone (H2Bub1) and a decrease in transcriptional activity [3]. It has also been shown that H2Bub1 suppresses tumor progression, and a decrease in its level has been observed in the late stages of cancer [4, 5]. Based on our results (improved protein-binding capacity of the G allele, increased reporter gene and *ATXN7LB* gene expression in mononuclears also in the case of the G allele), and given that the G allele was associated with a reduced risk of colorectal cancer, and taking into account that according to the literature increased expression of *ATXN7L3B* is indirectly associated with the suppression of carcinogenesis, the oncoprotective effect of the G allele rs590352 of the *ATXN7L3B* gene can be assumed.

*Acknowledgements:* This work was supported by the State Budget Project FWNR-2022-0016.

### References

1. Korbolina E.E. et al. Novel approach to functional SNPs discovery from genome-wide data reveals promising variants for colon cancer risk. *Hum Mutat.* 2018;39(6):851-85.
2. Leberfarb E.Yu. et al. Potential regulatory SNPs in the *ATXN7L3B* and *KRT15* genes are associated with gender-specific colorectal cancer risk. *Per Med.* 2020;17(1):43-54.
3. Li W. et al. Cytoplasmic *ATXN7L3B* Interferes with Nuclear Functions of the SAGA Deubiquitinase Module. *Mol Cell Biol.* 2016;36(22):2855-2866.
4. Hahn M.A. et al. The tumor suppressor *CDC73* interacts with the ring finger proteins *RNF20* and *RNF40* and is required for the maintenance of histone H2B monoubiquitination. *Hum Mol Genet.* 2012;21(3):559-68.
5. Urasaki Y. et al. Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors. *PLoS One.* 2012;7(5):e36775.

## **Prioritization of pathologies based on the analysis of gene networks of various macrophage phenotypes using ANDSystem**

Antropova E.A.<sup>1\*</sup>, Demenkov P.S.<sup>1,2</sup>, Ivanisenko V.A.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *nzhenia@bionet.nsc.ru*

**Key words:** macrophage, gene networks, prioritization, automatic analysis of texts

*Motivation and Aim:* Macrophages (immune system cells) are the central control element of immune responses [1]. Depending on the stimulus received, the plasticity of the macrophage response leads it to a pro-inflammatory phenotype (M1) or to one of the anti-inflammatory phenotypes of the M2 group, which promote wound healing and tissue regeneration [2]. In publications, there are observations about the relationship of M1 or M2 macrophages phenotypes with certain pathologies, or the relationship of the outcome of the disease with a certain macrophages phenotype. Understanding of the molecular genetic processes that ensure the plasticity of the macrophages can make it possible to control their functions.

*Methods and Algorithms:* Automatic analysis of a large amount of scientific data (published articles and factual databases information) was carried out by using of the ANDSystem system, which contains the ANDVisio program for visualization and analysis of gene networks [3]. ANDSystem software tools was also used to search for associated biological processes and pathologies, and their further prioritization. Cluster analysis of biological processes and pathologies ontologies was carried out by using the MCL method [4], with using Wang's measure of semantic similarity of biological processes [5].

*Results:* We have reconstructed the gene networks of five macrophage phenotypes (M1, M2a, M2b, M2c, M2d). The gene networks were based on groups of genes and proteins, the expression of which is significantly increased in the studied macrophage phenotypes according to published experimental data. Relationships between objects are reconstructed based on automatic analysis of texts and database contents. The search for associations of gene networks of M1, M2a, M2b, M2c, M2d macrophages with biological processes and their subsequent prioritization was carried out, the most significant ones for each macrophage phenotype were identified. The next step was to identify pathologies associated with different macrophage phenotypes using the built-in ANDSystem tools and prioritize them with the identification of the most typical macrophages for each phenotype.

*Conclusion:* The identified biological processes and pathologies that are specific for different macrophage phenotypes can complete the understanding of the causes and consequences of the experimentally shown facts of their relationships.

*Acknowledgements:* The study is supported by budget project 0259-2021-0009.

## References

1. Mills C.D. M1 and M2 Macrophages: Oracles of Health and Disease. *Crit Rev Immunol.* 2012;32(6):463-88.
2. Gurvich O.L. et al. Transcriptomics uncovers substantial variability associated with alterations in manufacturing processes of macrophage cell therapy products. *Sci Rep.* 2020;10(1):14049.
3. 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.
4. van Dongen S. et al. Using MCL to extract clusters from networks. *Methods Mol Biol.* 2012;804:281-295.
5. Wang J.Z., Du Z., Payattakool R., Yu P.S., Chen C.F. A new method to measure the semantic similarity of GO terms. *Bioinformatics.* 2007;23(10):1274-1281.



## cf mtDNA as promising biomarker of radon-induced lung cancer

Bersimbaev R.\*, Bulgakova O., Kussainova A., Aripova A.

*L.N. Gumilyov Eurasian National University, Institute of Cell Biology and Biotechnology, Nur-Sultan, Kazakhstan*

\* *ribers@mail.ru*

**Key words:** liquid biopsy, radon, cf mtDNA

**Motivation and Aim:** In recent years, liquid biopsy is becoming an important tool for assessing the progress or onset of lung disease. Liquid biopsy is a non-invasive procedure able to isolate free circulating DNA from body fluids [1]. Cell free circulating mitochondrial DNA (cf mtDNA) have been used not only as a biomarker of carcinogenesis but also as a biomarker of exposure to radiation. Radon is a colorless, odorless, tasteless gas that is the radioactive decay product of uranium, and source of high linear energy transfer  $\alpha$ -particles, which can affect the human body. According to the World Health Organization, radon is the second leading cause of lung cancer after smoking [2]. Many areas of Kazakhstan are characterized by an increased level of radon hazard [3]. However, the influence of environmental factors, including radon, on the level of cf mtDNA is still not studied at the moment. Research in this area can contribute to the development of "preventive diagnostics" aimed at identifying potential health risks associated with the influence of an adverse environmental factors or an unhealthy lifestyle.

**Methods and Algorithms:** 207 subjects were examined including 41 radon-exposed lung cancer patients, 40 lung cancer patients without radon exposure and 126 healthy controls exposed/not exposed to high level of radon. Total cell free circulating DNA from blood samples was extracted and used to detect cf mtDNA copy number by quantitative real-time polymerase chain reaction (qRT-PCR).

**Results:** Among the 62 healthy individuals (control), the median of cf mtDNA copy number was  $1.21 \times 10^5$  copies/ml. The median of cf mtDNA copy number in 64 healthy individuals living in the areas with high level of radon (radon) was  $5.14 \times 10^5$  copies/ml. The median of cf mtDNA copy number in 41 lung cancer (LC) patients living in the areas with low level of radon was  $3.05 \times 10^5$  copies/ml. The median of cf mtDNA copy number in the 40 radon-induced lung cancer (RLC) patients was  $4.14 \times 10^6$  copies/ml. The data indicate that the level of cf mtDNA in the radon-induced lung cancer patients was significantly higher than that of the other study participants. In addition, a correlation was found between the cf mtDNA copy number in the blood plasma of patients with radon-induced lung cancer and annual effective doses of radon.

**Conclusion:** We have shown for the first time that cf mtDNA can be an indicator of the impact of radon on the human body. Moreover, we believe that cf mtDNA is a promising biomarker for liquid biopsy of radon-induced lung cancer.

**Acknowledgements:** This study was supported by the Ministry of Science and Education of the Republic of Kazakhstan (Grant No. AP08856116).

### References

1. Bersimbaev R., Pulliero A., Bulgakova O., Asia K., Aripova A., Izzotti A. Radon Biomonitoring and microRNA in Lung Cancer. *Int J Mol Sci.* 2020;21(6):2154. doi: 10.3390/ijms21062154.
2. Bersimbaev R., Bulgakova O. Residential radon exposure and lung cancer risk in Kazakhstan. In: Adrovic F. (Ed.). London: Radon. InTech, 2017. doi: 10.5772/intechopen.71135.
3. Bersimbaev R.I., Bulgakova O. The health effects of radon and uranium on the population of Kazakhstan. *Genes Environ.* 2015;37:18. doi: 10.1186/s41021-015-0019-3.

# Swaveform: a genome-wide survey of structural variation profiles

Bezdvornyykh I.\*, Kanapin A., Samsonova A.

*Institute for Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia*

\* [latur@me.com](mailto:latur@me.com)

**Key words:** structural variation, waveforms, motif discovery, whole genome sequencing, signal processing, machine learning

*Motivation and Aim:* Genome structural variation (SV) is a potent source of genetic diversity and may have a profound effect upon human health [1, 2]. Typically, discovery algorithms rely on uniformity and evenness of sequencing coverage profile, as well as read depth information for accurate detection of structural variants [3]. However, as sequencing coverage signal is discontinuous, heterogeneous, and irregular, often even erratic existing SV detection tools still generate highly discordant results [4–6]. A comprehensive survey of signal coverage profiles in the vicinity of SV breakpoints may significantly improve robustness and reliability of SV discovery. Here, for the first time we present a detailed catalogue of various waveforms and patterns observed in the sequence coverage signal associated with different types of SVs, as well as a toolkit for coverage data management and analytics.

*Methods and Algorithms:* We used HGDP sequencing project data [7] which includes 929 whole-genome sequenced human samples with an average depth of coverage of about 30x. Structural variation data generated by the consortium contains annotated breakpoints for the following types of the events: CNV gain and loss, insertions, inversions, deletions and duplications [8]. We have developed a software suite to extract depth of coverage (DOC) profiles in a  $\pm 256$  bp vicinity of the breakpoints. Furthermore, to speed up coverage data processing and optimize storage we introduced a simple binary format (.bcov) for recording of coverage values. The extracted profiles and sequencing samples metadata are stored in a relational database (SQLite) to facilitate data search, retrieval and visualization.

A profound variability of waveforms associated with different classes of SVs has long impeded the reliability and reproducibility of the discovery algorithms. We, therefore, sought to identify repeated patterns found within DOC profiles (i.e., motifs) and characterize conserved structures in a signal. Coverage profiles attributed to specific classes of SVs were compressed, normalised and clustered using dynamic time warping distance. To reveal predominant waveforms and to discover the most significant motifs from the profiles, the signal was transformed into a SAX (Symbolic Aggregate approximation) representation [9]. The SAX-based motifs, stored in the database may have an important utility for development of novel improved approaches for breakpoint detection.

*Results:* In this study we present Swaveform – a recourse providing easy access to 15705161 DOC signal profiles extracted from 911 human sequencing samples generated by the HGDP. A portable database architecture and provided API facilitate easy and seamless on-premises deployment encompassing data processing routines on all levels i.e., from raw aligned data to visual representation of coverage profiles. We also propose

a new binary format to manage sequence coverage data. Finally, as motif discovery has been successfully applied throughout a large range of domains such as medicine, finance, robotics and DNA analyses we designed an algorithm for motif extraction from the coverage signal.

*Conclusion:* Robust and reproducible structural variation discovery still poses significant computational and algorithmic challenges and as of today we are nowhere near to resolving structural variation in personal genomes with accuracy required for translational research. The developed catalogue and accompanying toolkit form an indispensable resource that will facilitate development and honing of computational tools for discovery of specific genomic rearrangement events. It is expected that Swaveform will be of immense value to the machine learning and biomedical communities.

*Acknowledgements:* This work is supported by the Russian Scientific Foundation grant 20-14-00072.

### References

1. Collins R.L. et al. A structural variation reference for medical and population genetics. *Nature*. 2020;581(7809):444-451. doi: 10.1038/s41586-020-2287-8.
2. Liu Z., Roberts R., Mercer T.R., Xu J., Sedlazeck F.J., Tong W. Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.* 2022;23(1):68. doi: 10.1186/s13059-022-02636-8.
3. Cameron D.L., Stefano L., Papenfuss A.T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun.* 2019;10(1):1-11. doi: 10.1038/s41467-019-11146-4.
4. Parikh H. et al. Svclassify: a method to establish benchmark structural variant calls. *BMC Genomics*. 2016;17(1):64. doi: 10.1186/s12864-016-2366-2.
5. Chaisson M.J.P. et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun.* 2019;10(1):1-16. doi: 10.1038/s41467-018-08148-z.
6. Chapman L.M. et al. A crowdsourced set of curated structural variants for the human genome. *PLoS Comput Biol.* 2020;16(6):e1007933. doi: 10.1371/journal.pcbi.1007933.
7. Bergström A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 2020;367(6484):eaay5012. doi: 10.1126/science.aay5012.
8. Almarri M.A. et al. Population structure, stratification, and introgression of human structural variation. *Cell*. 2020;182(1):189-199.e15. doi: 10.1016/j.cell.2020.05.024.
9. Malinowski S., Guyet T., Quiniou R., Tavenard R. Advances in Intelligent Data Analysis XII. *Lect Notes Comput Sci.* 2013;8207:273-284. doi: 10.1007/978-3-642-41398-824.

## Examination of myocardium and adrenal glands structure in patients with COVID-19

Bgatova N.P.<sup>1\*</sup>, Savchenko S.V.<sup>1,2</sup>, Lamanov A.N.<sup>2</sup>, Miguel A.A.<sup>2</sup>, Letyagin A.Yu.<sup>1</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State Medical University, Novosibirsk, Russia*

\* *nataliya.bgatova@yandex.ru*

**Key words:** COVID-19, myocardium, adrenal glands, histology

*Motivation and Aim:* The new coronavirus, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has caused numerous deaths worldwide [1]. However, little is known about the causes of death [2]. Along with the development of viral pneumonia in COVID-19, extrapulmonary lesions, in particular the heart, are observed. The contractile activity of the heart is coordinated by the hypothalamic-pituitary-adrenal axis. In the terminal period with COVID-19, cardiac function is supported by adrenomimetics. The aim of this study was to analyze the structure of the myocardium and adrenal glands in patients who died from COVID-19.

*Methods and Algorithms:* A histological study was carried out on autopsy material from patients ( $n = 8$ ) died from Covid-19. As a control, autopsy material was used from "blunt trauma of the body" patients ( $n = 3$ ). Tissue samples were fixed in a 4 % paraformaldehyde solution. The formalized samples were subjected to a standard procedure in the histological tissue processor STP 120. Paraffin slices were stained with hematoxylin and eosin and examined under a Leica DME microscope.

*Results:* The study of the heart revealed acute circulatory system disorders, dystrophic and necrobiotic changes in cardiomyocytes. Foci of contractures of I-III degrees, myocytolysis of cardiomyocytes were observed in the myocardium. Intracellular lipid inclusions were detected in cardiomyocytes. In the zona glomerulosa of the adrenal cortex, dilated lumens of blood capillaries, the presence of large lipid droplets in cortical adrenolocyte and necrosis of individual cells were revealed. The medulla was characterized by medullary cells swelling and destructive changes, significantly enlarged of blood capillaries lumens. In the zona reticularis of the adrenal cortex, swelling of cortical adrenolocyte, accumulations of erythrocytes, lymphocytes and plasma cells, and large lipid droplets in the lumen of capillaries were observed. The most pronounced structural changes were noted in the zona fasciculate of the adrenal cortex. The heterogeneity of cortical adrenolocyte of the content of lipid inclusions was revealed. Cells with a small amount of lipid droplets, medium and hypertrophied cells with a large amount of lipids in the cytoplasm were observed. Signs of cell apoptosis were noted.

*Conclusion:* The data obtained indicate significant changes in the structure of the myocardium and adrenal gland in patients with COVID-19. The revealed structural changes can be informative in the development of cardioprotection to reduce mortality.

*Acknowledgements:* The study is supported by Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS (NFWNR-2022-0012).

### References

1. Menter T. et al. Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology*. 2020;77(2):198-209.
2. Wichmann D. et al. Autopsy findings and venous thromboembolism in patients with COVID-19: A prospective cohort study. *Ann Intern Med*. 2020;173(4):268-277.

# Mathematical model dynamics of pancreatic cancer cells with interaction of immune system. Approach by replicator quasispecies system

Bratus A.<sup>1\*</sup>, Leslie N.<sup>2</sup>, Bocharov G.<sup>3</sup>, Chamo M.<sup>1</sup>, Yurtchenko D.<sup>4</sup>

<sup>1</sup> Russian University of Transport (MIIT), Moscow, Russia

<sup>2</sup> Harriot-Watt University, Edinburgh, UK

<sup>3</sup> Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia

<sup>4</sup> ISVR University, Southampton, UK

\* bratus.alexander@yandex.ru

**Key words:** Dynamics of growth pancreatic cancer, differentiated description by replicator quasispecies system, interaction with immune system

*Motivation and Aim:* Pancreatic cancer one of the difficult problems of medicine. Development illness goes very slowly during long time in special sake (stroma) and appears at final stage. Another feature of this illness is coexistence (symbiotic) effect for cancer cells and normal cells inside stroma [1]. All this subjects considerable complicate the process of discover and treatment illness.

*Methods and algorithms:* Formation of pancreatic pre-cancer and cancer cells represent branching stochastic process contained 64 cells. Mathematical model considered mutation matrix, fineness landscape matrix and death rates. Each element of mutation matrix present probability of appears mutation as a result to arise next cell in sequence of cancer and pre-cancer cells. Interaction immune system and of cancer cells and symbiotic effect of coexistence are considered.

*Results:* Differentiated description for each species pancreatic cancer cells with help of quasispecies system make possible to investigate dynamics of all set of cancer cells and give value additional information on development illness. The result of mathematical model well agreement with experimental data.

*Conclusion:* Applied approach can be be used to other kinds of cancer if there exist information on mutation and proliferation of cancer and pre-cancer cells.

*Acknowledgements:* The study is supported by Russian Fond of Basic Research, grant No. 21-51-10006.

## References

1. Louzoun Y., Xue C., Lesiniski G., Friedman A. A Mathematical model for pancreatic cancer sells growth and treatment. *J Theor Biol.* 2014;351:74-82.

## Medical genetics and cancer studies based on next-gen sequencing presented at the bioinformatics conferences and young scientists schools

Chen M.<sup>1\*</sup>, Feng C.<sup>1</sup>, Ivanisenko V.A.<sup>2,3</sup>, Klimontov V.V.<sup>2</sup>, Korobeynikova A.V.<sup>4</sup>, Orlov Y.L.<sup>2,3,4,5</sup>

<sup>1</sup> College of Life Sciences, First Affiliated Hospital of Medical School, Zhejiang University, Hangzhou, China

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

<sup>5</sup> Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia

\* mchen@zju.edu.cn

**Key words:** medical genetics, biomedicine, bioinformatics, education, oncology, sequencing

**Motivation and Aim:** Next-Generation Sequencing (NGS)-driven analysis and Systems Biology approaches commonly serve as a backdrop for a study of a tumor genome. This field became extremely important for biomedical studies to be discussed at the series of international events [1, 5]. Thus, recent issue of *BMC Medical Genomics* presents recent works discussed at the 11th Young Scientists School SBB-2019 “Systems Biology and Bioinformatics”, held in Novosibirsk, Russia [8, 9]. The SBB school series on bioinformatics proceeds annually since 2008 under the joint steerage of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences and Novosibirsk State University. We had publications in special topic issues after the Schools in *BMC Genomics*, *BMC Medical Genomics* and related BioMed Central family journals since 2014 [7].

**Methods and Algorithms:** Recently we have organized the Research Topic (special journal issue) at *Frontiers in Genetics* to collect the papers focused on the ancestry studies [3]. We have organized special journal issue in *Frontiers in Genetics* considering genetic background and molecular mechanisms of the human diseases [5]. We collated some cancer gene expression studies, some mutation profiling studies as well as some insightful case reports [8]. The SBB Schools in Novosibirsk were initially conceived as satellite event for young scientists held at the same time as BGRS\SB (Bioinformatics of Genome Regulation and Structure\Systems Biology) conference series, since 1998 taking place biannually. The articles comprising the issue of *BMC Medical Genomics* are focused on cancer genomics [8]. Using transcriptomic data, bioinformatic models can be built for patient-oriented ranking of cancer drugs.

**Results:** We have considered the themes on gene network studies and next-gen sequencing for the special issues [8, 6]. The articles comprising this issue of *BMC Medical Genomics* are focused on cancer genomics [10]. Using transcriptomic data, bioinformatic models can be built for patient-oriented ranking of cancer drugs. Thus, the database for cancer gene expression profiles associated with clinical outcomes of the chemotherapy treatments [8]. Authors mined Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA) and Tumor Alterations Relevant for GENomics-driven Therapy (TARGET) repositories to pull a database of gene expression profiles

associated with clinical responses on chemotherapy. The cases represented breast cancer, lung cancer, low-grade glioma, endothelial carcinoma, multiple myeloma, adult leukemia, pediatric leukemia and kidney tumors and suitable for machine learning study of these malignancy. A.Lavrov and co-authors reviewed pathogenic variants targetable by single base editing [4]. E.Pudova and colleagues discussed miRNA expression in cancer [10]. The trends in cancer genomics and bioinformatics [1] extend towards ncRNA studies, gene network and associative network models [11, 6].

*Conclusion:* We hope for continuing international exchange and education via the schools and competitions for young scientists in Russia. We support systems biology meetings and young scientists school looking for new organizational formats (online and offline, hybrid forms). Digital Medicine Forum (Digital Medicine Forum) and MGNGS-2022 (Medical Genetics – Next-Generation Sequencing) (<http://ngs.med-gen.ru>) events are organized in spring 2022. Importance of Big Data analysis methods based on NGS continue to be challenging biomedical problems [2].

*Acknowledgements:* This publication has been supported by the Sechenov University and Potanin Foundation grant for masters' teachers (ГК22-000797).

### References

1. Anashkina A.A., Leberfarb E.Y., Orlov Y.L. Recent trends in cancer genomics and bioinformatics tools development. *Int J Mol Sci.* 2021;22(22):12146. doi: 10.3390/ijms222212146.
2. Chen M., Harrison A., Shanahan H., Orlov Y. Biological big bytes: integrative analysis of large biological datasets. *J Integr Bioinform.* 2017;14(3):20170052. doi: 10.1515/jib-2017-0052.
3. Das R., Tatarinova T.V., Galieva E.R., Orlov Y.L. Editorial: association between individuals' genomic ancestry and variation in disease susceptibility. *Front Genet.* 2022;13:831320. doi: 10.3389/fgene.2022.831320.
4. Lavrov A.V., Varenikov G.G., Skoblov M.Y. Genome scale analysis of pathogenic variants targetable for single base editing. *BMC Med Genomics.* 2020;13(Suppl 8):80. doi: 10.1186/s12960-020-00735-8.
5. Orlov Y.L., Anashkina A.A., Tatarinova T.V., Baranova A.V. Editorial: Bioinformatics of genome regulation. Volume II. *Front Genet.* 2021a;12:795257. doi: 10.3389/fgene.2021.795257.
6. 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.* 2021b;22(18):9962. doi: 10.3390/ijms22189962.
7. Orlov Y.L., Baranova A.V., Markel A.L. Computational models in genetics at BGRS\SB-2016: introductory note. *BMC Genetics.* 2016;17(Suppl 3):155. doi: 10.1186/s12863-016-0465-3.
8. Orlov Y.L., Voropaeva E.N., Chen M., Baranova A.V. Medical genomics at the Systems Biology and Bioinformatics (SBB-2019) school. *BMC Med Genomics.* 2020;13(Suppl 8):127. doi: 10.1186/s12920-020-00786-x.
9. Orlov Y.L., Fernandez-Masso J.R., Chen M., Baranova A.V. Medical genomics at Belyaev Conference-2017. *BMC Med Genomics.* 2018;11(Suppl 1):11. doi: 10.1186/s12920-018-0324-3.
10. Pudova E.A., Krasnov G.S., Nyushko K.M. et al. miRNAs expression signature potentially associated with lymphatic dissemination in locally advanced prostate cancer. *BMC Med Genomics.* 2020;13(Suppl 8):129. doi: 10.1186/s12920-020-00788-9.
11. Saik O.V., Nimaev V.V., Usmonov D.B. et al. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med Genet.* 2019;12(Suppl 2):47. doi: 10.1186/s12920-019-0492-9.
12. Tatarinova T.V., Chen M., Orlov Y.L. Bioinformatics research at BGRS-2018. *BMC Bioinformatics.* 2019;20(Suppl 1):33. doi: 10.1186/s12859-018-2566-7.

## The pattern of YAP1 distribution in the normal and pathological state of the human epidermis

Cherkashina O.L.\*, Beilin A.K., Kosykh A.V., Rippa A.L., Chermnykh E.S., Kalabusheva E.P., Vorotelyak E.A.

*Koltzov Institute of Developmental Biology, RAS, Moscow, Russia*

\**olgalcher@gmail.com*

**Key words:** YAP1, skin regeneration, skin xenograft, living skin equivalent

*Motivation and Aim:* YAP1 is a coactivator that binds to TEAD transcription factors and in the skin it controls the maintenance of homeostasis and regeneration due to interaction with the signaling pathways Wnt/b-catenin, RhoA/ROCK, NOTCH, SHH. In skin pathologies associated with hyperproliferation of keratinocytes, such as psoriasis, basal cell carcinoma, squamous cell carcinoma, there is an increase in the expression level and content of the activated form of YAP1 in the epidermis, while its reduced level and lack of activated form are associated with impaired expansion of basal keratinocytes. Currently available studies of human skin regeneration describe certain aspects of YAP1 participation in the regenerative process, however, it is necessary to conduct a comprehensive analysis of the role of YAP-signaling in the process of skin regeneration and the relationships between processes of proliferation and differentiation in normal and pathological conditions.

*Methods and Algorithms:* The following *in vivo* and *in vitro* models were used: xenografting of human skin to NOD/SCID mice (specimens were studied 40, 75 and 110 days after transplantation), live skin equivalents, biopsies of cutaneous papilloma, chronic wound during osteomyelitis, as well as healthy human skin. Immunohistochemically stained cryosections were used for measurement of morphometric parameters of model objects (epidermis thickness, thickness of dermal layers, adipocyte sizes) and the distribution of markers were calculated using ImageJ and Cell Profiler programs based on data on colocalization of signals from various channels and fluorescence intensity. The expression levels of the genes of interest were studied using quantitative PCR analysis. Statistical processing was carried out using GraphPad Prism 8.

*Results:* To study the dynamics of skin regeneration and assess the distribution of YAP1, we have developed a unique model of human skin xenotransplantation to mice, which reproduces the structure of human skin and contains all the main components: epidermis, dermis, adipose tissue, as well as hair follicles, sebaceous and sweat glands. Morphometric analysis of the grafts revealed a number of differences from healthy skin, including thickening of the epidermis in the graft on day 40 (the average thickness reaches 100 microns, while the thickness before transplantation averaged 19 microns). By 75 and 110 days, its thickness gradually reduces, but is still increased compared to normal skin, which indicates an incomplete regenerative process.

For quantitative evaluation, an algorithm was developed for analyzing the content of the nuclear and cytoplasmic form of YAP1 in the basal and suprabasal layers of the epidermis in ImageJ based on image segmentation and determining the presence of colocalization of signals in channels with YAP1 and with DAPI. YAP1 is normally



located mainly in the basal layer of the epidermis and has a nuclear localization in single cells of the basal layer. Analysis of skin xenografts on days 75 and 110 as well as live skin equivalents obtained from primary keratinocytes showed the following pattern: YAP1 has a nuclear localization in most cells of the basal layer of the epidermis, while analysis of markers of the basal (K15, K19, Inta6) and suprabasal (K10) layers shows that keratinocytes with a basal phenotype form several layers above the basement membrane, unlike one layer in normal skin. The distribution pattern of YAP1 and markers of epidermal differentiation similar to that detected in xenograft and in skin equivalents, is observed in the case of cutaneous papilloma. These data indicate a correlation between an increase in the content of the activated form of YAP1 in the epidermis and an impairment of the normal process of keratinocyte differentiation, which may be associated with a violation of the polarization of keratinocytes or a change in the microenvironment. In a chronic wound, on the contrary, the content of YAP1 in the epidermis was low, the activated form of YAP was not detected, and there was also no expression of K10, which may be due to the inhibited differentiation process.

*Conclusion:* An increase in the content of activated YAP1 in the epidermis in the xenograft, skin equivalent and in the cutaneous papilloma may be associated with an active regeneration process associated with hypertrophy of basal layer of keratinocytes. A decrease in its content and the absence of an activated form in the epidermis of a chronic wound may indicate a slowdown in the regenerative process, which is accompanied by a violation of keratinocyte differentiation. Further study of YAP-signaling will allow us to analyze in more detail the signaling pathways that can be involved in the process of regeneration of human skin, and which can be therapeutically influenced to correct impairments of the regenerative process.

*Funding:* The work was performed with the financial support of a grant within the framework of the Russian Science Foundation, No. 21-74-30015.

*Acknowledgements:* We thank the Core Centrum of Koltzov Institute of Developmental Biology, Russian Academy of Sciences, for providing access to Keyence BZ-9000 and Leica DMI 6000 fluorescence microscopes and cryostat Leica CM1950.

## Specific allelic diversity of the *SLC26A4* gene associated with hearing loss in the indigenous peoples of Southern Siberia (Russia)

Danilchenko V.Yu.<sup>1,2\*</sup>, Zytsar M.V.<sup>1</sup>, Maslova E.A.<sup>1,2</sup>, Posukh O.L.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *danilchenko\_valeri@mail.ru*

**Key words:** hearing loss, gene *SLC26A4*, allelic diversity, indigenous peoples of Southern Siberia

*Motivation and Aim:* Mutations in the *SLC26A4* gene (MIM 605646, 7q22.3) are considered to be the second common cause of hereditary hearing loss (HL) in many world populations. The *SLC26A4* mutation spectrum and prevalence were found to be very diverse around the world. We aimed to analyze for the first time the *SLC26A4* allelic diversity in deaf patients belonging to indigenous Turkic-speaking Siberian peoples.

*Methods and Algorithms:* Analysis of the *SLC26A4* gene sequence including all 21 exons and flanking regions was performed by Sanger sequencing in ethnically matched cohorts of deaf patients (170 Tuvinians and 62 Altaians).

*Results:* Contrast differences in the proportion of the *SLC26A4*-related HL caused by pathogenic variants c.919-2A>G, c.2168A>G, c.2027T>A, c.170C>A, c.2034+1G>A, c.1545T>G were found in Tuvinian (28.2 %) and Altaian (4.3 %) patients, despite their similar ethnogenesis and close residence in Southern Siberia. Along with pathogenic variants, both already known SNPs (c.225C>G, c.1790T>C, c.165-626T>A, c.165-546T>C, c.165-119A>T, c.165-113C>T, c.601-266T>G, c.1001+131G>T, c.1150-135C>T, c.1264-187C>T, c.1341+47T>C, c.1545-85G>C, c.1707+398T>C, c.1708-18T>A, and c.2319+243T>C) and novel variants (NC\_000007.14: 107662833A>T, 107663118A>T, 107663208C>T, 107683478A>G, 107697874A>G, 107700976T>C, and 107701796C>T) were identified. We evaluated their frequencies in our samples and compared with available information from the Deafness Variation Database (<https://deafnessvariationdatabase.org/>). The frequencies of c.2319+243T>C (3.57 %), c.165-546T>C (0.40 %) in Tuvinians and c.225C>G (12.0 %), c.1545-85G>C (0.68 %), c.1707+398T>C (0.76 %) in Altaians significantly exceeded their world MAFs (0.04 %, 0.24, 0.09, 0.01, and 0.24 % respectively). The frequencies of c.601-266T>G (53.36 %), c.1150-135C>T (2.70 %), c.1341+47T>C (3.57 %) in Tuvinians, and c.1790T>C (2.27 %) in Altaians were comparable with their frequencies in East Asian populations (61.35, 3.30, 11.17, and 2.81 % respectively).

*Conclusion:* This work is the first study of the *SLC26A4* allelic diversity in Siberian indigenous populations. Analysis of the *SLC26A4* allelic frequencies in comparison with world populations revealed the affinity of Tuvinians and Altaians with East Asian populations. Obtained data significantly contribute to the information on the *SLC26A4* allelic diversity worldwide.

*Acknowledgements:* The study was supported by the RCF grant No. 21-75-00030.

## Preliminary bioinformatic analysis to optimize the testing of the *SLC26A4* pathogenic variants involved in hearing loss

Danilchenko V.Yu.<sup>1,2\*</sup>, Zytzar M.V.<sup>1</sup>, Maslova E.A.<sup>1,2</sup>, Posukh O.L.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [danilchenko\\_valeri@mail.ru](mailto:danilchenko_valeri@mail.ru)

**Key words:** hearing loss, gene *SLC26A4*, pathogenic variants, diagnostics

*Motivation and Aim:* Over 5 % of the world's human population affects by hearing loss (HL), and more than a half of all HL cases due to genetic causes. To date, 134 genes are causally implicated in nonsyndromic HL with different types of inheritance (Hereditary Hearing Loss Homepage: <https://hereditaryhearingloss.org>).

Pathogenic variants in the *SLC26A4* (7q22.3, OMIM 605646) gene are one of the common causes of hereditary HL in many populations. Mutations in the *SLC26A4* gene cause non-syndromic recessive deafness (DFNB4, OMIM 600791) and Pendred syndrome (PDS, OMIM 274600) which combines sensorineural HL and goiter. The *SLC26A4* gene encodes pendrin, an anion transporter protein, which is mostly expressed in tissues of inner ear, thyroid and kidneys. Numerous studies revealed a diverse spectrum of *SLC26A4* mutations and their varying prevalence in patients from different regions of the world. The region- or ethno-specific prevalence of the *SLC26A4* mutations has not yet been fully elucidated.

*SLC26A4* is relatively large gene covering about 57 kb of genomic DNA. The canonical transcript ENST00000644269.2 (ENSEMBL: <https://www.ensembl.org>), 4737 bp long, includes 20 coding exons within 2343-bp cDNA. Screening of pathogenic variants in the *SLC26A4* gene is an important part of molecular genetic testing for HL. Sequential analysis of the *SLC26A4* gene in a particular patient is continued until two recessive pathogenic *SLC26A4* variants are detected and, therefore, diagnosis could be made. A large size of *SLC26A4* predetermines the difficulties of its whole mutational analysis, which can be a labor-intensive, a time-consuming, and an expensive, especially in large patient cohorts.

*Methods and Algorithms:* Our study aims to optimize the testing of the *SLC26A4* gene by preliminary selection of the most diagnostically important regions of this gene for their priority analysis by direct sequencing.

To solve this task, we performed a thorough bioinformatic analysis of variations in the *SLC26A4* sequence based on the data from the Deafness Variation Database (DVD: <https://deafnessvariationdatabase.org/gene/SLC26A4>) [1]. The DVD, integrating all available genetic, genomic, and clinical data together with expert curation, generates the classification of variants: benign (B), likely benign (LB), likely pathogenic (LP), pathogenic (P), and variant of unknown significance (VUS) in many genes implicated in HL (including *SLC26A4*).

*Results:* To date, in total, 8647 different *SLC26A4* variants are reported in the DVD. The proportion of P and LP variants (PLP variants) is 7.0, 32.4, and 1.4 % in whole sequence, in coding and intronic regions of *SLC26A4* respectively. We selected all PLP variants within the *SLC26A4* coding and intronic regions ( $n = 603$ ), and analyzed their distribution across the *SLC26A4* sequence. Among PLP variants, the common molecular

alterations are the missense variants, followed by the variants affecting splicing, and the frameshift variants. The number of PLP variants across the coding exons of *SLC26A4* varies substantially, as well as the physical size of the exons. For selection of exons with the highest load of PLP variants, the number of these variants was normalized by exon size. Ten “hot” exons had the variation rates higher than a mean (median value = 21.39) of a number of PLP variants per an exon size. Among PLP variants located in the *SLC26A4* intronic regions, the variants recognized by the DVD as the splice donor variants, are the most frequent followed by the splice acceptor variants.

In order to identify the *SLC26A4* regions with the highest load of PLP variants, we combined the data on PLP variants located in coding exons with those in adjacent intronic regions and revealed ten *SLC26A4* regions with the highest diagnostic value (exons 4, 6, 10, 11, 13-17, 19 and their flanking regions). Mutational analysis of these *SLC26A4* sequence regions can provide the diagnostic rate of 61.9 % (373/603) of all currently known PLP variants in the *SLC26A4* sequence.

Observed uneven prevalence of PLP variants in the *SLC26A4* gene in different populations or geographical regions could be influenced by ethnicity of the studied patients, history of a population to which they belong, as well as by a variable sensitivity of diagnostics methods applied in a particular group of patients. The accumulation of specific PLP variants in certain populations can be the result of founder effect as it was evidenced for c.2168A>G (p.His723Arg) in Japanese and Koreans and c.919-2A>G in Chinese.

*Conclusion:* The bioinformatic analysis of the distribution of PLP variants across *SLC26A4* gene sequence revealed the most diagnostically important regions of the *SLC26A4* gene (exons 4, 6, 10, 11, 13–17, 19 with flanking intronic regions), sequencing of which can provide the diagnostic rate of 61.9 % of all currently known PLP variants in the *SLC26A4* sequence. This approach can be used as an initial effective diagnostic testing in samples of patients of unknown ethnicity, or as a subsequent step after targeted testing of already known ethno- or region-specific *SLC26A4* pathogenic variants.

*Acknowledgements:* This work was supported by the RSF grant No. 21-75-00030 (to M.V.Z., E.A.M., and V.Y.D.) and by the Ministry of Education and Science of Russian Federation (grant No. FSUS-2020-0040) (to O.L.P.).

## References

1. Azaiez H., Booth K.T., Ephraim S.S., Crone B., Black-Ziegelbein E.A., Marini R.J., Shearer A.E., Sloan-Heggen C.M., Kolbe D., Casavant T., Schnieders M.J., Nishimura C., Braun T., Smith R.J.H. Genomic landscape and mutational signatures of deafness-associated genes. *Am J Hum Genet.* 2018;103(4):484-497.

## Note on genomic ancestry and variation in disease susceptibility

Das R.<sup>1\*</sup>, Elistratova M.G.<sup>2</sup>, Orlov Y.L.<sup>2,3</sup>

<sup>1</sup> *Yenepoya Research Centre, Yenepoya University, Mangalore, India*

<sup>2</sup> *I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>3</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

\* *das.ranajit@gmail.com*

**Key words:** genetics, biomedicine, disease susceptibility, bioinformatics, e-Health

*Motivation and Aim:* Human DNA holds the key to uncovering our true ancestral past and determines who we are. A plethora of information stored inside the genome reflects our uniqueness and proximity to different ancestral and modern-day populations. Given the high correspondence between our genetic make-up and the geographical origin of our forefathers, it is possible to glean into precise ancestral origin using the genetic information [1, 9]. Understanding one's ancestry is not only a 'homing' tool bringing someone closer to human evolutionary past but also holds the key in determining population-specific variability in disease susceptibility, drug responsiveness, and other health and fitness related traits. Recently we have organized the Research Topic (special journal issue) at *Frontiers in Genetics* to collect the papers focused on the ancestry studies [4].

*Methods and Algorithms:* We have organized special journal issue (Research Topic) in "Frontiers in Genetics" considering genetic background and molecular mechanisms of the human diseases [5]. Understanding one's ancestry can play a monumental role in understanding variation in disease susceptibility across various populations and glean into the complex gene and environment interplay in ancestry-specific disorders [8]. Thus, traditional high fat and protein diets in cold regions of Siberia and North Asia have consequences on obesity and diabetes-related diseases [2]. Understanding one's genomic ancestry can facilitate the development of novel therapeutics or repurposing of existing treatment strategies, particularly aiding in identifying population-specific therapies. Subsequently, our knowledge of ancestry-specific variation in complex disorders can be used towards developing personalized and ancestry-specific precision medicine approaches to ameliorate several complex disorders [9].

*Results:* We have considered the following themes for this special issue:

- Unravel predisposition of various modern-day populations towards common disorders and conditions, including but not limited to cancers, heart diseases, and infectious diseases.
- The association of alleles with complex disorders, evaluated at a population level; discovery of novel disease marker panels.
- Identification of novel medically relevant genetic variants that can be used as diagnostic markers in genetic diagnostics and healthcare.
- Selection dynamics of the genes. Investigation of the spatial and temporal distributions of positively selected alleles in response to population specific disease susceptibility.

The papers published in the Research Topic correspond to the themes stated above and extend the studies presented in *Frontiers in Genetics Topics* (see

<https://www.frontiersin.org/research-topics/17947/bioinformatics-of-genome-regulation-volume-ii>). We had participated in the organization of series of conferences on human population genetics and computational genomics in Russia that allowed formalize the idea of special journal issues [6]. Thus, conference "Century of Human Population Genetics" was held in Moscow in 2019. It was focused on the discussion of the research on gene pools of the world's nations, ancient DNA analysis, possibilities of judicial genetics, population-genetic database development, biobanks, and new genomics technologies [7]. We had series of publications on human ancestry based on new genomics data initially discussed at this meeting [3].

**Conclusion:** Overall, the Research Topic at Frontiers in Genetics raised actual problems in the genetics studied. The Research Topics in Frontiers in genetics (<https://www.frontiersin.org/research-topics/21036/high-throughput-sequencing-based-investigation-of-chronic-disease-markers-and-mechanisms>) continues collection of papers on human diseases' markers.

**Acknowledgements:** This publication has been supported by the RUDN University Scientific Projects Grant System, project No. R.3-2022-ins.

### References

1. Babenko V., Chadaeva I., Orlov Y. Genomic landscape of CpG rich elements in human genome. *BMC Evol Biol.* 2017;17(Suppl 1):19. doi: 10.1186/s12862-016-0864-0.
2. Bai H., Liu H., Suyalatu S. et al. Association Analysis of Genetic Variants with Type 2 Diabetes in a Mongolian Population in China. *J Diabetes Res.* 2015;2015:613236. doi: 10.1155/2015/613236.
3. Das R., Ivanisenko V.A., Anashkina A.A., Upadhyai P. The story of the lost twins: decoding the genetic identities of the Kumhar and Kurcha populations from the Indian subcontinent. *BMC Genetics.* 2020;21(Suppl 1):117. doi: /10.1186/s12863-020-00919-2.
4. Das R., Tatarinova T.V., Galieva E.R., Orlov Y.L. Editorial: Association Between Individuals' Genomic Ancestry and Variation in Disease Susceptibility. *Front Genet.* 2022;13:831320. doi: 10.3389/fgene.2022.831320.
5. Orlov Y.L., Anashkina A.A., Tatarinova T.V., Baranova A.V. Editorial: Bioinformatics of Genome Regulation, Volume II. *Front Genet.* 2021;12:795257. doi: 10.3389/fgene.2021.795257.
6. Orlov Y.L., Baranova A.V., Markel A.L. Computational models in genetics at BGRS/SB-2016: introductory note. *BMC Genetics.* 2016;17(Suppl 3):155. doi: 10.1186/s12863-016-0465-3.
7. Tatarinova T.V., Tabikhanova L.E., Eslami G., Bai H., Orlov Y.L. Genetics research at the "Centenary of human population genetics" conference and SBB-2019. *BMC Genetics.* 2020;21(Suppl 1):109. doi: 10.1186/s12863-020-00906-7.
8. Tiis R.P., Osipova L.P., Galieva E.R., Lichman D.V., Voronina E.N., Melikhova A.V., Orlov Y.L., Filipenko M.L. N-acetyltransferase (NAT2) gene polymorphism and gene network analysis. *Biomeditsinskaya Khimiya.* 2021;67(3):213-221. doi: 10.18097/PBMC20216703213.
9. Zinchenko R.A., Ginter E.K., Marakhonov A.V. et al. Epidemiology of rare hereditary diseases in the european part of russia: point and cumulative prevalence. *Front Genet.* 2021;12:678957. doi: 10.3389/fgene.2021.678957.

## Hemoglobin-binding 2'-F-modified RNA aptamers: structure studies and analytical application

Davydova A.\*, Kabilov M., Lomzov A., Pyshnyi D., Vorobyeva M.

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

\* [anna.davydova@niboch.nsc.ru](mailto:anna.davydova@niboch.nsc.ru)

**Key words:** aptamer, G-quadruplex, human hemoglobin, colorimetry microplate assay

*Motivation and Aim:* DNA and RNA aptamers specifically recognize various molecular targets due to the formation of a specific three-dimensional structure. Such unique characteristics of aptamers as stability, ease of synthesis and modification as well as low immunogenicity and toxicity promoted the development of aptamer-based diagnostic and therapeutic tools. Particularly, aptamers are widely used as molecular recognizing elements for biosensors (also called aptasensors). Nowadays, there are plenty of electrochemical, fluorescent, colorimetric, and surface plasmon resonance aptasensors for the detection of a wide target spectrum [1]. Colorimetric detection systems are quite simple and easy to perform, provide reliable results, and suits well for point-of-care testing. Enzyme-linked immunosorbent assays (ELISA), which rely on plate reader colorimetry, are routinely used for laboratory diagnostics. Nevertheless, this method has its drawbacks determined by the use of antibodies: the high cost of diagnostic kits, their relatively short shelf-life, and lot-to-lot variability, which hampers the long-term studies [2]. The development of ELISA-like systems based on aptamers is a very promising approach to biomarker detection. In the present study, we aimed to select hemoglobin-binding 2'-F-modified RNA aptamers, analyze their structure and demonstrate their possible application as biorecognizing elements of microplate colorimetric assays.

*Methods and Algorithms:* We utilized SELEX method for *in vitro* selection of hemoglobin-binding aptamers using combinatorial library of 2'-fluoro pyrimidine modified RNAs. The final enriched library was sequenced using the NGS platform MiSeq (Illumina). A series of G-rich individual aptamers were chemically synthesized and screened for their binding activity against human hemoglobin. For detailed analysis of G-quadruplex formation, we employed CD spectroscopy, enzymatic structure probing and G-quadruplex specific fluorophore binding assays. We also performed enzyme-linked aptamer-based microplate assays both in direct and sandwich formats for hemoglobin detection.

*Results:* We selected guanine-rich 2'-fluoro pyrimidine modified RNA aptamers against human hemoglobin. Three 2'-F-RNA aptamers contained G-stretches, the analysis of their secondary structures by the Vienna algorithm revealed a possibility of G-quadruplex formation. The formation of parallel quadruplexes for these aptamers was confirmed by a combination of experimental assays. We established an aptamer-based colorimetric detection of human hemoglobin in microplate assays and demonstrated that quadruplex-forming 2'-F-RNA aptamers are suitable for this detection in direct and sandwich formats. We also revealed that the sandwich-type detection system based on two aptamers could generate a nonspecific colorimetric signal, probably due to complex formation between two aptamers used in the assay.

*Conclusion:* In the present study, we selected hemoglobin-specific G-rich modified RNA aptamers and confirmed the formation of G-quadruplex for these aptamers. We successfully used obtained aptamers for hemoglobin detection in microplate colorimetric assay.

*Acknowledgements:* The work was supported by the Russian state-funded project for ICBFM SB RAS (grant No. 121031300042-1).

### *References*

1. Feng F. et al. A computational solution to improve biomarker reproducibility during long-term projects. *PLoS One*. 2019;14(4):1-15.
2. Lashin S.A., Matushkin Yu.G. Haploid evolutionary constructor: new features and further challenges. *In Silico Biol*. 2012;11(3):125-135.
3. Zhang Y. et al. Recent advances in aptamer discovery and applications. *Molecules*. 2019;24(5):941.



## Functional study of rs2072580 (*SART3* and *ISCU* genes) associated with breast cancer

Degtyareva A.<sup>1\*</sup>, Antontseva E.<sup>1</sup>, Merkulova T.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *degtyareva\_rso@mail.ru*

**Key words:** rSNPs, *SART3*, *ISCU*, EMSA, breast cancer

*Motivation and Aim:* Single nucleotide polymorphism (SNP) rs2072580 T>A, located in the overlapping promoter regions of the *SART3* and *ISCU* genes, was identified by us as potentially regulatory (rSNP) using a bioinformatics approach developed in the laboratory for the analysis of genome-wide allele specific ChIP-seq and RNA-seq data [1]. An association of this polymorphism with an increased risk of developing breast cancer in the case of the A allele was also established [2]. The aim of this work is a functional analysis of the regulatory potential of rs2072580 to understand its significance in the molecular mechanisms of breast cancer development.

*Methods and Algorithms:* Electrophoretic mobility shift assay (EMSA) – method of DNA-probes retention in the gel by nuclear extract proteins was used to study the ability of this SNP to destroy or create binding sites for certain transcription factors (TFs). For its implementation, 31 bp double-stranded oligonucleotides V1 (T-allele) and V2 (A-allele) corresponding to allelic variants of the SNP location in the genome were used. Human cell lines HepG2, MCF7 were chosen as a source of nuclear extract proteins. Using the MotifbreakR software package, five TFs were selected, the binding sites of which, with a high probability, could change their affinity as a result of the rs2072580 T>A substitution. These are CREB1, AP1, FOXK1, PAX3 and REST. Oligonucleotides containing experimentally confirmed binding sites for these TFs were used in EMSA for competitive analysis. To evaluate the effect of the rs2072580 substitution on the expression of the Luc2 luciferase reporter gene, plasmid constructs containing a large promoter region (~1500 bp) both in the orientation of the *SART3* gene and in the orientation of the *ISCU* gene were created. Allelic variants of the promoter region were obtained by site-directed mutagenesis. The pGL3-Basic plasmid was used as a vector. The HepG2 cell line was used for transfection.

*Results:* The method of DNA-probes retention in gel with proteins of the nuclear extract of HepG2 cells showed that V1 forms more intense delay bands than V2, which was confirmed during cross-competition of these oligonucleotides, i.e. the substitution T>A leads to a decrease in the affinity of the given region for certain TFs. As a result of competitive analysis, it turned out that oligonucleotides containing the binding sites of CREB1, PAX3 and AP1 specifically compete with V1 and V2 for protein binding, since they attenuate certain bands on radioautograph. Moreover, the use of the CREB1-mut oligonucleotide as a competitor, with the destroyed site of this TF, did not lead to any change in the binding pattern. Thus, the presence of rs2072580 T>A affects the affinity of the transcription factors CREB1, PAX3, and AP1 for this region and, as a result, can determine the level of expression of nearby genes. It should be noted that, according to

the data of the ENCODE project, this SNP falls into the ChiP-seq peaks of the same TFs obtained on HepG2 and MCF7 cell lines.

The regulatory potential of rs2072580 in the *in vivo* system was studied in a reporter system on the HepG2 cell line. It revealed that the T>A substitution affects the expression of the Luc2 luciferase gene. Using data from the GTEx portal (<https://www.gtexportal.org/>), which contains information on the association of SNPs with the level of gene expression in postmortem samples of various human tissues, it was found that rs2072580 is associated with changes in the expression level of the *FICD* and *ISCU* genes in some tissues.

*Conclusion:* Thus, rs2072580 associated with breast cancer, localized in the common promoter region of the *SART3* and *ISCU* genes, has a regulatory potential and can affect the expression of target genes and finally to phenotype. In addition, the analysis of literature data revealed that *SART3* and *ISCU* are shown to be involved in carcinogenesis processes [3, 4].

*Acknowledgements:* This work was supported by the State Budget Project FWNR-2022-0016.

### References

1. Korbolina E.E. et al. Novel approach to functional SNPs discovery from genome-wide data reveals promising variants for colon cancer risk. *Hum Mutat.* 2018;39(6):851-854. doi: 10.1002/humu.23425.
2. Degtyareva A.O. et al. rs2072580T>A Polymorphism in the overlapping promoter regions of the *SART3* and *ISCU* genes associated with the risk of breast cancer. *Bull Exp Biol Med.* 2020;169(1):81-84.
3. Whitmill A. et al. Tip110: Physical properties, primary structure, and biological functions. *Life Sci.* 2016;149:79-95.
4. Ullmann P. et al. Hypoxia-responsive miR-210 promotes self-renewal capacity of colon tumor-initiating cells by repressing *ISCU* and by inducing lactate production. *Oncotarget.* 2016;7(40):65454-65470.

## **In search for the peptide/protein ligands of human serum albumin able to affect its interaction with amyloid $\beta$ peptide**

Deryusheva E.\*, Nemashkalova E., Shevelyova M., Permyakov S., Litus E.

*Institute for Biological Instrumentation, Pushchino Scientific Center for Biological Research, RAS, Pushchino, Moscow Region, Russia*

\**janed1986@ya.ru*

**Key words:** human serum albumin, amyloid  $\beta$  peptide, peptide/protein ligands

*Motivation and Aim:* Human serum albumin (HSA) is the most abundant protein in blood, comprising ca 60 % of its total protein content. HSA is a depot for amyloid  $\beta$  peptide (A $\beta$ ), one of the key factors in progression of Alzheimer's disease (AD). An imbalance between production and excretion of A $\beta$  in AD patients leads to A $\beta$  accumulation in the brain and formation of noxious amyloid oligomers/plaques. A promising approach to AD prevention is to reduce free A $\beta$  concentration by directed stimulation of the HSA-A $\beta$  interaction.

*Methods and Algorithms:* In this study, we systematically searched for the peptide/protein HSA ligands able to affect A $\beta$  interaction with HSA. The bioinformatic analysis of DrugBank, BioGRID and IntAct databases, as well as Alzforum online resource gave rise to a panel of 11 peptides and 34 proteins, potentially able to modulate the HSA-A $\beta$  interaction. As the candidate selection criteria, we used the values of molecular weight, subcellular location and strong association with AD, according to DisGeNET database and Open Targets Platform. The algorithm for search of the potential candidates and their in-depth analysis was implemented using the high-level programming language Python.

*Results:* The resulting protein candidates include apolipoprotein E, apolipoprotein C-III, cystatin-C, S100A8, serotransferrin, and others. The lead peptide candidates comprise vancomycin, enalapril, semaglutide, octreotide.

*Conclusion:* The most promising candidates are subject to experimental validation regarding their effect on the HSA-A $\beta$  interaction. They serve as a basis for further development of the first-in-class medications for prevention of AD onset.

*Acknowledgements:* Funded by Russian Science Foundation, grant No. 20-74-10072 (Litus E.).

## **Transcriptome of failing human heart reveals atrial myocytes reprogramming**

Deviatiarov R.<sup>1\*</sup>, Gusev O.<sup>1,2</sup>

<sup>1</sup> *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>2</sup> *Graduate School of Medicine, Juntendo University, Tokyo, Japan*

\* *rmdevyatiarov@kpfu.ru*

Heart diseases are the leading cause of death worldwide, however genome regulation behind it remains barely studied in human. In the way to fulfil lacking information we created an atlas of transcribed promoters and enhancers in healthy and failing adult human hearts. Total RNA was extracted from atriums and ventricles, left and right sides, and submitted to cap analysis of gene expression (CAGE) libraries protocol. In total 109 samples from 31 donors were sequenced and analyzed. We observed to general expression patterns in our data representing atrial and ventricular types respectively. While all failing ventricle heart samples keep ventricular expression profiles and showed activation of immune system processes, on the contrary failing atrium samples loosen immune responses and activated sarcomere, myofibril, and oxidation associated processes. Interestingly, most of failing atrium samples had ventricular type expression profiles, which might be explained by atrial-ventricular transdifferentiation. To confirm our observation, we reanalyzed a relevant subset of heartcellatlas.org single cell data (left/right, ventricle/atrium), and use it for deconvolution of bulk CAGE libraries. In atrium samples with ventricular type expression atrial myocytes were replaced with ventricular myocytes. Expression profiles of known cardiomyocytes marker genes like MYL2 and MYL7 confirms such reorganization, but from another side reprogrammed atrium samples could be clearly distinguished from other ventricle samples by NPPA gene activity, which is significantly higher in atrial samples.

*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2021-601).

# Research and analysis of genes associated with hereditary predisposition to Crohn's disease and reconstruction of the gene network using bioinformatics tools

Efremova A.D.<sup>1\*</sup>, Orlov Y.L.<sup>1,2</sup>

<sup>1</sup> I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* [nastya.efremova.2000@gmail.com](mailto:nastya.efremova.2000@gmail.com)

**Key words:** medical genetics, Crohn's disease, gene network, gene ontologies, education, bioinformatics

*Motivation and Aim:* Genomic association studies and computerized meta-analyses have identified and confirmed 71 loci of susceptibility to Crohn's disease on 17 chromosomes [1]. Understanding the role of the innate immune response in Crohn's disease was the result of the discovery of the corresponding susceptibility loci in genetic studies, which still explains only a little more than 20 % of the heritability of Crohn's disease [2]. In this work, we compiled a list of genes associated with the development of Crohn's disease, analyzed the categories of gene ontologies for such a list and reconstructed the gene network. For the key genes of the disease obtained by analyzing the structure of the gene network, options for the search for drugs interacting with the studied proteins could be considered.

*Methods and Algorithms:* To build a list of genes associated with hereditary predisposition to Crohn's disease and to analyze the disease genes, the online resource OMIM (Online Mendelian Inheritance in Man, <https://omim.org>) was used [3]. To search and analyze significant categories of gene ontologies for this group of genes, the resulting list of human genes was downloaded via the DAVID interface (Database for Annotation, Visualization and Integrated Discovery) (<https://david.ncifcrf.gov/summary.jsp>) and PANTHER [4]. The categories of gene ontologies for molecular functions were also calculated. GeneMANIA (<https://genemania.org>) and STRING-DB (<https://string-db.org/>) tools were used to reconstruct the gene network of interactions of genes associated with Crohn's disease. We had follow scheme of applications of online tools recently published in [5] for glioblastoma analysis.

*Results:* The GO analysis revealed that the most significant categories for genes associated with Crohn's disease are signaling pathways mediated by tumor necrosis factor and caused by cytokines, as well as reactions to the drug and cytosolic processes. DAVID and PANTHER for Crohn's disease genes confirm the categories of gene ontologies of protein binding, membranes and intracellular processes. The most significant categories of gene ontologies for molecular functions are the categories of protein binding, binding of enzymes and cytokines, and other binding of protein-containing complexes. The usage of online bioinformatics tools [6] proved be reliable approach for Crohn's disease genes analysis. The presented bioinformatics techniques were discussed at the master-classes and students workshops at I.M. Sechenov First Moscow State Medical University.

*Conclusion:* Due to the identification of susceptibility loci, it is possible to conclude about violations of immune regulation in the intestine, as well as the presence of categories of cytosolic processes and signaling pathways allows us to determine the relationship of the disease with intracellular inflammation. The resulting gene network is quite coherent, although the connections were exposed only by the parameters protein-protein interactions and co-expression.

Current treatments are ineffective at counteracting disease progression. Overall, search for novel drug targets in Crohn's disease remains challenging problem [7].

*Acknowledgements:* The study is supported by the Potanin Foundation grant for masters' teachers (ГК22-000797).

### *References*

1. Khor B., Gardet A., Xavier R.J. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011;474:307-317.
2. Baumgart D.C., Sandborn W.J. Crohn's disease. *Lancet*. 2012;380:1590-605.
3. Amberger J.S., Bocchini C.A., Scott A.F., Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res*. 2019;47(D1):D1038-D1043. doi: 10.1093/nar/gky1151.
4. Mi H., Muruganujan A., Thomas P.D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013;41(Database issue):D377-D386.
5. Gubanova N.V., Orlova N.G., Dergilev A.I., Oparina N.Y., Orlov Y.L. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools. *J Integr Bioinform*. 2021;18:20210031. doi: 10.1515/jib-2021-0031.

## Effect of DNA-hydrolyzing catalytic IgGs from schizophrenia patients on cell viability of the SH-SY5Y cell line

Epimakhova E.V.<sup>1\*</sup>, Ermakov E.A.<sup>2</sup>, Kazantseva D.V.<sup>3</sup>, Vasilyeva A.R.<sup>3</sup>,  
Ivanova S.A.<sup>1</sup>, Smirnova L.P.<sup>1</sup>

<sup>1</sup> *Mental Health Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

<sup>2</sup> *Institute of Chemical Biology and Fundamental Medicine, Laboratory of Repair Enzymes, RAS, Novosibirsk, Russia*

<sup>3</sup> *Siberian State Medical University, Tomsk, Russia*

\* *ElenaEpimakhova@mail.ru*

**Key words:** Schizophrenia, DNA-hydrolyzing catalytic IgGs, neuroblastoma, SH-SY5Y, cell viability

*Motivation and Aim:* DNA-hydrolyzing catalytic IgGs have cytotoxic effects in autoimmune diseases [1, 2]. Such IgG antibodies penetrate the cell membrane, affect intracellular processes and activate cell death processes [3]. Recently, DNA-hydrolyzing IgGs have been discovered in schizophrenia [4]. However, their cytotoxic properties have not been studied. The aim of this study was to investigate the effects of IgG isolated from the serum of schizophrenic patients with DNA-hydrolyzing activity on cell viability of the SH-SY5Y human neuroblastoma cell line *in vitro*.

*Methods:* Serum of patients with paranoid schizophrenia (F 20.00, F 20.01, F20.02) in the acute phase and mentally and somatically healthy persons matched by sex and age was used. IgG was purified from serum by affinity chromatography on Protein-G-Sepharose columns. The homogeneity of the preparations was confirmed by 12.5 % PAGE. The DNA hydrolyzing activity of IgG was assessed by the degree of hydrolysis of the pBluescript plasmid. The cell viability of the SH-SY5Y cell line after exposure to purified IgG preparations was assessed by high-throughput screening on the CellInsight CX7 platform (Thermo Scientific, USA) using the fluorescent dyes propidium iodide and Hoechst. The final antibody concentration was 0.1 and 0.2 mg/ml. Statistical analysis was performed using the SPSS software, release 20.0, for Windows. Median, first and third quartiles were calculated. Between-group differences were evaluated using Mann–Whitney U test.

*Results:* Four of the eight obtained IgG preparation had high DNA-hydrolyzing activity. All tested IgG preparations from healthy donors were inactive. The percentage of dead cells SH-SY5Y after exposure to IgG (0.1 mg/ml) with DNA-hydrolyzing activity after 24 hours was 4.5 [3.9; 5.1] %. When IgG patients without activity was added to the cell culture, the percentage of dead cells was 4.5 [4.3; 5.1] %. When IgG of healthy donors was added to the cell culture, this indicator was 4.9 [4.7; 5.4] %. One-way ANOVA analysis showed no significant differences in the percentage of dead cells between groups. Similar results were obtained at a higher concentration of antibodies –0.2 mg/ml.

*Conclusion:* Thus, it has been shown *in vitro* that IgGs isolated from the serum of schizophrenia patients with or without DNA-hydrolyzing activity does not exhibit cytotoxic properties against the SH-SY5Y human neuroblastoma cell line. It is possible that the absence of the cytotoxic effect of DNA-hydrolyzing IgGs is due to the high level of protective properties of neuroblastoma [5].

*Acknowledgements:* This work was supported by the Russian Science Foundation, grant No. 18-15-00053P (2021-2022). Cell viability was assessed using the equipment of the on the base The Core Facility “Medical genomics” of the Tomsk National Medical Research Center.

### *References*

1. Sabirzyanova A.Z., Nevzorova T.A. Cytotoxicity and genotoxicity of antibodies to native deoxyribonucleic acid in systemic lupus erythematosus and rheumatoid arthritis. *RJI*. 2013;7(4):428-436.
2. Nevinsky G.A. Catalytic antibodies in norm and systemic lupus erythematosus. In: Khan W.A. (Ed.). *Lupus* [Internet]. London: IntechOpen, 2017;41-101. doi: 10.5772/67790.
3. Rivadeneyra-Espinoza L., Ruiz-Argelles A. Cell penetrating anti-native DNA antibodies trigger apoptosis through both the neglect and programmed pathways. *J Autoimmun*. 2006;26(1):52-56. doi: 10.1016/j.jaut.2005.10.008.
4. Ermakov E.A., Smirnova L.P., Parkhomenko T.A., Dmitrenok P.S., Krotchenko N.M., Fattakhov N.S., Bokhan N.A., Semke A.V., Ivanova S.A., Buneva V.N., Nevinsky G.A. DNA-hydrolysing activity of IgG antibodies from the sera of patients with schizophrenia. *Open Biol*. 2015;5(9):150064. doi: 10.1098/rsob.150064.
5. Kondakova I.V., Kakurina G.V., Smirnova L.P., Borunov E.V. Regulation of proliferation and apoptosis of tumor cells by free radicals. *Sib J Oncol*. 2005;1:58-61.



## Multiplex analysis of serum cytokine profiles in patients with multiple sclerosis and systemic lupus erythematosus

Ermakov E.<sup>1\*</sup>, Melamud M.<sup>1</sup>, Boiko A.<sup>2</sup>, Parshukova D.<sup>2</sup>, Sizikov A.<sup>3</sup>, Ivanova S.<sup>2</sup>, Nevinsky G.<sup>1</sup>, Buneva V.<sup>1</sup>

<sup>1</sup>*Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Mental Health Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

<sup>3</sup>*Institute of Clinical Immunology, SB RAS, Novosibirsk, Russia*

\* [evgeny\\_ermakov@mail.ru](mailto:evgeny_ermakov@mail.ru)

**Key words:** cytokine profiles, serum, multiple sclerosis, systemic lupus erythematosus

*Motivation and Aim:* Dysregulation of cytokine networks is known to be associated with various autoimmune diseases including multiple sclerosis (MS) and systemic lupus erythematosus (SLE) [1]. Many studies have shown changes in serum cytokine levels in MS and SLE [2, 3]. However, the detailed changes in serum cytokine profiles in MS and SLE are poorly understood. Besides, comparisons of cytokine profiles in these diseases have not been performed. The aim of this work was to investigate changes in the serum cytokine profiles in MS and SLE in comparison with healthy donors in one series of experiments.

*Methods and Algorithms:* The analysis was carried out in the following groups: a group of SLE patients – 60 people, a group of MS patients – 55 people, a group of healthy donors – 36 people. The concentration of 41 cytokines was determined using a MAGPIX multiplex analyzer (Luminex, USA) using the MILLIPLEX MAP Human Cytokine/Chemokine Magnetic Bead Panel – Premixed 41 Plex – Immunology Multiplex Assay (Merck Millipore, USA). Statistical analysis (Kruskal-Wallis ANOVA, Dunn's test, Spearman correlation analysis, Bonferroni correction for multiple hypothesis testing) was carried out in STATISTICA 10 and ORIGIN 2021 programs. Volcano plots that reflect the dependence of the Log<sub>2</sub> fold changes of the median values on the -Log<sub>10</sub> p-value were plotted. Analysis of changes in cytokine profiles was performed by principal component analysis and partial discriminant least squares analysis using the CytokineExplore online resource (<http://al-saleh.cc/exabx.com/apps/cytokineexplore/>).

*Results:* Principal component analysis and discriminant least squares analysis showed that the combined multicytokine profiles of the studied groups differed. The following cytokines changed significantly in the group of healthy individuals compared with the group of SLE patients (in decreasing order of significance): IL-15, IL-1RA, IL-6, TNF $\alpha$ , Fractalkine, IL-9, MDC, sCD40L, IL-10, IL-1b, PDGF-AA, MCP-1, MIP-1a, MIP-1b, G-CSF, GM-CSF, IL-1a. Most of these cytokines were higher in SLE patients. The concentrations of the following cytokines changed significantly in healthy donors compared with MS patients (in decreasing order of significance): sCD40L, EGF, PDGF-AA, TGF- $\alpha$ , PDGF-AB/BB, IL-8, VEGF, MIP-1b, MCP-1, MDC, Eotaxin, GRO, TNF $\beta$ , IL-7, IP-10. Most of these cytokines were lower in MS patients. Interestingly, IFN $\alpha$ 2 and PDGF-AB/BB levels were significantly different in patients with primary progressive MS compared with patients with relapsing-remitting MS ( $p < 0.03$ ). The following cytokines changed significantly in the group of MS patients compared with the group of

SLE patients (in decreasing order of significance): IL-8, EGF, TGF- $\alpha$ , MIP-1b, sCD40L, MCP-1, VEGF, PDGF-AB/BB, PDGF-AA, IL-1RA, TNF $\alpha$ , MIP-1a, IL-15, IL-7, Eotaxin, IFN $\gamma$ , GRO, IL-6, IL-1b, IL-17A, G-CSF, IL-10, GM-CSF, Fractalkine, IL-12P70, IL-3, IL-2, RANTES, IL-9, IL-4, IFN $\alpha$ 2. Thus, the levels of the 31st cytokine (out of 41 analyzed) differed in SLE and MS. Most of these cytokines were higher in SLE patients.

*Conclusion:* Thus, cytokine profiles in SLE and MS differ significantly both in comparison with healthy individuals and between these diseases. These data will help to better understand the differences in cytokine network dysregulation in MS and SLE.

*Acknowledgements:* This work was supported by the Russian Science Foundation under grant number 20-15-00162.

### *References*

1. O'Shea J.J., Ma A., Lipsky P. Cytokines and autoimmunity. *Nat Rev Immunol.* 2002;2(1):37-45.
2. Kallaur A.P. et al. Cytokine profile in patients with progressive multiple sclerosis and its association with disease progression and disability. *Mol Neurobiol.* 2017;54(4):2950-2960.
3. Melamud M.M. et al. Serum cytokine levels of systemic lupus erythematosus patients in the presence of concomitant cardiovascular diseases. *Endocr Metab Immune Disord Drug Targets.* 2022. Online ahead of print. doi: 10.2174/1871530322666220304214512.

## ***Toxoplasma gondii* proteins: development of bioinformatics analysis for vaccine design against toxoplasmosis**

Eslami G.<sup>1\*</sup>, Mehrani M.J.B.<sup>2</sup>, Lesani S.<sup>3</sup>, Askari V.<sup>4</sup>

<sup>1</sup> Department of Parasitology and Mycology, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

<sup>2</sup> Department of Parasitology and Mycology, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

<sup>3</sup> Department of Parasitology and Mycology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

<sup>4</sup> Research Center for Food Hygiene and Safety, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

\* [eslami\\_g2000@yahoo.com](mailto:eslami_g2000@yahoo.com)

**Key words:** *Toxoplasma gondii*, Vaccine, toxoplasmosis, Epitope

**Motivation and Aim:** Toxoplasmosis is one of the ubiquitous zoonotic diseases in the world, caused by intracellular parasite *Toxoplasma gondii*. This parasite causes infection in humans described as congenital and or postnatal transmission. Due to the lack of appropriate vaccines and drug to prevent toxoplasmosis, this study intends to develop bioinformatics screening of epitope-based vaccine.

**Methods and Algorithms:** The protein sequences of *T. gondii* candidate antigens were retrieved in FASTA format via a high-quality web server of protein sequence called Vaxquery, available at <http://www.violinet.org/vaxquery/> for further analysis. Antigens have chosen for this study included; MIC3, GRA5, p30, ROP8, ROP2, SAG2, MIC13, hsp70, and OMPDC. The immunogenicity of the selected antigens were predicted using VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). Non-toxic epitopes were chosen and further analysis was done with antigen that had the highest immunogenicity and lack allergenicity. ABCpred online server ([https://webs.iitd.edu.in/raghava/abcpred/ABC\\_submission.htm](https://webs.iitd.edu.in/raghava/abcpred/ABC_submission.htm)) was employed to predict the 16 amino acids linear B cell epitopes with a specificity threshold of 0.75 %. We used overlaps filter. This server recruits a combination of two approaches (subsequence kernel and support vector machine) by 74.57 % performance accuracy. Epitopes with e highest immunogenicity and lack of allergenicity and toxicity are considered for further study. Prediction of potential epitopes of 9-mer length for MHC I was provided using Immune Epitope Database (IEDB) accessible at <https://www.iedb.org>. The screening of antigenicity and allergenicity performed via VaxiJen v2.0 and AllerTOP v2.0 servers. Epitopes with highest immunogenicity, non-allergenicity and non- toxicity, which are high-binder epitopes were selected to design vaccine. Chimer protein design was accomplished for the construction of multi-epitope vaccine. The screening of toxicity and allergenicity of chimer performed via ToxinPred and AllerTOP v2.0 servers respectively. The secondary structure of multi-epitope vaccine, PRABI online server at [https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_server.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html) and also, Garnier-Osguthorpe-Robson (GOR IV) method were utilized. SWISS-MODEL online service (<https://swissmodel.expasy.org/>) was employed for tertiary or three-dimensional (3D) structure of designed vaccine and a

homology modeling technique was recruited. Molecular docking method was performed to assess the binding high affinity of target vaccine construct to respective MHC molecules. The strong binding affinity of the target vaccine towards MHC was validated by these scores.

*Results:* All the selected conserved protein sequences were subjected to VaxiJen v. 2.0 for prediction of immunogenicity probabilities. Allergenicity prediction carried out by AllerTOP v. 2.0 server and the findings revealed that seven chosen proteins include, GRA5, ROP8, Rop2, MIC13, SOD, hsp70, OMPDC have non-allergic nature and the assessment of toxicity with ToxinPred online server showed that GRA5 had no toxicity and was thus selected for the final vaccine construct. According to the results, one 16-mer potent epitope (GVAGSTRDTGSGGDDS) with a cut-off binding score > 0.9, qualified to be included in the final vaccine sequence. Out of all predicted epitopes, "RDTGSGGDD" was chosen as a target based on the results of epitope screening by using VaxiJen v2.0, AllerTOP v2.0, and ToxinPred servers. The constructed vaccine sequence (274 amino acid residues) validated for immunogenicity, allergenicity, and toxicity and fulfilling all the criteria. The overall investigations revealed that the designed vaccine had appropriate criteria for production.

*Conclusion:* As current vaccination against *T. gondii* infection is not satisfactory, in the present study, a novel epitope-based vaccine was designed and we employed several bioinformatics tools. GRA4 which express in two infectious stages (tachyzoite and bradyzoite) with strong immunogenicity and non-allergenic nature was selected to obtain better result.

## Profiling of targeted miRNAs (8-mer) for the genes involved in type 2 diabetes mellitus and cardiac hypertrophy

Hussain K.\*, Ishtiaq A., Mushtaq I., Khan M., Karim S., Gul R., Azhar I., Murtaza I.\*

*Signal transduction laboratory, Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan*

\* [khadam@bs.qau.edu.pk](mailto:khadam@bs.qau.edu.pk), [irambch@qau.edu.pk](mailto:irambch@qau.edu.pk)

Type 2 diabetes mellitus and cardiac hypertrophy are among the top ten leading cause of deaths, worldwide. Type 2 diabetes mellitus and cardiac hypertrophy are the chronic diseases, have close association and direct life-threatening implications like stroke, myocardial infarction, retinopathy, nephropathy and limb amputation. In addition to other medical approaches, microRNAs -based strategy is considered a game-changer for early detection of these chronic diseases and also has potential for the treatment of Type 2 diabetes mellitus and cardiac hypertrophy like it is being used for cancer in clinical trials. MicroRNAs are the non-coding single stranded, almost 20 to 22 nucleotides sequences which upon the complimentary basis bind to their target mRNA to silencing the protein expression at post transcriptional level. Bioinformatic databases are used like gene testing registry, online mendelian inheritance in man, TargetScan and ShinyGO for validation of disease linked genes and sorting the common microRNAs in both diseases, such as *miR-30-5p/101-3p.2/190-5p/506-3p/9-5p/128-3p/137/96-5p/7-5p/107/101-3p.1, let-7-5p, miR-98-5p/124-3p.2/124-3p.1/15-5p/16-5p/195-5p/424-5p/497-5p/1271-5p*. Aforementioned databases also used for the microRNAs that have more than one disease linked genes target in each pathological condition. Such microRNAs for cardiac hypertrophy are: *miR-19-3p/183-5p.2/153-3p/302-3p/372-3p/373-3p/520-3p/129-5p/144-3p/139-5p* and for Type 2 diabetes mellitus are: *miR-1-3p/206/27-3p/181-5p*. These findings will be of merit after future validation by using their expression analysis, mimic/anti-microRNA approach to check their potential against cardiac hypertrophy and Type 2 diabetes mellitus.

## ANDDigest: An AI-enhanced web-tool for finding information in the biological literature

Ivanisenko T.V.<sup>1,2\*</sup>, Demenkov P.S.<sup>1,2</sup>, Kolchanov N.A.<sup>1,3</sup> Ivanisenko V.A.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [itv@bionet.nsc.ru](mailto:itv@bionet.nsc.ru)

**Key words:** text-mining, ANDDigest, ANDSySystem, knowledge retrieval, machine learning, neural network

*Motivation and Aim:* Finding relevant information is an extremely important issue in almost any scientific research. Well-known scientific search engines such as Google Scholar, Scopus, PubMed, etc. [1–3] allow finding relevant documents based on user-defined keywords, but do not provide the user with tools for the automated extraction and analysis of information from search results, which can reach thousands of documents. Also, such queries do not take into the account the synonyms for the objects of interest and their linkage with external databases.

The vast amounts of published scientific literature made the task of searching for the information relevant to the field of the researcher's study one of the most important problems. Well-known search engines, such as Google Scholar, Web of Science, PubMed, Scopus, and others, are highly effective in finding documents; however, they do not allow the use of an ontology of the subject area [1–3]. Programs based on automatic text analysis provide the extraction of knowledge and its presentation in the form of semantic networks [4]. Of particular interest among them are systems providing a full cycle of knowledge engineering, including their automatic extraction, integration, presentation in the form of semantic networks, visualization, and analysis. Examples of such software products are STRING [5], Pathway Studio [6], MetaCore [7, 8], ANDSySystem [9–12], etc.

Earlier, we developed the ANDDigest tool [13] designed for the finding of the information from the textual sources based on the ANDSySystem ontology, presented in the form of the dictionaries. However, due to the use of the dictionary-based approach one of the main drawbacks of this system, was a high number of the false positive recognitions of the molecular-genetics objects in texts. The most of such errors were associated with incorrect recognition of the short names (5 or less symbols).

*Methods and Algorithms:* To deal with a described problem we used a fine-tuning strategy [14] to the pretrained BioBERT transformer model [15], previously trained on the large number of the PubMed abstracts and full-text articles. The fine-tuned models were used as binary classifiers for the 5 groups of objects from the ontology of ANDSySystem. Each classifier was trained on the separated set of abstracts, generated from the mapped data, where only the long names of objects were considered.

*Results:* The developed new version of the ANDDigest system designed for the intelligent search, analysis, and graphical representation of the information in the scientific literature. The system was improved with five deep learning classification models, providing the automated context-based filtration of the short names in natural language texts.

The models were created for the classification of short names of objects of the following types: genes/proteins, metabolites/drugs, cell components, diseases/side effects, and pathways. For the assessment of the quality of each model, a verification dataset was created. Each of the sets consisted of about 15,000 of sentences, containing short names of objects from the described group, mapped by the dictionary. Each object was manually assigned as positively or negatively recognized. The AUC values appeared to be the following: diseases/side effects classification model – 0.96; genes/proteins classification model – 0.94; cell components model – 0.91; drugs/metabolites model – 0.91; pathways model – 0.88.

The new version of ANDDigest is available by <http://anddigest.sysbio.ru>

*Acknowledgements:* Work 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.

### References

1. Beel J., Gipp B. Google Scholar's Ranking Algorithm: An Introductory Overview. In: In Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09). 2009;1:230-241.
2. McEntyre J., Ostell J. The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US). 2002.
3. Jacso P. As we may search - Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Curr Sci.* 2005;89:1537-1547.
4. McEntyre J., Ostell J. The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US). 2013.
5. Szklarczyk D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45(D1):D362-D368.
6. Nikitin A. et al. Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics.* 2003;19(16):2155-2157.
7. Nikolsky Y. et al. Biological networks and analysis of experimental data in drug discovery. *Drug Discov Today.* 2005;10(9):653-62.
8. Ekins S. et al. Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica.* 2006;36(10-11):877-901.
9. Demenkov P.S. et al. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSytem. *In Silico Biol.* 2012;11(3, 4):149-161.
10. Ivanisenko V.A. et al. ANDSytem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol.* 2015;9(S2):S2.
11. Ivanisenko V.A. et al. A new version of the ANDSytem 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.
12. Saik O.V. et al. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSytem. *BMC Med Genomics.* 2019;12(2):47.
13. Ivanisenko T.V. et al. ANDDigest: a new web-based module of ANDSytem for the search of knowledge in the scientific literature. *BMC Bioinformatics.* 2020;21(11):1-21.
14. Ji Z. et al. (2020) Bert-based ranking for biomedical entity normalization. In: AMIA Summits on Translational Science Proceedings. 2020:269-277.
15. Lee J. et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* 2020;36(4):1234-1240.

# Phylostratigraphic analysis of human cancers transcriptomic data

Ivanov R.<sup>1\*</sup>, Mustafin Z.<sup>1,2,3</sup>, Lashin S.<sup>1,2,3</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [ivanovromanart@bionet.nsc.ru](mailto:ivanovromanart@bionet.nsc.ru)

**Key words:** cancer, transcriptomics, phylostratigraphy

*Motivation and Aim:* The emergence and evolution of human cancer is a topical issue of modern biology. One of the promising modern approaches to the evolutionary analysis of genes is phylostratigraphic analysis. It allows to determine the important stages of genome evolution, at which there was a rapid increase in the number of new genes, to identify genes specific to certain organisms [1]. In particular, with the help of phylotranscriptome analysis, which correlates phylostratigraphic indices with the level of gene expression, a pattern was revealed between the level of gene expression and their age in the process of embryogenesis, both in plants and animals [2, 3]. In the course of our work, we set ourselves the task to analyze the available transcriptomic data of various types of cancer using the phylotranscriptomic method.

*Methods and Algorithms:* Raw expression data from tumor RNA-seq were processed using STAR and SAMtools. Phylostratigraphic and phylotranscriptomic analysis of PAI (Phylostratigraphic Age Index), DI (Divergence Index), TAI (Transcriptome Age Index) and TDI (Transcriptome Divergence Index) indices were carried out with Orthoweb web service.

*Results:* In our study we decided to analyze data of lung, uterus, prostate, thyroid, kidney and colon adenocarcinomas, breast ductal cancer and glioma. We obtained several gene expression datasets of these tumor tissues from open-sources, mainly TCGA and GEO NCBI. Raw counts were normalized by calculating TPM and FPKM. Phylostratigraphic indices shows the enrichment peak of genes formed early in the evolution of Eukaryote and the peak at the age of Vertebrata occurrence. The evolutionary indices data were also overlaid on pathway networks obtained from KEGG databases.

*Acknowledgements:* The study was supported by the budgetary projects of the Institute Cytology and Genetics of SB RAS (numbers of State registration: AAAA-A19-119101090031-8, 121031300209-8).

## References

1. Domazet-Lošo T., Brajković J., Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 2007;23:533-539. doi: 10.1016/j.tig.2007.08.014.
2. Quint M., Drost H.-G., Gabel A., Ullrich K.K., Bönn M., Grosse I. A transcriptomic hourglass in plant embryogenesis. *Nature.* 2012;490:98-101. doi: 10.1038/nature11394.
3. Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 2010;468:815-819. doi: 10.1038/nature09632.



## N-Nitroso-BODIPY derivatives as effective light activated NO donors

Karogodina T.Yu.<sup>1\*</sup>, Panfilov M.A.<sup>2</sup>, Tretyakova I.S.<sup>3,4</sup>, Vorob'ev A.Yu.<sup>2</sup>, Moskalensky A.E.<sup>1</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia

<sup>3</sup> Voevodsky Institute of Chemical Kinetics and Combustion, SB RAS, Novosibirsk, Russia

<sup>4</sup> Institute of Solid State Chemistry and Mechanochemistry, SB RAS, Novosibirsk, Russia

\* karogodina@gmail.com

**Key words:** nitric oxide, light-controllable NO releasers, light removal group

*Motivation and Aim:* Nitric oxide (NO) is a unique biochemical mediator involved in the regulation of a huge number of vital processes. NO level is critical for the regulation of vascular and muscle tone. Lack of nitric oxide can cause serious disorders. Compensation for the lack of NO by taking medications has several disadvantages. A promising alternative to this approach would be to use medications that can release nitric oxide directly in the area needed. Such technique can be based on the photolabile compounds that release NO under illumination by light. Such molecules usually consist of a light-absorbing antenna (a chromophore) attached to an NO-carrying group. The purpose of this work is the design, synthesis, and study of such compounds.

*Methods and Algorithms:* In this work we present a family of novel green-light-controllable NO releasers. We designed and synthesized several compounds using BODIPY dye as a light-harvesting antenna linked to an N-Nitroso NO-carrying group.

*Results:* The compounds was shown to be light sensitive. We have studied the photoproducts which were formed under green light irradiation. Using the DAR-2 fluorescent NO probe, we observed the photoinduced formation of NO. We also studied the generation of singlet oxygen which can be useful for photodynamic therapy.

*Conclusion:* The presented N-Nitroso-BODIPY derivatives are promising as green-light-controllable NO releasers. These compounds could serve as the basis for the development of more advanced photodonors and their use on test biochemical objects.

*Acknowledgements:* The study was supported by the Russian Science Foundation (grant No. 18-15-00049).

## ***In silico* determination of the risk haplotype for developing AMD**

Karpova N.

*Institute of General Pathology and Pathophysiology, Moscow, Russia*

\* *nataliakarpova.sp@gmail.com*

**Key words:** ARMS2, AMD, age-related diseases, pathology of eyes

*Motivation and Aim:* Age-related macular degeneration (AMD) is a multifactorial disease and a prevalent cause of visual impairment in developed countries. Risk factors include environmental components and genetic determinants. 50 % or more of the risk of developing AMD is due to mutations in the region of 2 genes, *CFH* and *ARMS2/HTRA1*. However, the exact mechanisms of pathology development, in particular the role of mutations in the *ARMS2* gene region, remain not fully understood, as well as the functions of the protein [1]. A number of polymorphisms related to *ARMS2* associated with the risk of developing AMD were identified in several GWASs at once [2–4]. In this regard, we decided to determine the putative AMD risk haplotype, its prevalence, and compare it with the prevalence of early and late stages of AMD.

*Methods and Algorithms:* We used the GWAS catalog and PubMed to identify SNPs in *ARMS2* associated with AMD, followed by mapping these SNPs to the human genome in the UCSC genome browser to identify overlapping polymorphisms in genes and regulatory regions [5,6]. We selected SNPs with a MAF of 1 % in the *ARMS2* gene region and identified possible haplotypes in European populations, followed by comparison of the putative risk haplotype with the prevalence of early and late AMD [7].

*Results:* In the GWAS-catalog, we found 11 polymorphisms in the *ARMS2* region: rs3750846, rs10490924, rs3750847, rs3750848, rs370974631, rs10490923, rs36212732, rs61871744, rs72834437, rs222308. An analysis of the literature made it possible to reduce the number of studied polymorphisms to 5 (rs3750846, rs10490924, rs3750847, rs3750848, rs61871744), since only they are associated with AMD and are located in the region of the gene itself.

When assessing the prevalence of possible haplotypes, a possible AMD risk haplotype was determined, for which the prevalence was 0,1899 or 18,99 % (taking into account the AMD risk alleles, rs10490924-T and rs3750847-T) [8] (Fig. 1).

*Conclusion:* As a result, we have *in silico* determined a possible AMD risk haplotype: rs61871744-C, rs10490924-T, rs3750848-G, rs3750847-T, rs3750846-C (C\_T\_G\_T\_C). Based on the data obtained, it can be assumed that the C\_T\_G\_T\_C haplotype in homozygous condition is a risk factor for early AMD, but it remains unclear what triggers the development of the disease. And in the case of heterozygotes, there must be a second risk factor that will increase the detrimental effect of the risk haplotype. The resulting hypothesis will then be tested on clinical samples obtained in a case-control study.

RS Number	Position (GRCh37)	Allele Frequencies	Haplotypes	
rs61871744	chr10:124203787	T=0.808, C=0.192	T	C
rs10490924	chr10:124214448	G=0.805, T=0.195	G	T
rs3750848	chr10:124215315	T=0.805, G=0.195	T	G
rs3750847	chr10:124215421	C=0.805, T=0.195	C	T
rs3750846	chr10:124215565	T=0.805, C=0.195	T	C
<b>Haplotype Count</b>			<b>808</b>	<b>191</b>
<b>Haplotype Frequency</b>			<b>0.8032</b>	<b>0.1899</b>

**Fig. 1.** The prevalence of possible haplotypes, according to the LDhap Tool for polymorphisms rs61871744, rs10490924, rs3750848, rs3750847, rs3750846 of the *ARMS2* gene, in EUR populations (SEU, TSI, FIN, GBR, IBS. Moreover, the risk haplotype occurs with a frequency of 0,1899 (18,99 %). In addition, this risk haplotype will occur in the homozygous state with a frequency of 3,6 % (which is close to the prevalence of early AMD 3,5 % aged 55–59 years) and in the heterozygous state with a frequency of 30,5 % (which is 2 times higher than the prevalence of late AMD at the age of  $\geq 85$  years) [9]

**Acknowledgements:** The study was performed in the framework of the Program for Basic Research of State Academies of Sciences for 2013–2022.

### References

1. DeAngelis M.M. et al. Genetics of age-related macular degeneration (AMD). *Hum Mol Genet* 2017;26(R1):R45-R50. doi:10.1093/hmg/ddx228.
2. Sobrin L. et al. Heritability and genome-wide association study to assess genetic differences between advanced age-related macular degeneration subtypes. *Ophthalmology*. 2012;19(9):1874-85. doi:10.1016/j.ophtha.2012.03.014.
3. Ruamviboonsuk P. et al. Genome-wide association study of neovascular age-related macular degeneration in the Thai population. *J Hum Genet* 2017;62(11):957-962. doi: 10.1038/jhg.2017.72.
4. Guindo-Martínez M. et al. The impact of non-additive genetic associations on age-related complex diseases. *Nat Commun*. 2021;12(1):2436. doi: 10.1038/s41467-021-21952-4.
5. Buniello A., MacArthur J.A.L., Cerezo M., Harris L.W., Hayhurst J., Malangone C., McMahon A., Morales J., Mountjoy E., Sollis E., Suveges D., Vrousou O., Whetzel P.L., Amode R., Guillen J.A., Riat H.S., Trevanion S.J., Hall P., Junkins H., Flicek P., Burdett T., Hindorf L.A., Cunningham F., Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(Database issue):D1005-D1012.
6. Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
7. Machiela M.J., Chanock S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-3557.
8. Manhan T. et al. rs10490924 surrounding HTRA1/ARMS2 regulates the susceptibility of age-related macular degeneration. *J Recept Signal Transduct Res*. 2021;41(2):188-195.
9. Colijn J.M. et al. Prevalence of age-related macular degeneration in Europe: The past and the future. *Ophthalmology*. 2017;124(12):1753-1763. doi: 10.1016/j.ophtha.2017.05.035.

## ***In silico* determination of the risk haplotype for developing preeclampsia**

Karpova N., Dmitrenko O.

*Institute of General Pathology and Pathophysiology, Moscow, Russia*

\* [nataliakarpova.sp@gmail.com](mailto:nataliakarpova.sp@gmail.com)

**Key words:** FLT1, preeclampsia, risk haplotype, pathology of pregnancy

*Motivation and Aim:* Preeclampsia (PE) refers to hypertensive disorders during pregnancy, which poses a threat to the life of the mother and fetus. PE complicates about 2–8 % of pregnancies and is diagnosed as a combination of high blood pressure and proteinuria [1]. Trisomy for chr13 is associated with PE, and this chromosome contains the FLT1 gene, polymorphisms in which are also associated with PE (rs4769612-C, rs4769613-C, rs7318880-T). Moreover, the polymorphism of the fetus, rather than the mother, plays an important role [2]. Phupong et al. showed that sFlt1/PlGF is not associated with other pregnancy complications other than PE [3]. Based on these data, we decided to investigate the presence of regulatory regions in polymorphisms and determine the prevalence of the haplotype, which includes risk alleles.

*Methods and Algorithms:* We used the GWAS catalog and PubMed to identify SNPs in FLT1 associated with preeclampsia in the maternal and child genomes, followed by mapping these SNPs to the human genome in the UCSC genome browser to identify overlapping polymorphisms in regulatory regions [4, 5]. Next, we analyzed the presence of periods in which the enhancer signature is detected in the tissues of the placenta and embryo, using the cCRE details in ENCODE SCREEN [6]. For SNPs in the regulatory region, we selected SNPs with a MAF of 1 % and identified possible haplotypes in European populations.

*Results:* In the GWAS-catalog, we found 2 polymorphisms in FLT1 associated with preeclampsia: rs4769612-C (p-value  $4 \times 10^{-14}$ ) and rs7318880-T (p-value  $8 \times 10^{-8}$ ), with rs4769612 associated with preeclampsia in analysis of the child's genotype, and rs7318880 in the analysis of the mother's genotype. According to the UCSC genome browser (oRegAnno) data, there are 4 regulatory elements in the rs4769612 region: OREG1191996, OREG1658246, OREG1688336, OREG1537828. According to the cCRE details of ENCODE SCREEN, this region contains the putative regulatory element EH38E1663332, the largest distal enhancer signature of which sharply increases at 16 weeks of gestation in the placenta and fetal tissues, which can lead to changes in FLT1 expression. In addition to rs4769612 and rs7318880, the EH38E1663332 region contains 7 more polymorphisms with MAF > 1 %: rs7320190, rs12867370, rs4769613, rs74623647, rs7321138, rs76592233, rs9579193. Of these, only rs7320190 and rs4769613 are mentioned in scientific articles and rs4769612 and rs7318880 in the GWAS database, as well as rs12867370 is associated with the risk of developing schizophrenia in offspring born to mothers with PE [7]. When assessing the prevalence of 4 possible haplotypes, a possible haplotype of the risk of developing preeclampsia was determined, for which the prevalence was 8.25 % (considering the risk alleles of maternal and fetal PE). In addition, this risk haplotype will occur in the homozygous

state with a frequency of 0.68 %, which is close to the prevalence of early-onset preeclampsia rate of 0.38 % [1] (Fig. 1).

RS Number	Position (GRCh37)	Allele Frequencies	Haplotypes			
rs7320190	chr13:29138256	T=0.791, C=0.209	T	T	C	C
rs7318880	chr13:29138285	T=0.539, C=0.461	C	T	T	T
rs12867370	chr13:29138398	G=0.917, A=0.083	G	G	G	A
rs4769612	chr13:29138498	C=0.542, T=0.458	T	C	C	C
rs4769613	chr13:29138609	C=0.544, T=0.456	T	C	C	C
rs74623647	chr13:29138632	G=1.0, T=0.0	G	G	G	G
rs7321138	chr13:29138705	T=0.793, C=0.207	T	T	C	C
rs76592233	chr13:29138761	C=1.0, T=0.0	C	C	C	C
rs9579193	chr13:29138768	G=0.794, A=0.206	G	G	A	A
<b>Haplotype Count</b>			<b>458</b>	<b>333</b>	<b>124</b>	<b>83</b>
<b>Haplotype Frequency</b>			<b>0.4553</b>	<b>0.331</b>	<b>0.1233</b>	<b>0.0825</b>

**Fig. 1.** The prevalence of possible haplotypes, according to the LDhap Tool for polymorphisms rs7320190, rs7318880, rs12867370, rs4769612, rs4769613, rs74623647, rs7321138, rs76592233, rs9579193, calculated for EUR populations (SEU, TSI, FIN, GBR, IBS). Moreover, the risk haplotype occurs at 8.25 % (C\_T\_A\_C\_C\_G\_C\_C\_A)

**Conclusion:** As a result, we have *in silico* determined a possible preeclampsia risk haplotype (C\_T\_A\_C\_C\_G\_C\_C\_A), with a prevalence of 0.68 % for homozygotes, which is comparable to the development of early-onset preeclampsia rate of 0.38 %. The resulting hypothesis will be further tested on clinical samples obtained from mother-child pairs in a case-control study.

**Acknowledgements:** The study was performed in the framework of the Program for Basic Research of State Academies of Sciences for 2013-2022.

### References

1. Lisonkova S., Joseph K.S. Incidence of preeclampsia: risk factors and outcomes associated with early-versus late-onset disease. *Am J Obstet Gynecol.* 2013;209(6):544.e1-544.e12.
2. Zhao L., Bracken M.B., DeWan A.T. Genome-Wide association study of pre-eclampsia detects novel maternal single nucleotide polymorphisms and copy-number variants in subsets of the hyperglycemia and adverse pregnancy outcome (HAPO) study cohort. *Ann Hum Gene.* 2013;77(4):277-287.
3. Phupong V. et al. Soluble fms-like tyrosine kinase 1 and placental growth factor ratio for predicting preeclampsia in elderly gravida. *Hypertension Pregnancy.* 2020;39(2):139-144.
4. Buniello A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012.
5. Kent W.J. et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996-1006.
6. Moore J.E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699-710.
7. Malaspina D. et al. Short duration of marriage at conception as an independent risk factor for schizophrenia. *Schizophr Res.* 2019;208:190-195.

## Autophagy inducer trehalose positive effect in db/db mice, genetic model of diabetes

Korolenko T.A.<sup>1\*</sup>, Bgatova N.P.<sup>2</sup>, Korolenko E.T.<sup>3</sup>, Yuzik E.I.<sup>4</sup>, Akopyan A.A.<sup>1</sup>

<sup>1</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *University Canada West, Vancouver, Canada*

<sup>4</sup> *Institute of Molecular Pathology and Pathomorphology, Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia*

\* [t.a.korolenko@physiol.ru](mailto:t.a.korolenko@physiol.ru)

Autophagy suppression was shown in aging, diabetes [1]. Db/db mice is a model of diabetes, related to leptin receptor deficiency, was used in experiments. The aim – to evaluate effect of autophagy inducer trehalose on autophagic activity in db/db mice with significant lipid storage syndrome in liver and other organs. The study was carried out on a specific pathogen-free mice of db/db strain, 3 months old (SPF-vivarium of the Institute of Cytology and Genetics, SB RAS, Novosibirsk). Mice were subdivided into four groups: 1) WT mice + H<sub>2</sub>O; 2) WT mice drinking 2 % trehalose (24 days); 3) db/db mice + H<sub>2</sub>O; 4) db/db mice drinking 2 % trehalose (24 days). It was shown that treatment by trehalose reduced weight of db/db mice, improved glycemic profile and behavioral characteristics of db/db mice. Immunohistochemical analysis revealed significant decrease of LC3-II expression in hippocampal areas in db/db mice compared to the control WT mice, while trehalose treatment significantly increased this index. We have shown positive effect of trehalose in liver and heart of db/db mice, revealing activation of lipophagy by trehalose. According to electron microscopic study, trehalose significantly decreased volume density of lipid inclusions in cardiomyocytes of db/db mice. Trehalose can restore the suppressed autophagy induced by high glucose *in vivo* in db/db mice. Autophagy activation may be useful for treatment of diabetes, regulating the balance between apoptosis and autophagy by inhibiting mTOR signaling. Mechanism of protective effect of trehalose includes activation of hepatic transcription factor EB (TFEB), increasing TFEB nuclear translocation, elevating level of LC-3-II (marker of autophagosome) in mouse liver. These findings provide the basis for trehalose usage in pathology.

*Acknowledgements:* The study was supported by budgetary funding for basic scientific research, Institute of Neurosciences and Medicine (theme No. 122042700001-9 (2021-2025)).

### References

1. Korolenko T.A., Dubrovina N.I., Ovsyukova M.V., Bgatova N.P., Tenditnik M.V., Pupyshev A.B., Akopyan A.A., Goncharova N.V., Lin C.L., Zavjalov E.L., Tikhonova M.A., Amstislavskaya T.G. Treatment with autophagy inducer trehalose alleviates memory and behavioral impairments and neuroinflammatory brain processes in db/db mice. *Cells*. 2021;10(10):2557. doi: 10.3390/cells10102557.

## A new approach to stimulating the wound healing based on *Opisthorchis felineus* proteins

Kovner A.<sup>1\*</sup>, Zaparina O.<sup>1</sup>, Kapushchak Ya.<sup>1</sup>, Tarasenko A.<sup>1,2</sup>, Mordvinov V.<sup>1</sup>, Pakharukova M.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *anya.kovner@gmail.com*

**Key words:** *Opisthorchis felineus*, lysate, excretory-secretory product, wound healing, extracellular matrix, angiogenesis

**Motivation and Aim:** Skin regeneration is a natural homeostatic process, however, in some cases, chronic non-healing wounds or abnormal scarring develop. This situation is typical for such pathological conditions as deep mechanical damage, severe burns, old age, obesity, diseases of the circulatory system, as well as diabetes mellitus and other autoimmune diseases [1]. Currently, the correction of non-healing wounds is extremely difficult, which dictates the search for new promising therapeutic agents. It should be noted the difficult task of finding drugs that can non-specifically stimulate the processes of skin regeneration, regardless of endo- and/or exogenous triggers of pathological non-healing of wounds. In this regard, the search for new non-specific stimulators of skin regeneration of plant and animal origin seems to be highly relevant. The trematode *Opisthorchis felineus* infects the hepatobiliary system of fish-eating mammals [2]. These parasites have the ability to reduce acute inflammation, and also cause not only damage, but also stimulate the proliferation of bile duct cholangiocytes [3, 4]. Based on this, the aim of the study is to investigate the properties of *O. felineus* proteins as a wound healing agent.

**Methods and Algorithms:** 80 male mice of the C57Bl/6 line were inflicted with superficial wounds with a diameter of 8 mm. Further, the animals were divided into 8 groups: control (chlorhexidine, 1.5 % methylcellulose, BSA) and experimental (ESP 1, 10, 10 µg without endotoxin; lysate 10 and 50 µg). After applying the solutions, the wound was treated with Luxplast spray plaster (Farmac-zabban, Italy), which created a water-repellent film to protect against contamination and fix the solutions. All groups will receive treatment every 3 days of the experiment with simultaneous detection of the wound area. Animals will be withdrawn from the experiment on days 7 and 10 of treatment. Damaged skin samples were taken for histological examination (formalin and freezing) and for the evaluation of gene expression (RNA-later).

**Results:** ESP and *O. felineus* lysate proteins significantly increased wound healing in mice ( $p < 0.05$ ). Histological methods showed that accelerated overgrowth was accompanied by a significant decrease in the area of inflammatory infiltration. On 10<sup>th</sup> day of the experiment, there was no wet crust in the treatment groups, and re-epithelization was detected in the lysate groups. In addition, by day 10, an increase in the number of newly formed CD34<sup>+</sup> vessels was diagnosed in the treatment groups, which indicates an improvement in tissue trophism.

According to the results of analysis of the expression of genes-markers of inflammation (Bl<sub>4</sub>, Nos2, Arg1), organization of the extracellular matrix (Acta2, Coll $\alpha$ , Col 3,

MMP9, Tgf $\beta$ , FGF2, Fn1), the state of the vascular bed (VEGF $\alpha$ ) and nervous tissue (NGF, Nestin, NG2) when wounds are treated with lysate and ESP proteins of the trematode *O. felineus*, all of the above processes are completed faster than in animals of the control groups.

Analysis of ESP and lysate of the *O. felineus* proteome revealed not only heme-binding proteins, but also GST, TPX, HDM1 and annexin A2 among the main proteins, which require further individual study as potential participants in wound healing processes.

*Conclusion:* Thus, preparations based on the trematodes *O. felineus* contribute to faster and better healing of superficial wounds. Apparently, they contain bioactive molecules that can be considered as potential agents for stimulating the regeneration of mammalian tissues and requiring further study.

*Acknowledgements:* The study is supported by the Russian Science Foundation (No. 22-25-20018).

### References

1. Aitchison S.M. et al. Skin wound healing: normal macrophage function and macrophage dysfunction in diabetic wounds. *Molecules*. 2021;26(16):4917.
2. Pakharukova M.Y., Mordvinov V.A. The liver fluke *Opisthorchis felineus*: biology, epidemiology and carcinogenic potential. *Trans R Soc Trop Med Hyg*. 2016;110(1):28-36.
3. Maksimova G.A. et al. Effect of *Opisthorchis felineus* infection and dimethylnitrosamine administration on the induction of cholangiocarcinoma in Syrian hamsters. *Parasitol Int*. 2017;66(4):458-463.
4. Kovner A.V. et al. Characteristics of liver fibrosis associated with chronic *Opisthorchis felineus* infection in Syrian hamsters and humans. *Exp Mol Pathol*. 2019;110:104274.



## **Pharmacogenetic influence of complement genes genotypes on the response to anti-VEGF treatment for age-related macular degeneration in a Russian population**

Kozhevnikova O.S.\*, Fursova A.Zh., Devyatkin V.A., Rumyantseva Y.V., Telegina D.V., Derbeneva A.S., Nikulich I.F., Kolosova N.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *oidopova@bionet.nsc.ru*

Age-related macular degeneration (AMD) is the leading cause of vision loss in the aging population. The current standard of care for neovascular AMD is the administration of vascular endothelial growth factor (VEGF) inhibitors. However, a high degree of variability in treatment response has been observed. The aim of the study was to determine the effect of gene polymorphisms related to the complement system on the dynamics of changes in the functional and anatomical parameters of the retina according to optical coherence tomography during anti-VEGF therapy. The genotype frequencies of the rs2285714 (CFI), rs10490924 (ARMS2), rs2230199 (C3), rs800292 (CFH), and rs6677604 (CFH) loci in the Russian cohort were determined. The height pigment epithelium detachment and neuroepithelial detachment depended on the rs2285714 genotype and were significantly higher in carriers of the risk allele initially, after 3 and after 5 doses of drug. The chances of the presence of anastomoses and loops and the activity of the neovascular membrane in the end of observation period in individuals homozygous for the risk allele of rs2285714 were significantly higher. Also association of rs2230199 with the presence of intraretinal cysts was revealed. According to the regression analysis considering gender, age, and baseline values, it was found that rs2285714 was associated with the dynamics of a decrease in the central retinal thickness (CRT) during treatment: the CRT decline was significantly less in patients with risk allele. Thus, the severity of morphological changes and the extent of macular lesions are associated with polymorphisms rs2285714 (CFI), rs2230199 (C3) in genes that play a key role in regulating the activity of the alternative pathway of the complement system. Patients with the risk allele rs2285714 poorly respond to anti-angiogenic therapy. *Acknowledgements:* Supported by RSF 21-15-00047.

## ReLEx SMILE lenticules as a source of obtaining stromal corneal cells

Krasner K.U.<sup>1,2\*</sup>, Poveshenko O.V.<sup>1</sup>, Surovtseva M.A.<sup>1</sup>, Kim I.I.<sup>1</sup>, Bondarenko N.A.<sup>1</sup>, Trunov A.N.<sup>2</sup>, Chernykh V.V.<sup>2</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk Branch of S. Fyodorov Eye Microsurgery Federal State Institution, Novosibirsk, Russia*

\* [kityli@mail.com](mailto:kityli@mail.com)

**Key words:** cornea, stromal cells, keratocyte, fibroblast, lenticules

*Motivation and Aim:* It is well-known that the corneal blindness takes the third leading place in the structure of the causes of eye diseases [1]. The impaired function of keratocytes in degenerative diseases of the cornea is accompanied by its clouding [2]. The transparency is a key property of the cornea, providing its optical properties. Corneal transplantation remains the treatment for patients with severe corneal lesions, but has some limitations, including the deficiency of donor corneal tissue and post-transplant complications. Cellular therapy using human corneal stromal cells (hCSC) is an alternative approach for treating the cornea. Aim of this study to isolate hCSCs from a new source – ReLEx SMILE lenticules and conduct a comparative of the phenotype and functional properties for fibroblasts and keratocytes, with the prospect in the treatment of diseases associated with corneal opacity.

*Methods and Algorithms:* hCSCs were isolated from lenticules and cultivated in DMEM/F12 media with 10 % FCS, L-glutamin, antibiotics. The differentiation of corneal fibroblasts into keratocytes was performed in a selective KGM (keratocyte growth medium). The cell culture phenotypes were determined by flow cytometry. The proliferation and migration of cells were assessed using a real time cell analyzer RTCA. The cell production of EPO, BDNF, VEGF, TNF-alpha, IGF-1, SDF-1a, sICAM-1, vimentin, fibronectin, and total collagen was determined immunosorbent assay (ELISA).

*Results:* The hCSCs had the morphology and phenotype of fibroblasts in the growth medium with the addition of 10 % FCS and the expressed markers of mesenchymal cells CD73 ( $99.2 \pm 1.62$  %), CD105 ( $22.8 \pm 2.91$  %) and were negative for CD90, CD34, CD45, HLA-DR, as well as keratocan and lumican. Keratocytes cultured in KGM expressed high levels of keratocyte markers: keratocan ( $98.8 \pm 1.93$  %) and lumican ( $81.4 \pm 26.64$  %) but not fibroblast markers. Corneal fibroblasts had more pronounced proliferative and migratory activity than keratocytes according to real time cell analyzer RTCA. The fibroblast conditioned medium contained the higher levels of VEGF and vimentin. The keratocytes produced higher levels of EPO and BDNF. The level of fibronectin, collagen, SDF-1a, IGF-I, TNF-alpha, and sICAM-1 was similar in the conditioned mediums of the fibroblast and keratocytes.

*Conclusion:* Our data indicate that lenticules obtained with the ReLEx SMILE can be a source of hCSCs and keratocytes 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. Hassell J.R., Birk D.E. The molecular basis of corneal transparency. *Exp Eye Res.* 2010;91(3):326-335. doi: 10.1016/j.exer.2010.06.021.

## Exploring lncRNA-mRNA signatures in relapsed acute myeloid leukemia compared with primary tumors

Kulaeva E.\*, Mashkina E.

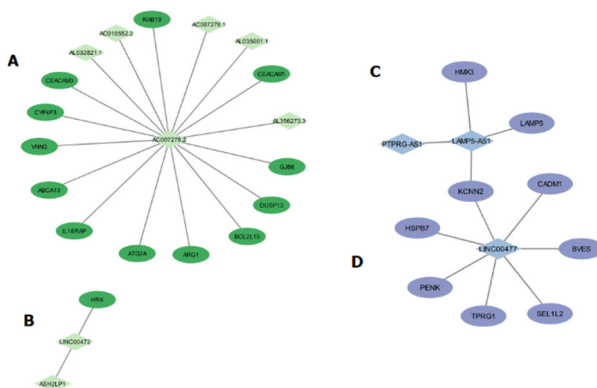
Academy of Biology and Biotechnology, Southern Federal University, Rostov-on-Don, Russia

\* ekulaeva@sfsedu.ru

**Key words:** lncRNA, pediatric acute myeloid leukemia, TARGET-AML

**Motivation and Aim:** Long non-coding RNAs (lncRNAs) play a role in the pathogenesis of acute myeloid leukemia (AML) by affecting gene expression, including in pediatric AML. Deregulation of lncRNAs patterns has the potential to regulate tumor drug resistance and sustaining of leukemic stem cells. Here we investigated the expression patterns of lncRNAs and mRNAs in relapsed tumors compared to primary tumors, using TARGET-AML project data.

**Methods and Algorithms:** We examined the molecular pathways for differentially expressed lncRNAs and mRNAs (using GO and LncPath) and constructed lncRNA-mRNA co-expression networks (Fig. 1, A, B for upregulated genes and Fig. 1, C, D for downregulated genes).



**Fig. 1.** lncRNA-mRNA co-expression networks of upregulated and downregulated genes. Upregulated mRNAs are shown as dark green ellipses; upregulated lncRNAs are shown as light green diamonds. Downregulated mRNAs are shown as lavender ellipses; downregulated lncRNAs are shown as light blue diamonds

**Results:** The lncRNA-mRNA landscape of relapsed tumors compared to primary tumors in pediatric AML is extremely inconsistent. In both upregulated networks and downregulated co-expression networks, there are parallel processes associated with the effects of therapy and mechanisms of drug resistance and immune system escape. Most of the genes involved in the interaction networks are nonspecific for a particular tumor type; intriguingly, some of them are associated with invasion processes that are not typical for hematological malignancies.

**Conclusion:** Detailed studies of the resulting interactions using other pediatric AML datasets are needed to better understanding the current results.

## Artificial intelligence (AI) of 3D MRI images for neurooncology

Letyagin A.Yu.<sup>1,3\*</sup>, Amelina E.V.<sup>1</sup>, Tuchinov B.N.<sup>1</sup>, Tolstokulakov N.Yu.<sup>1</sup>,  
Amelin M.E.<sup>1,2</sup>, Pavlovsky E.N.<sup>1</sup>, Groza V.V.<sup>1,4</sup>, Golushko S.K.<sup>1</sup>

<sup>1</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>2</sup> *Federal State Budgetary Institution "Federal Neurosurgical Center", Novosibirsk, Russia*

<sup>3</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>4</sup> *Median Technologies, Valbonne, France*

\* *letyagin-andrey@yandex.ru*

In structural and statistical analysis of neurooncological lesions, there are problems of detection, classification, contouring of tumors, and segmentation of its subregions (Gd-contrast-accumulating, Gd-contrast-non-accumulating, necrotic part of the tumor, perifocal edema), treatment prognosis. Almost all of these tasks are usually solved diagnosis neuro-oncology diseases with use the Artificial Intelligence (AI) applications. AI models training is based on the "Federal Neurosurgical Center" in Novosibirsk, Russia data on patients, who underwent MRI imaging of the brain on a 1.5 T (Siemens Magnetom Avanto) and 3.0 T (Ingenia, Philips) magnetic resonance tomographs: astrocytomas (70; 39 % male; median = 37, IQR (31; 56) years), glioblastomas (140; 53 % male; median = 57, IQR (48; 63) years), meningiomas (160; 26 % male; median = 58, IQR (50; 64) years), neurinomas (130; 32 % male; median = 53, IQR (40; 59) years). MRI sequences: T1-WI, T1-WI-Gd+, T2-WI, T2-FLAIR and DWI; scans in a single isovoxel anatomical template with a resolution of 1 mm<sup>3</sup>. The data were processed in the open-source 3D Slicer ([www.slicer.org](http://www.slicer.org)) cross-platform system with manual tumor segmentation and post-correction by two expert neuroradiologists.

Firstly, we elaborated 2D approach for segmentation and classification tasks. As part of the PyTorch machine learning framework, two convolutional neural networks have been developed based on the se-resnext50-32x4d, se-resnext101-32x4d networks [1]. Segmentation of tomograms is implemented according to a multilayer model of a brain tumor (layer = disease) of logically connected 4 blocks (= variant or stage of the disease), consisting of a series of subblocks of the 2D convolution layer, batch normalization, and activation, as well as channel attention blocks, compression and excitation blocks, followed by a Pooling layer, a DropOut layer, and a Linear Fully Connected layer to enhance tumor-specific features. Sections including only perifocal edema are not informative and were excluded from the test set.

The task of tumor type classification was solved on the basis of our dataset. We performed tumor classification using two different CNN developed in the PyTorch framework, that are slightly changed variants of the well-known se-resnext50-32x4d, se-resnext101-32x4d networks. We obtained results using a test set of 40 independent patients with balanced representation of tumor classes. There were only 3 misclassified patients among given 40 in the test set, so it was possible to analyze them individually as well as some other difficult cases that were tricky to resolve properly. We analyzed the observed classification errors from both machine learning and medical perspectives. We also developed recommendations how to improve the results of the algorithm.

Now we expand our SBT dataset and extend our research work to 3D models for account the all context along the axial, sagittal and coronal planes. We have used a well-proven (won on BraTS 20) the nnU-net framework as a 3D baseline. This is a symmetric encoder-decoder architecture using skip-connection. Downsampling is performed with stridden convolutions and upsampling is performed with convolution transposed. Each decoder output is used for dip supervision. A total of five downsampling operations are performed. The initial number of convolution kernels is set to 32, which is doubled with each downsampling up to a maximum of 320. The non-linearity function is Leaky ReLU and normalization is instance normalization

The second 3D approach is DMFnet (Dilated Multi-fiber network). Their main idea is to divide the convolution operation into several paths, called Fiber. Information between the fibers is distributed by means of multiplexer, consisting of 1x1 reapplication of convolution. The difference between DMF and MF units is that the first one applies not only conventional convolutions, but also dilated convolution with  $d = 1, 2, 3$  and different weights. This is done to increase the receptive field and capture spatial correlations at different scales.

The main body of the network is composed of the MF/DMF units, excluding the first and last convolution layers. In the feature encoding stage, we apply the DMF unit in the first six encoding units to achieve multi-scale representation. In the decoding stage, the high-resolution features from the encoder are concatenated with the upsampled features. All upsampling operations are performed using 3-linear interpolation. The nonlinearity function is standard ReLU and normalization in the form of Batch Normalization.

TransBTS is a hybrid architecture of CNN-Transformers type encoder-decoder with skip-connection and bottle-neck transformer sequence. The standard encoder block consists of a 3x3x3 convolution, activation and normalization functions in the form. Downsampling done with strided convolutions. After the encoder feature map contains 128 channels. Then, the 3x3x3 convex representation of each volume is applied again and the number of channels is increased to 512. Since transformers work with 2D patches, the feature map turns into a vector. After that, 8 layers of Vision Transformer (ViT) \cite{vit} are followed to find global dependencies in the obtained features. The ViT output is also a sequence, so it is recorded into a volumetric matrix to obtain an image. The resulting 3-dimensional weighted matrix goes to the input of the decoder. The building block of the decoder consists of 3x3x3 usual and transposed convolutions

**Table 1.** Results on BraTS dataset. All metrics are average results for 5 folds

Models	Dice ET	Dice TC	Dice WT
LinkNet	0,614	0,758	0,871
nnU-net	0,771	0,836	0,916
DMFnet	0,742	0,812	0,893
TransBTS	0,764	0,817	0,897

**Table 2.** Results on SBT testing dataset

Models	Dice ET	Dice TC	Dice WT
LinkNet	0,826	0,826	0,911
nnU-net	0,846	0,867	0,917
DMFnet	0,796	0,827	0,900
TransBTS	0,756	0,753	0,858

Based on nnU-net and DMFnet we developed the new techniques – Brain Tumor Segmentation with Self-supervised Enhance Region Post-processing. It is based on the clinical radiology hypothesis and presents an efficient approach to combining and matching 3D methods to search for areas of comprised the GD-enhancing tumor in order to significantly improve the model's performance of the particular applied numerical problem of brain tumor segmentation.

The obtained stability and robustness results of this method was confirmed by the Brats 2021. The proposed method also demonstrates significant improvement on the segmentation problem with respect to Dice and Hausdorff metrics compared to similar training / validation procedures based on any regardless of the chosen neural network architecture. These particular results are very important from the clinical point of view and can strongly increase the quality of the computer-aided systems where similar solutions are deployed.

*Acknowledgements:* The study was supported by the Russian Foundation for Basic Research project No. 19-29-01103.

### *References*

1. Hu J., Shen L., Sun. G. Squeeze-and-Excitation Networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. doi: 10.1109/CVPR.2018.00745.

## Extracellular matrix in experimental hepatocarcinoma-29

Makarova V.\*, Bgatova N.

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

\*shedina\_vika@mail.ru

**Key words:** hepatocellular carcinoma, extracellular matrix, PAS-reactoin

*Motivation and Aim:* Hepatocellular carcinoma (HCC) is one of the most aggressive human tumor with a high mortality rate [1]. Despite the fact that this is a frequently occurring tumor, the mechanisms of its pathogenesis and metastasis are not fully understood, and the existing therapy often fails [2]. There is evidence that in patients with HCC, changes occur in the microenvironment of the tumor tissue, including the extracellular matrix (ECM), stimulating the growth and metastasis of HCC cells [3]. Therefore, the aim was to study form of ECM in the tumor during spontaneous development and lithium carbonate treatment.

*Methods and Algorithms:* The research was carried out on male mice of the CBA line. Hepatocarcinoma-29 (G-29) cells were used to induce the tumor. Animals were divided into 2 groups: 1 – animals with tumor growth; 2 – animals with tumor and lithium carbonate treatment. The material was taken on the 30th day of the experiment. Staining of tissues was carried out by the method of PAS-reaction. Morphometric analysis of tumor tissue was performed.

*Results:* It was found that part of the PAS-positive patterns had gaps and contained erythrocytes. While the other part did not have visible gaps, therefore it is impossible to judge their functional state. The volume density of PAS-positive patterns in tumor in condition spontaneous development was significantly higher than in lithium carbonate treated animals. There is evidence that ECM is actively involved in the processes of carcinogenesis and metastasis of HCC. At the same time, the levels of proteoglycans increase [4]. Apparently, lithium carbonate inhibits the formation of proteoglycans in the ECM, thereby suppressing tumor growth and development.

*Conclusion:* The state of the ECM is an indicator of tumor growth. A decrease of the PAS-positive components volume density in the tumor tissue after the lithium carbonate using may be the result of G-29 growth suppression.

*Acknowledgements:* The work was carried out at the expense of budget financing within the framework of the state task on the topic No. FWNR-2022-0012.

### References

1. Chedid M.F., Krueel C.R.P., Pinto M.A., Grezzana-Filho T.J.M., Leipnitz I., Krueel C.D.P., Scaffaro L.A., Chedid A.D. Hepatocellular carcinoma: diagnosis and operative management. *Arq Bras Cir Dig.* 2017;30(4):272-278.
2. Ogunwobi O.O., Harricharran T., Huaman J., Galuza A., Odumuwaogun O., Tan Y., Ma G.X., Nguyen M.T. Mechanisms of hepatocellular carcinoma progression. *World J Gastroenterol.* 2019;25(19):2279-2293. doi: 10.3748/wjg.v25.i19.2279.
3. Pickup M.W., Mouw J.K., Weaver V.M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* 2014;15:1243-1253.
4. Nault J.-C., Guyot E., Laguillier C., Chevret S., Ganne-Carrie N., N'Kontchou G., Beaugrand M., Seror O., Trinchet J.-C., Coelho J. Serum proteoglycans as prognostic biomarkers of hepatocellular carcinoma in patients with alcoholic cirrhosis. *Cancer Epidemiol Biomarkes Prev.* 2013;22:1343-1352.

## New generation dressings for wound healing: combined antibacterial activity and boosted regeneration process

Manakhov A.<sup>1\*</sup>, Permyakova E.<sup>1,2</sup>, Solovieva A.<sup>1</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *National University of Science and Technology “MISiS”, Moscow, Russia*

\* *ant-manahov@ya.ru*

**Key words:** wound dressings, surface functionalization, regeneration, plasma

*Motivation and Aim:* Type 2 diabetes mellitus has been awarded the status of a 21st-century non-infectious pandemic. Complications of this disease lead to a significant decrease in the quality of life, disability of a fairly young, capable population. One of the complications is chronic non-healing wounds, leading to amputations in 50 %. The therapy of chronic wounds is difficult since this pathology is characterized by multifactorial homeostatic disorders at the systemic and local levels.

The low efficiency of therapy for complications of type 2 diabetes (diabetic foot syndrome, DFS) is an urgent problem and standard treatment protocols do not lead to its solution, which is manifested by a high frequency of amputations in this category of patients. The low efficiency of therapy for complications of type 2 diabetes (diabetic foot syndrome, DFS) is an urgent problem and standard treatment protocols do not lead to its solution, which is manifested by a high frequency of amputations in this category of patients.

The result of active scientific research and technological solutions is the creation of effective tissue-engineered structures for the treatment of chronic wounds [1, 2]. Foreign clinical trials have shown the high efficiency of using bioengineered skin substitutes for the treatment of non-healing wounds: cell-free human skin matrices – AlloDerm® (LifeCell, USA), GraftJacket® (Wright Medical Technology, USA), pig skin – Permacol® (Covidien, USA), Oasis® (Cook Biotech, USA).

Excessive degradation of the extracellular matrix, proteins, growth factors leads to impaired adhesion and functional activity of cells, to a decrease in the proliferative activity of fibroblasts and their susceptibility to growth factors.

The extracellular matrix (ECM) provides structural support, intercellular contact, serves as a reservoir for signaling molecules, thereby regulating cell migration, proliferation and angiogenesis. ECM failure is a key link in the pathogenesis of chronic ulcers in DFS, and it is ECM that is the goal of innovative approaches to the treatment of chronic wounds [3].

Thus, the development of tissue-engineered structures that replace defective ECM is an effective way to accelerate wound healing. Today, analogues of human skin dermal equivalents are polymer matrices, which are represented by both biopolymers (polypeptides, hydroxyapatites, glycosaminoglycans, fibronectin, collagen, chitosan, and alginates) and synthetic, completely degradable polymers (polyglycolides, polylactides, polytetrafluoroethylenes), polycaprolactone) [4–6]. Synthetic polymeric materials have less expensive and more reliable sources of raw materials [7].



*Methods and Algorithms:* We utilized the plasma functionalized nanofiber prepared by electrospinning of polycaprolactone modified by low temperature plasma and grafting of Platelet-rich plasma species and curdlan/chitosan foams in immobilized Ag nanoparticles to test *in vivo* the efficiency of new wound dressings.

*Results:* Herein we present the innovative two-component system for wound care that contains: 1. plasma treated poly( $\epsilon$ -caprolactone) nanofiber layer bonded with platelet-rich plasma (PRP) proteins (PRP PCL NFs – covering directly on the wounds) to enhance regeneration rate and to prevent scar tissue formation; 2. antibacterial superabsorbent foam layer (outer layer) which provides good oxygenation, prevents bacterial infection, and absorbs exudate, forming a soft gel (moist environment). These foams were prepared from a mixture of hydrolyzed curdlan and chitosan by lyophilization process. To enhance antibacterial properties AgNO<sub>3</sub> solution was added to curdlan/chitosan mixture during the polymerization process and then reduced by UV irradiation. The membranes were further investigated for their structure and composition using optical microscopy, scanning electron microscopy, EDX-analysis, FT-IR spectroscopy and XPS modelling. The influence of using obtained dressings on the regeneration process was studied *in vivo* and compared with commercial dressings.

*Conclusion:* The results obtained in this research proves high potential of nanofibers and news superabsorbing foams as wound dressing for type 2 diabetes DFS.

*Acknowledgements:* The study is supported by Russian Science Foundation (grant No. 18-75-10057).

### References

1. Liu X., Xu H., Zhang M., Yu D.-G. Electrospun Medicated Nanofibers for Wound Healing: Review. *Membranes (Basel)*. 2021;11:770.
2. Lai H.J., Kuan C.H., Wu H.C., Tsai J.C., Chen T.M., Hsieh D.J., Wang T.W. Tailored design of electrospun composite nanofibers with staged release of multiple angiogenic growth factors for chronic wound healing. *Acta Biomater*. 2014;10:4156.
3. Lev-Tov H., Li C., Dahle S., Isseroff R.R. Cellular versus acellular matrix devices in treatment of diabetic foot ulcers: study protocol for a comparative efficacy randomized controlled trial. *Trials*. 2013;14:8.
4. Pachence J.M., Kohn J. Biodegradable polymers In: Principles of Tissue Engineering. Elsevier, 2000:263-277.
5. Yoo H.S., Kim T.G., Park T.G. Surface-functionalized electrospun nanofibers for tissue engineering and drug delivery. *Adv Drug Deliv Rev*. 2009;61:1033.
6. Permyakova E.S., Kiryukhantsev-Korneev P.V., Gudz K.Y., Konopatsky A.S., Polčák J., Zhitnyak I.Y., Glouhankova N.A., Shtansky D.V., Manakhov A.M. Comparison of different approaches to surface functionalization of biodegradable polycaprolactone scaffolds *Nanomaterials*. 2019;9:1769.
7. Zhong S.P., Zhang Y.Z., Lim C.T. Tissue scaffolds for skin wound healing and dermal reconstruction. *WIREs Nanomed Nanobiotechnol*. 2010;2:510.

## Melatonin effect on the thymus cell composition in mice with functional pinealectomy

Michurina S.V.<sup>1</sup>, Miroshnichenko S.M.<sup>1,2</sup>, Ishchenko I.Yu.<sup>1\*</sup>, Serykh A.E.<sup>1,3</sup>, Rachkovskaya L.N.<sup>1</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Research Institute of Biochemistry – Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

<sup>3</sup> *Research Institute of Experimental and Clinical Medicine – Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

\* *irenisich@mail.ru*

**Key words:** functional pinealectomy, continuous lighting, melatonin, thymus, CD3<sup>hi</sup> and CD3<sup>low</sup> T-lymphocytes, flow cytometry

*Motivation and Aim:* Prolonged continuous illumination contributes to the development of functional pinealectomy, a decrease in adaptive cellular and humoral immune response. The study aim was to investigate the melatonin effect on the content of young CD3<sup>low</sup> and mature CD3<sup>hi</sup> T-lymphocyte forms in the thymus of C57Bl/6j mice with a model of functional pinealectomy (FP).

*Methods and Algorithms:* Four animal groups were formed. "CL" – mice with the FP model, which was created by keeping animals under continuous lighting (CL) for 14 days (photoperiod light:dark 24:0 hours). "CL+MT" – animals with FP which were intragastrically injected with melatonin (MT, 0.664 g/kg body weight) in 200µl of distilled water daily against a background of CL. "CL+Placebo" – mice with FP which received water 200µl intragastrically daily. "Control" – intact mice kept under standard lighting mode (14:10 h). Immunophenotyping of CD3<sup>hi</sup> and CD3<sup>low</sup> thymocyte subpopulations was performed by staining using CD3-ARS antibodies and analyzed using a flow cytometer. Statistical analysis was carried out using Mann-Whitney test ("Statistica 12").

*Results:* In FP mice, both a decrease in the relative number of CD3<sup>low</sup> and CD3<sup>hi</sup> thymocytes and a significant decline in the CD3<sup>low</sup>/CD3<sup>hi</sup> ratio were found. Treatment of animals with melatonin under CL restored the percentage of CD3<sup>low</sup> and CD3<sup>hi</sup> thymocytes, as well as the CD3<sup>low</sup>/CD3<sup>hi</sup> ratio to control values.

*Conclusion:* Thus the results indicate that prolonged continuous illumination has a depressing effect on differentiation and maturation of young thymocytes into mature forms. Melatonin treatment in this model situation helps to compensate for the CL effects by triggering the process of cell proliferation in the thymus from the earliest stages. This indicates that melatonin has immunotropic properties and can be used to correct the consequences of functional pinealectomy.

*Acknowledgements:* This study was supported by the budget project of the Research Institute of Clinical and Experimental Lymphology – Branch of the ICG SB RAS (FWNR-2022-0009). This work involved the use of the equipment of the Center for Genetic Resources of Laboratory Animals of the ICG SB RAS and the Center for collective use of scientific equipment "Spectrometric measurements" FRC FTM Research Institute of Biochemistry.

## Effect of aluminum oxide and polydimethylsiloxane on thymus cellular composition in mice kept under 24-hour lighting

Miroshnichenko S.M.<sup>1,2\*</sup>, Michurina S.V.<sup>1</sup>, Rachkovskaya L.N.<sup>1</sup>, Serykh A.E.<sup>1,3</sup>,  
Ishchenko I.Yu.<sup>1</sup>, Rachkovsky E.E.<sup>1</sup>, Letyagin A.Yu.<sup>1</sup>, Zavyalov E.L.<sup>4</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Research Institute of Biochemistry – Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

<sup>3</sup> *Research Institute of Experimental and Clinical Medicine – Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

<sup>4</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [svmiro@yandex.ru](mailto:svmiro@yandex.ru)

Lighting at night is considered as "light pollution" that disrupts the circadian regulation of melatonin synthesis, violates metabolism and other hormone-controlled systems, and also increases the risk of developing cancer. Currently, a sorbent matrix created on the basis of aluminum oxide and polydimethylsiloxane is actively used as a carrier of biologically active molecules of melatonin, lithium, etc. The study aim was to investigate the sorbent effect on the thymus T-cell composition in C57Bl/6j mice kept under 24-hour illumination for 14 days. Experiments included groups: "CL" – continuous lighting (24/0 h); "Control" – animals under a standard light regime (14/10 h); "CL+Sorbent" – mice under CL with administration of an aqueous (200 µl) sorbent suspension 0.664 g/kg body weight. Thymocytes CD3<sup>low</sup> and CD3<sup>hi</sup> were determined by flow cytofluorimetry. The Mann-Whitney criterion ("Statistica12") was used for statistical analysis. The median, first and third quartiles were determined. The differences were considered statistically significant at  $p < 0.05$ . Prolonged CL, reducing the thymocyte proliferation and activating apoptosis, led to a decrease in the percentage of CD3<sup>low</sup> and CD3<sup>hi</sup> thymocytes compared with the control. Sorbent treatment increased the relative number of young CD3<sup>low</sup> T-cells to 47.0 % [46.08; 47.25] compared with the "CL" group (38.7 % [35.35; 39.5]). The content of mature CD3<sup>hi</sup> thymocytes after sorbent administration did not differ from the "CL" group. The CD3<sup>low</sup>/CD3<sup>hi</sup> ratio of thymocytes significantly exceeded the "CL" group and reached the control level. So, the inhibitory effect of CL on the mice thymus is partially compensated by the sorbent introduction that protects young and sensitive to apoptotic factors CD3<sup>low</sup> T-lymphocytes from death and supports their proliferation.

*Acknowledgements:* This study was supported by the budget project of the Research Institute of Clinical and Experimental Lymphology – Branch of the ICG SB RAS (FWNR-2022-0009).

# The current status of mechanosensitive molecular interactions in atherogenic regions of the arteries: development of atherosclerosis

Mishchenko E.<sup>1</sup>, Ivanisenko V.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *elmish@bionet.nsc.ru*

**Key words:** atherogenesis; shear stress; transcription factor *NF-κB*; mechanosensitive receptors; cell adhesion molecules; signaling pathways; mechanotransduction

*Motivation and Aim:* A terrible disease of the cardiovascular system – atherosclerosis, develops in the areas of bends and branches of arteries, where the direction and modulus of the blood flow velocity vector changes and, consequently, the mechanical effect on endothelial cells in contact with the blood flow [1]. The aim of this work focuses on topical researches on the development of atherosclerosis – mechanobiochemical events that transform the pro-atherogenic mechanical stimulus of blood flow – low and low/oscillatory arterial wall shear stress in the chains of biochemical reactions in endothelial cells, leading to the expression of specific proteins that cause the progression of the pathological process.

*Results:* The stages of atherogenesis, systemic risk factors for atherogenesis and its important hemodynamic factor – low and low/oscillatory wall shear stress exerted by blood flow on the endothelial cells lining the arterial walls, interactions of cell adhesion molecules responsible for the development of atherosclerosis under low and low/oscillatory wall shear stress conditions, activation of the regulator of the expression of cell adhesion molecules, the transcription factor *NF-κB* and the factors regulating its activation under these conditions, mechanosensitive signaling pathways leading to the expression of *NF-κB* in endothelial cells have been analyzed [2].

*Conclusion:* Studies of the mechanobiochemical signaling pathways and interactions involved in the progression of atherosclerosis provide valuable information for the development of approaches that delay or block the development of this disease.

*Acknowledgements:* Work 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.

## References

1. Morbiducci U., Kok A.M., Kwak B.R., Stone P.H., Steinman D.A., Wentzel J.J. Atherosclerosis at arterial bifurcations: evidence for the role of haemodynamics and geometry. *Thromb Haemost.* 2016;115(3):484-492. doi: 10.1160/TH15-07-0597.
2. Mishchenko E.L., Ivanisenko V.A. Mechanosensitive molecular interactions in atherogenic regions of the arteries: development of atherosclerosis. *Vavilov J Genet Breed.* 2021;25(5):552-561. doi 10.18699/VJ21.062.

## Lipids lateral diffusion study associated with structural changes in cytoplasmic membranes

Mokrushnikov P.V.<sup>1\*</sup>, Rudyak V.Ya.<sup>1, 2, 3</sup>

<sup>1</sup> Novosibirsk State University of Architecture and Civil Engineering (SIBSTRIN), Novosibirsk, Russia

<sup>2</sup> Institute of Thermophysics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

\* pavel.mokrushnikov@bk.ru

**Key words:** lipid diffusion, structural changes in cytoplasmic membranes, Fourier transform method

*Motivation and Aim:* To date, it has been experimentally established that the diffusion of lipids in cytoplasmic membranes is complex. There are several qualitatively different types of lipid diffusion: (slowed, confined, simple-Brownian, directed and hop diffusion) [1]. This paper presents a model of lipid diffusion in heterogeneous native cytoplasmic membranes. It has been shown that the appearance of various types of lipid diffusion mentioned above is associated with structural changes in cytoplasmic membranes [2–5].

*Methods and Algorithms:* A model of an infinite flat layer with an inhomogeneous, periodic diffusion coefficient is considered. Two cases were studied – one and two sublattices of harmonic dependences of the diffusion coefficient  $D$  on the coordinate:

$$D = D_0 - D_1 \cos(k_0 x) \cos(k_0 y) \quad \text{and} \quad D = D_0 - \tilde{D}_1 \cos(k_0 x) \cos(k_0 y) - \tilde{D}_2 \cos(k_2 x) \cos(k_2 y)$$

where  $D_0, \tilde{D}_1, \tilde{D}_2$  are the first, second and third diffusion coefficients respectively,  $\lambda_1 = 100$  nm is the distance between the nearest membrane proteins that have slightly changed their conformational state,  $\lambda_2$  is the distance between the nearest membrane proteins that have greatly changed their conformational state  $k_0 = 2\pi/\lambda_1, k_2 = 2\pi/\lambda_2$ .

The diffusion equations for the  $\Psi$  distribution function of lipids in an inhomogeneous

membrane  $\frac{\partial \Psi}{\partial t} = \frac{\partial}{\partial x} \left( D(x, y) \frac{\partial \Psi(x, y, t)}{\partial x} \right) + \frac{\partial}{\partial y} \left( D(x, y) \frac{\partial \Psi(x, y, t)}{\partial y} \right)$  have been solved

by the Fourier transform method. An analytical analysis of the solution is given.

*Results:* It is analytically shown that there is an advection-diffusion mode for these models. The alternating modes of nonlinear diffusion and joint advection-diffusion are analytically shown. It is shown that from zero to sometime  $t_0$ , nonlinear diffusion is observed, then there comes a period of advection-diffusion, then again a period of nonlinear diffusion, etc. In the experiment, this is not always observed in the form of hop diffusion. When the drift during advection is less than the dispersion, the advection of lipids is not visible against the background of their chaotic thermal motion; the trajectory of lipids in the membrane corresponds to nonlinear diffusion. Analytically, criteria have been obtained, under which the advection of lipids can be observed in the experiment. It is analytically shown that lipid advection is not observed in the case of a single lattice of inhomogeneity's, since the coefficient of membrane inhomogeneity  $\beta = D_1/D_0$  must be very high ( $\beta > 0.8$ ), which is rarely performed in an experiment. At lower values of the lattice inhomogeneity coefficient,  $\beta$  advection will not be visible against the background of chaotic thermal motion of the lipid.

In the second case, when the membrane proteins associated with the spectrin-actin-ankyrin network change their conformation in different ways, they attract the lipids surrounding them in different ways. The protein-lipid domains formed in this case will be different, the density of the lipid bilayer can be represented as the superposition of two or more sublattices of inhomogeneous lipid density on top of each other, it is shown that the nature of lipid diffusion remains qualitatively the same, as for the model with a single inhomogeneity lattice. But in this case, hop diffusion can already be observed, in which nonlinear diffusion is replaced by periods of advection-diffusion. Advection is local in membrane, concentrated near protein-lipid domains, in which the increase in lipid density is greatest. The average displacement of particles during advection is higher for molecules that started from the region of lipid-lipid interactions. Analytically, a criterion has been obtained, under which hop diffusion can be observed in the experiment:

$$\frac{(\beta_2 + \beta_3 q)^2}{\beta_2 + \beta_3 q^2} \geq 4(1 - \beta_2 - \beta_3)$$

where  $\beta_2 = \tilde{D}_1/D_0$  is the second lattice inhomogeneity coefficient, the third lattice inhomogeneity coefficient  $\beta_3 = \tilde{D}_2/D_0$  and the ratio of the wave numbers of the two sublattices  $q = k_2/k_0$ .

*Conclusion:* This model explains well on a qualitative level the experimental results of measuring the movement of lipids in the cytoplasmic membrane. It has been shown that the lateral diffusion of lipids in the cytoplasmic membrane is determined by structural changes in the membrane, the appearance or disappearance of a stationary periodic network of protein-lipid domains associated with the cytoskeleton.

*Acknowledgements:* The paper was supported by the mega grant of the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-575) and Russian Foundation of Basic Research (Grant No. 20-01-00041).

### References

1. Fujiwara T.K., Iwasawa K., Kalay Z., Tsunoyama T.A. et al. Confined diffusion of transmembrane proteins and lipids induced by the same actin meshwork lining the plasma membrane. *Mol Biol Cell.* 2016;27(7):1101-1119.
2. Mokrushnikov P.V., Panin L.E., Panin V.E., Kozelskaya A.I., Zaitsev B.N. Structural transitions in erythrocyte membranes (experimental and theoretical models). Novosibirsk: Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), 2019. (In Russian).
3. Mokrushnikov P.V. Mechanical Stresses in the Lipid Bilayer of Erythrocyte Membranes. Lipid bilayers. Properties, behavior and interactions. New York: Nova Science Publishers, Inc. (USA), 2019.
4. Mokrushnikov P.V. Mechanical stresses in erythrocyte membranes (Theoretical Models). *Biophysics.* 2017;62(2):256-260.
5. Mokrushnikov P.V., Dudarev A.N., Tkachenko T.A., Gorodetskaya A.Y., Usynin I.F. Effects of native and oxidized apolipoprotein A-I on lipid bilayer microviscosity of erythrocyte plasma membrane. *J Biochemistry (Moscow). Supplement Series A: Membrane Cell Biol.* 2017;11(1):48-53.

## Study of platelet activation in a system continuously releasing nitric oxide

Moskalensky A.E. \*, Karogodina T.Yu., Panfilov M.A., Vorob'ev A.Yu.

*Novosibirsk State University, Novosibirsk, Russia*

\* a.mosk@nsu.ru

**Key words:** platelet activation, nitric oxide

*Motivation and Aim:* Platelet activation is considered as a cornerstone in pathogenesis of cardiovascular disease. Laboratory assessment of platelet function is usually done using blood samples, for instance, in platelet-rich plasma [1]. However, the behavior of platelets in physiological conditions is different, mainly because activation is constantly inhibited. The main inhibitor is nitric oxide (NO), which is continuously released by endothelial cells.

*Methods and Algorithms:* NO is short-lived metabolite, therefore it should be continuously generated to maintain constant concentration and to mimic physiological conditions. We used specially synthesized molecules capable of light-controlled release of NO [2], or “NO photodonors”. This approach allows us to tune the rate of NO generation by changing the light intensity. Specifically, we use N-Nitroso derivatives of BODIPY dyes [3].

*Results:* NO photodonor was added to the sample containing platelets loaded with Fluo-4 calcium indicator. The sample was studied using fluorescent microscope. 490 nm source was used for the excitation of Fluo-4 fluorescence and at the same time for activation of NO photorelease. We observed significantly reduced activation compared to the control sample and to the sample with BODIPY dye without N-Nitroso bond.

*Conclusion:* Our results suggest that the use of light-controlled nitric oxide donors allows one to mimic physiological conditions and reduces the spontaneous activation of platelets.

*Acknowledgements:* The study was supported by the Russian Science Foundation (grant No. 18-15-00049).

### References

1. Spiryova D.V., Vorobev A.Y., Klimontov V.V., Koroleva E.A., Moskalensky A.E. Optical uncaging of ADP reveals the early calcium dynamics in single, freely moving platelets. *Biomedical Optics Express*. 2020;11(6):3319-3330.
2. Dranova T.Y., Vorobev A.Y., Pisarev E.V., Moskalensky A.E. Diaminorhodamine and Light-Activatable NO Donors: Photorelease Quantification and Potential Pitfalls. *J Fluoresc*. 2021;31(1):11-16.
3. Vorobev A.Y., Moskalensky A.E. Long-wavelength photoremovable protecting groups: On the way to *in vivo* application. *Comput Struct Biotechnol J*. 2020;18:27-34.

## Human water model: interstitium and meridians of traditional Chinese medicine

Nebrat V.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia  
academvlad@yandex.ru*

**Key words:** interstitium, water structures, acupuncture meridians, human model

*Motivation and Aim:* It is well known that the human body is a complex and still insufficiently studied biosystem. However, the network that runs through all the tissues of the interstitium in the human body that has been found quite recently became a complete surprise for the scientific community [1]. This network consists of interconnected visible structures – microcavities containing interstitial fluids, in which the water content is 99 % of the volume. These structures are constantly contracting and expanding as water moves through the network. This characterizes the interstitium network as a dynamic water network. The purpose, structure and distribution of the network in the human body are still unclear. Along with this, there are hypotheses that suggest that the network can serve as the anatomical basis of either a previously unknown organ or a long-known network of acupuncture meridians of traditional Chinese medicine [2]. The presence of an interstitium water network indicates the need to build new models of the human body, taking into account contemporary ideas about water and its role in biosystems. Previously, the author developed a human water model, which takes into account the processes of water self-organization in biosystems [3]. Thus, the purpose of this work is to present the formalized foundations of the human water model and their application to explain the structure of the interstitium network as well as to elucidate the possibility of the network to serve as an anatomical basis for the network of acupuncture meridians.

*Methods and Algorithms:* When presenting the basics of the human water model, we considered human body consisting of almost 70 % of water in a simplified way as a water environment without taking into account electrolytes and organic components of this environment. It was believed that its structural arrangement is determined by the processes of self-organization of water, which for different water systems occur according to their own rules. Therefore, we used the rules of water self-organization in biosystems, which were first established using the method of generalized crystallography [4]. According to these rules, a scheme of phased self-organization of the water environment was built. At the beginning, the formation of five simplest water clusters from water molecules was considered, in the form: dimer, trimer, tetramer, pentamer, and hexamer, from which five simplest dissipative structures consisting of clusters of the same type were further formed. Then these structures were combined into cyclic dissipative structures, from which subsequent fractal structures of the water environment of a higher hierarchy level were further formed. This scheme was formalized in the form of a graph of the water environment, similar to the Petersen graph. The graph was analyzed in order to identify the mechanism that ensures the dynamics of self-organization of the water environment. Next, the medium was divided, as is customary in physiology, into two parts: intracellular and extracellular, and the found



mechanism was used to identify the network outside the cellular part of the water environment. They built a graph of this network, analyzed its structure and properties, and the paths of the graph were interpreted as water channels of the interstitium and used to explain the networks of the interstitium and meridians.

*Results:* The structure of the interstitium network the model explains how the body's water circulation system consists of interconnected 12 main and 36 secondary water channels of three types. This system is characterized by three essential properties. The first property is due to the fact that in the system there is a directed movement of charged ions  $H^+$  and  $OH^-$ , which form ionic (electric) currents of two polarities, circulating through the water channels of the network simultaneously in two opposite directions. Therefore, this property manifests itself in a constant change in the direction of the total current of the network, depending on the current ratio of the contributions of the two ionic currents. The second property is due to the fact that moving ion currents excite two fields around the water channels – electric and magnetic. This property is manifested in the arrangement of the channels of the water environment in two perpendicular directions of these fields. Moreover, in each of these directions, the interaction of water channels will obey the Ampere law: parallel channels with currents of the same polarity, the networks flowing in one direction will attract, and in the opposite direction they will repel. The third property is due to the fact that all water channels are grouped into three groups within the network, which has a rotational symmetry of the third order. These three groups of channels can be considered as analogues of "sensors" of electromagnetic fields, which are arranged in a similar way and contain the same working substance in the form of water molecules and  $H^+$  and  $OH^-$  ions. It follows that the network will have three spectral windows of absorption and emission of electromagnetic energy. Moreover, each window will have two resonant frequency peaks that are multiples of the resonant frequencies of the molecular vibrations of the  $H^+$  and  $OH^-$  ions. Comparison of the models of the interstitium network and the meridian network reveals an almost complete coincidence of their structure. It makes it possible to count the interstitium as the anatomical basis of the meridian network of traditional Chinese medicine.

*Conclusion:* The results obtained in this paper confirm that the human water model and the Chinese medicine model are similar despite their different terminology. Both models describe the same process of accumulation and removal of  $H^+$  ions found in the model, which is realized in the water environment of the interstitium of the human body. The importance of the human water model being for medicine is due to its new content. It allows clinicians to take a fresh look at the causes of many human diseases associated with disorders in the structure and dynamics of the body's water environments.

### References

1. Benias P.C. et al. Structure and distribution of an unrecognized interstitium in human tissues. *Sci Rep.* 2018;8:4947.
2. Tomov N. et al. Is the newly described interstitial network the anatomical basis of acupuncture meridians? A Commentary. *Anat Rec (Hoboken)*. 2020;303(8):2169-2170. doi: 10.1002/ar.24114.
3. Nebrat V.V. Thermodynamic model of human water self-organization. *J Problems Evol Open Syst.* 2008;1(10):118-122.
4. Bulienkov N.A. On the possible role of hydration as a leading integration factor in the organization of biosystems at various levels of their hierarchy. *Biophysics*. 1991;36(2):181-243.

## **The low-molecular-weight ligands of human serum albumin, promoting its interaction with amyloid $\beta$ peptide**

Nemashkalova E.\*, Deryusheva E., Kazakov A., Shevelyova M., Permyakov S., Litus E.

*Institute for Biological Instrumentation, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Moscow Region, Russia*

*\*elnemashkalova@gmail.com*

One of the functions of human serum albumin (HSA) in blood and cerebrospinal fluid is a sequestration of amyloid  $\beta$  peptide ( $A\beta$ ), the important participant of Alzheimer's disease (AD) progression. The therapeutically advantageous enhancement of  $A\beta$  trapping can be achieved via directed improvement of HSA affinity to  $A\beta$ . Previously, we demonstrated that arachidonic/linoleic acids (ArA/LA) favor  $A\beta$  binding to HSA. In this study, we explored a panel of the low-molecular-weight HSA ligands associated with AD (ibuprofen (IBU), serotonin (SRO) and tryptophan (TRP)) with regard to their ability to affect the HSA- $A\beta$  interaction. Molecular modelling predicts an overlap between the HSA-binding sites of IBU and ArA in the region of fatty acid binding site 3 (FA3). Similarly, the intersection between the  $A\beta$ -binding sites of HSA and its sites for ArA (FA6), SRO, TRP, IBU binding suggests their influence on HSA- $A\beta$  interaction. The equilibrium and kinetic parameters of HSA interaction with monomeric  $A\beta_{40/42}$  in the presence/absence of the ligands was studied by surface plasmon resonance spectroscopy. The therapeutic IBU level decreases the equilibrium dissociation constant for the HSA- $A\beta_{40/42}$  complexes by a factor of 3-5. SRO increases HSA affinity to  $A\beta_{40/42}$  by a factor of 7/17, whereas the structurally homologous TRP is inactive in this sense. Overall, we have shown high potency of the low-molecular-weight ligands to promote HSA interaction with  $A\beta$ , and laid the foundation for rational search of the substances with therapeutic potential in the treatment of AD.

*Acknowledgements:* Russian Science Foundation, grant No. 20-74-10072 (Litus E.).

# Novel aptamer-based targeted agents for boron neutron capture therapy

Novopashina D.<sup>1</sup>, Dymova M.<sup>1</sup>, Kuligina E.<sup>1</sup>, Silnikov V.<sup>1</sup>, Meschaninova M.<sup>1</sup>, Taskaev S.<sup>2,3</sup>, Richter V.<sup>1</sup>, Vorobyeva M.<sup>1\*</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [maria@vorobjeva.ru](mailto:maria@vorobjeva.ru)

**Key words:** cell-specific aptamers, boron clusters, boron neutron capture therapy, cancer treatment

*Motivation and Aim:* Glioblastoma is the most frequent malignant brain tumor which still remains incurable due to its rapid growth, invasive nature, and resistance to conventional therapies. Boron neutron capture therapy (BNCT) represents a promising radiotherapeutic approach to the treatment of malignant tumors in general and glioblastoma in particular [1]. Successful BNCT needs a boron-containing therapeutic agent for addressed delivery of <sup>10</sup>B isotope inside cancer cells, and a source of epithermal neutrons to irradiate and kill boron-containing targets. Polyhedral boranes, namely boron clusters, are very attractive for the purposes of BNCT since they combine high boron content with relative metabolic inertness and low toxicity. Otherwise, cell-specific nucleic acid aptamers seem to be prospective candidates for carrying <sup>10</sup>B to tumor cells. Here, we aimed to evaluate the potential of aptamers specific to human glioblastoma cells as boron delivery agents for BNCT.

*Methods and Algorithms:* The study was carried out using the human glioblastoma cell line U-87 MG as target cells and normal human fibroblasts hFF8 as controls. We formed a set of fluorescently labeled 2'-F-RNA and DNA aptamers that were reported to internalize into the U-87 MG human glioblastoma cells. Their cell penetration was assessed by confocal microscopy. The *closo*-dodecaborate residue was attached to the 5'-end of the aptamer through the click reaction between the 5'-alkyne-modified aptamer and azide-containing derivative of *closo*-dodecaborate [2]. Cell toxicity of aptamer conjugates was examined by the MTT test. Model BNCT experiments *in vitro* were performed on the Tandem-BNCT neutron source (Budker Institute of Nuclear Physics, Novosibirsk, Russia). Their results were assessed by two independent methods: real-time cell analysis and clonogenic assay.

*Results:* Two 2'-F-RNA aptamers demonstrated specific internalization into glioblastoma U-87 MG glioblastoma cells and entered cell nuclei. *Closo*-dodecaborate conjugates of the aptamers possessed low cell toxicity and showed the same intracellular localization as the parent aptamers. A pre-treatment of the cells by boron-containing aptamer conjugates resulted in the specific decrease of tumor cells viability after neutron irradiation. The effect was comparable to that of <sup>10</sup>B-boronophenylalanine taken as a control.

*Conclusion:* We demonstrated for the first time the specific inhibition of cancer cell proliferation in model BNCT experiments with 2'-F-RNA aptamers loaded by the *closo*-dodecaborate. Taking into account their target specificity, ease of synthesis and a wide

range of chemical approaches for high boron-loading, aptamers represent a very promising basis for engineering novel BNCT agents.

*Acknowledgements:* The research was supported by Russian Scientific Foundation (grant No. 19-74-20127).

### *References*

1. Dymova M.A. et al. Boron neutron capture therapy: Current status and future perspectives. *Cancer Commun.* 2020;40(9):406-421.
2. Meschaninova M.I. et al. Novel convenient approach to the solid-phase synthesis of oligonucleotide conjugates. *Molecules.* 2019;24(23):4266.

## ER stress in mice prefrontal cortex pyramidal neurons in peripheral tumor growth

Obanina N.A.<sup>1,2\*</sup>, Bgatova N.P.<sup>1</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

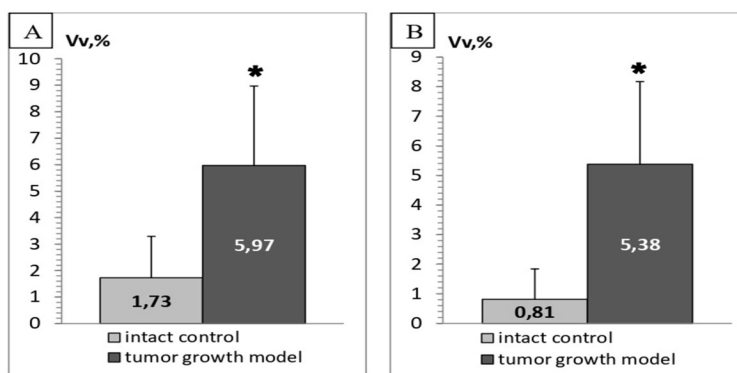
\* [n.obanina@g.nsu.ru](mailto:n.obanina@g.nsu.ru)

**Key words:** ER stress, pyramidal neurons, ultrastructure

*Motivation and Aim:* Endoplasmic reticulum (ER) plays an important role in the synthesis, processing and transport of many proteins, the biosynthesis of membrane lipids, the regulation of carbohydrate metabolism, the maintenance of calcium homeostasis, and other important cellular processes. Under pathological conditions leading to oxidative stress, including oncology, protein folding is disrupted and ER stress is triggered [1]. Purpose: to investigate the structure of endoplasmic reticulum of prefrontal cortex pyramidal neurons under conditions of peripheral tumor growth.

*Methods and Algorithms:* For tumor induction, cultured B16 cells were subcutaneously injected into the right inguinal area of mice. The material was sampled at 7 days. Transmission electron microscopy was used to examine the ultrastructural organization of prefrontal cortex pyramidal neurons in brain of C57BL/6 male mice from two groups: a tumor growth model and an intact control. Ultrathin sections thick 70–100 nm were studied using an microscope electron JEM 1400 (Japan). Morphometry was performed using Image J software (Wayne Rasband, USA) with an opened test system at x30K magnification. The mean (M) and standard deviation (SD) were calculated using Microsoft Excel software. The significance of differences between the parameters was determined by Statistica 6.0 software (StatSoft, Inc.) with Mann-Whitney U-test. The differences were considered significant at  $p < 0.05$ .

*Results:* Morphometric analysis of neurons in animals with a tumor revealed a significant increase in the volume density of the ER with widening of its cisterns and tubules ( $V_v = 5.97\%$ ), which is a structural sign of ER stress (Fig. 1, A). In addition, in the neurons of animals with a tumor, a high level of volume density of mitochondria with destructive changes of cristae was noted ( $V_v = 5.38\%$ ), this is 46.6 % of the total number of mitochondria (see Fig. 1, B). It is known that ER and mitochondria are functionally related to each other; they provide exchange of calcium, delivery of reactive oxygen species, and lipid sorting [2]. Cellular stress caused by external or internal signals activates processes aimed either at restoring cellular homeostasis or at cell death. These processes include unfolded protein response (UPR), autophagy, hypoxia, and mitochondrial function, which are part of the global ER response to stress [3]. Activation of autophagy limits the inflammatory response and provides cytoprotection by reducing ER stress [1]. Decreased levels of autophagy result in disruption of the intracellular homeostasis of neurons.



**Fig. 1.** Volume density of organelles in prefrontal cortex pyramidal neurons. A –  $V_V$  of ER with widening of its cisterns and tubules; B –  $V_V$  of mitochondria with destructive changes of cristae

*Conclusion:* The data obtained indicate an increase in ER stress in conditions of peripheral tumor growth. Further study of the structural features of pyramidal neurons will contribute to a more detailed understanding of the effect of peripheral tumor growth on them, and will also help to identify signaling pathways and targets for therapeutic use.

*Acknowledgements:* The study is supported by Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS (NFWNR-2022-0012).

### References

1. Chipurupalli S., Samavedam U., Robinson N. Crosstalk between ER Stress, Autophagy and Inflammation. *Front Med (Lausanne)*. 2021;8:758311. doi: 10.3389/fmed.2021.758311.
2. Stacchiotti A., Favero G., Lavazza A., Garcia-Gomez R., Monsalve M., Rezzani R. Perspective: Mitochondria-ER Contacts in metabolic cellular stress assessed by microscopy. *Cells*. 2018;8(1):5. doi: 10.3390/cells8010005.
3. Senft D., Ronai Z.A. UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. *Trends Biochem Sci*. 2015;40(3):141-148. doi: 10.1016/j.tibs.2015.01.002.

## Digital medicine education problems and online solutions

Orlov Y.L.<sup>1,2\*</sup>, Orlova N.G.<sup>3,4</sup>

<sup>1</sup> *The Digital Health Institute, I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

<sup>3</sup> *Financial University under the Government of the Russian Federation, Moscow, Russia*

<sup>4</sup> *Moscow State Technical University of Civil Aviation, Moscow, Russia*

\* [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** education, digitalization, bioinformatics, databases, e-Health

*Motivation and Aim:* Education in medicine base on modern internet-technologies is a priority for contemporary organization of medical care. Distant education in recent years raised series of problems in the materials availability, access by the students. The epidemic situation of the last two years has only accelerated the adoption of measures for remote organization of medical consultations; these problems have been discussed both at the RAS and at events held at Sechenov University in Moscow, at scientific and educational conferences [1]. Bioinformatics as a modern scientific discipline meets more and more applications in biology, medicine, requires re-training of the specialists. Of special interest is the computerization and automation of teaching informatics-related disciplines, including medicine – in the areas of telemedicine, e-Health [2, 3].

*Methods and Algorithms:* The issue of developing bioinformatics courses is related to the adaptation of the training to the educational profile of students and trainees [4]. Thus, according to the experience of the work of teaching bioinformatics students – mathematicians require not only a different presentation of the materials, but also the methodology itself, in contrast to students in medical and natural students who do not have enough skills in computer science, programming, and writing their own code. The assessment of students' skills is given on the basis of personal teaching experience in Novosibirsk State University for biologists of the Natural Sciences Department, for students of pharmacy specialization "Bioengineering and bioinformatics" at Sechenov University, and at PFUR (Peoples' Friendship University of Russia), as well as for the students of the Financial University under the Government of Russian Federation in Moscow.

The situation with the frequent transition to distance learning in Russia from 2020 required the development of new methods of teaching, with the possibility for students to use Internet resources and do independent work without access to a stationary computer class, in remote mode. The authors proposed a training course on basics bioinformatics, the use of elements of which began at Novosibirsk State University in the early 2000s (<http://lcg.nsu.ru/ru/home-2/>), then was adapted for students at Sechenov University in Moscow (<https://student.sechenov.ru/>), and later, since 2021, at the Agrarian and Technological Institute of PFUR (<https://www.rudn.ru/education/schools-and-departments/institutions/agricultural-technology-institute>) [5].

*Results:* The bioinformatics course was read for students at Sechenov University in Moscow; it was presented online only on the university's internal website (<https://student.sechenov.ru/>). Let us note a number of qualitatively new educational

tasks in the field of digital health, such as the use of blockchain technologies, the use of Artificial Intelligence (AI) methods in supporting medical decision-making [2, 3, 6].

There is great interest in the development of such a bioinformatics course from a number of universities in Russia. For replication, video tutorials need to be recorded and examples of scientific problems to be adapted to different directions. A project has been formed which was used in applying for grants to the Russian Science Foundation and the Potanin Foundation (<https://zayavka.fondpotanin.ru/ru/>). The concept of the proposed educational course includes a theoretical part (listening to the course in the form of lectures, video lessons) and a practical part – performing tasks on the use of computer programs and databases, to solve practical tasks, in written form, which are evaluated by the teacher and can be of independent value as reference scientific papers. Course content includes the fundamentals of bioinformatics databases, online tools, and web services for medical consultations. Successful examples of published students work are presented [5, 6].

*Conclusion:* Overall, the goal of the project is to teach digital specialties, bioinformatics in medicine using online and hybrid approaches. As already noted, the issue of developing bioinformatics courses is related to the need to adapt training depending on the educational profile of the trainees. The transition to distance learning requires the development of new methods of teaching, with the ability of students to perform independent work using only Internet resources, which in turn meets a number of challenges due to limited access to a number of online services in Russia (for example, data storage in Google cloud storages) [7].

*Acknowledgements:* The study is supported by the Potanin Foundation grant for masters' teachers (ГК22-000797).

## References

1. Orlov Y.L., Bakulina A.Yu. Development of education in bioinformatics based on student conferences ISSC-2018, School of molecular modeling and hackathon in Novosibirsk. *Vestnik Novosibirsk State University. Series: Information Technologies*. 2018;16(3):5-6 (in Russian).
2. Lebedev G. et al. Systematization of the principles and methods of applying for digital medicine in oncology. *Procedia Computer Sci*. 2021;192:3214-3224. doi: 10.1016/j.procs.2021.09.094.
3. Koshechkin K., Lebedev G., Radzievsky G., Seepold R., Martinez N. Review of blockchain technology projects to provide telemedical services. *J Medical Internet Res*. 2021;23(8):e17475. doi: 10.2196/17475.
4. Orlova N.G., Orlov Y.L. Problems of developing online training courses for students in digital disciplines using bioinformatics as an example In: Proceedings of the International Conference Scientific research of the SCO countries: synergy and integration. Part 3 – Reports in English. March 31, 2022. Beijing, PRC. Scientific publishing house Infinity. 2022:58-65.
5. Dergilev A.I., Orlova N.G., Dobrovolskay O.B., Orlov Y.L. Statistical estimates of transcription factor binding site clusters in plant genomes based on genome-wide data. *J Integr Bioinform*. 2021;18:20200036. doi: 10.1515/jib-2020-0036.
6. Orlov Y.L. et al. Reconstruction of the Parkinson's disease gene network to search for target genes. *Biomed Chem*. 2021;67(3):222-230. doi:10.18097/PBMC20216703222 (in Russian).
7. Gusev A.V., Ivshin A.A., Vladzimirsky A.V. Russian mobile health apps: systematic search in app stores. *Russian J Telemedicine E-Health*. 2021;7(3):21-31. doi: 10.29188/2712-9217-2021-7-3-21-31 (in Russian).



## Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients

Oshchepkov D.<sup>1\*</sup>, Chadaeva I.<sup>1</sup>, Kozhemyakina R.<sup>1</sup>, Zolotareva K.<sup>1</sup>, Khandaev B.<sup>1</sup>, Sharypova E.<sup>1</sup>, Ponomarenko P.<sup>1</sup>, Bogomolov A.<sup>1</sup>, Klimova N.<sup>1</sup>, Shikhevich S.<sup>1</sup>, Redina O.<sup>1</sup>, Kolosova N.<sup>1</sup>, Nazarenko M.<sup>2</sup>, Kolchanov N.<sup>1</sup>, Markel A.<sup>1</sup>, Ponomarenko M.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*

\* [diman@bionet.nsc.ru](mailto:diman@bionet.nsc.ru)

**Key words:** human, hypertension, stress reactivity, molecular marker, *Rattus norvegicus*, RNA-Seq, qPCR, differentially expressed gene, meta-analysis, correlation, principal component, bootstrap

Although half of hypertensive patients have hypertensive parents, known hypertension related human loci identified by genome-wide analysis explain only 3 % of hypertension heredity. Therefore, mainstream transcriptome profiling of hypertensive subjects addresses differentially expressed genes (DEGs) specific to gender, age, and comorbidities in accordance with predictive preventive personalized participatory medicine treating patients according to their symptoms, individual lifestyle, and genetic background. Within this mainstream paradigm, here, we determined whether, among the known hypertension-related DEGs that we could find, there is any genome-wide hypertension theranostic molecular marker applicable to everyone, everywhere, anytime. Therefore, we sequenced the hippocampal transcriptome of tame and aggressive rats, corresponding to low and high stress reactivity, an increase of which raises hypertensive risk; we identified stress-reactivity related rat DEGs and compared them with their known homologous hypertension-related animal DEGs. This yielded significant correlation between stress reactivity-related and hypertension-related fold changes (log<sub>2</sub> values) of these DEG homologs. We found principal components, PC1 and PC2, corresponding to a half-difference and half-sum of these log<sub>2</sub> values (Figure 1). Using the DEGs of hypertensive versus normotensive patients (as the control), we verified the correlations and principal components. This analysis highlighted downregulation of  $\beta$ -protocadherins and hemoglobin as whole-genome hypertension theranostic molecular markers associated with a wide vascular inner diameter and low blood viscosity respectively.

*Acknowledgements:* This study was supported by the Russian Science Foundation (project No. 19-74-10041). We are also thankful to the Multi-Access Center “Bioinformatics” for the use of computational resources as supported by Russian government project FWNR-2022-0020. We are appreciative of the Center for Genomic Research at the ICG SB RAS, where RNA-Seq was carried.

## Trilateral relationship: food-borne trematodes, microbiota, and experimental host

Pakharukova M.Y.<sup>1,2\*</sup>, Lishai E.<sup>1,2</sup>, Zaparina O.<sup>1</sup>, Sripa B.<sup>3</sup>, Hong S.-J.<sup>4</sup>, Mordvinov V.A.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> WHO Collaborating Centre for Research and Control of Opisthorchiasis (Southeast Asian Liver Fluke Disease), Tropical Disease Research Center, Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

<sup>4</sup> Department of Medical Environmental Biology, Chung-Ang University College of Medicine, Seoul, Korea

\* pmaria@yandex.ru

**Key words:** Food-borne trematodes, experimental model, parasite microbiome, *Helicobacter*

**Background:** Three epidemiologically significant food-borne trematodes (*Opisthorchis felineus*, *O. viverrini*, *Clonorchis sinensis*) have some similarities and differences. The definitive host are fish-eating mammals including humans. The flukes affect the hepatobiliary system inducing cholangitis, and bile duct neoplasia, and even cholangiocarcinoma among chronically infected individuals. Differences primarily concern the geographical range and main foci, where they are endemic to, and the carcinogenic potential to human. Two species, *O. viverrini* and *C. sinensis* are both recognized 1A group of biological carcinogens to human, whereas *O. felineus* is recognized to be potentially carcinogenic to animals. The mechanisms of carcinogenesis by the liver flukes are studied fragmentarily, the role of host and parasite microbiome is an unexplored aspect.

**Methods and Algorithms:** In order to characterize the microbial communities in adult parasites as well as in host bile and colon, the hamsters were infected with metacercariae of *C. sinensis* (South Korea), *O. viverrini* (Thailand) and *O. felineus* (Russia). We performed high-throughput sequencing (MiSeq, Illumina) of libraries constructed from V3 – V4 region of 16S ribosomal DNA isolated from adult worms and from colon faeces and bile from the hamsters. Furthermore, *ureA* and *cagA* genes of *Helicobacter pylori* were assessed by the real-time PCR method in the stomach, feces and bile of hamsters.

**Results:** As a result, 1,784,000 reads were assigned to 13,244 operational taxonomy units (OTUs) and, in turn, to 273 genera of Bacteria. Analysis revealed the significant phylogenetic diversity of the microbial communities among three liver flukes. Numerous bacterial species were identified in the bile of the infected animals, in particular, bile contains the same bacterial phyla as worms do. Prevalence of *H. pylori* and *ureA* gene copy number was significantly higher in the liver fluke-infected hamsters than in the uninfected ones.

**Conclusion:** The infection with any liver fluke significantly modified the bile and faecal microbiome, increasing the abundance of *H. pylori*. Mechanisms of host microbiome modification by the liver flukes are discussed.

**Acknowledgements:** This work was supported by the Russian Science Foundation (No. 22-24-20010).

# Nucleoside analogues with photoremovable protecting groups for efficient light-controlled inhibition of viral infection

Panfilov M.A., Vorob'ev A.Yu., Moskalensky A.E.\*

Novosibirsk State University, Novosibirsk, Russia

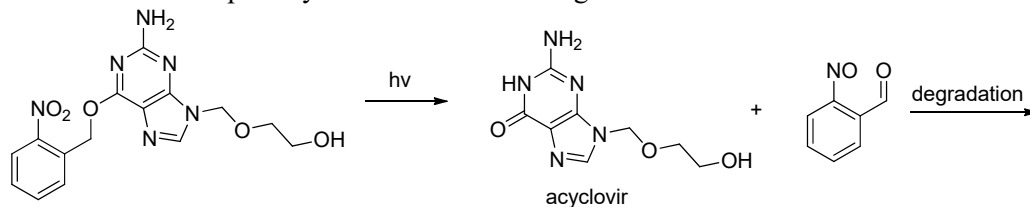
\* a.mosk@nsu.ru

**Key words:** nucleoside analogues, acyclovir, caged acyclovir

*Motivation and Aim:* Recent covid-19 pandemic showed that the development of antiviral drugs is of great importance. Many antiviral drugs are nucleoside analogues. They undergo phosphorylation inside cells and then inhibit the synthesis of viral nucleic acids. To begin its action, the nucleoside analogue should be able to enter the cell. It is often a bottleneck for drugs of this family.

*Methods and Algorithms:* Our idea is to conjugate the nucleoside analogue with the photoremovable protecting group (PPG, [1]), which renders the compound cell-permeable. Then the group can be dissociated by light inside cell, thereby liberating the antiviral drug. Apart from delivery into the cytoplasm, our approach will enable on-demand, local *in vivo* application. In this work, we used nitrobenzyl PPG responsive to UV light, but we also plan to use BODIPY-based PPG which efficiently absorbs green light.

*Results:* As a model compound, we used acyclovir in place of nucleoside analogue. It has been conjugated with PPG using the Mitsunobu reaction. The resulting compound and the scheme of photolysis are shown in the figure.



The presence of free acyclovir after photolysis was confirmed using the HPLC technique.

*Conclusion:* Our results indicate that the approach is viable. Further research will be aimed at the conjugation of visible-light PPGs with anti-coronavirus drugs such as molnupiravir, favipiravir and GS-441524 (metabolite of remdesivir).

*Acknowledgements:* The study was supported by the grant of the President of the Russian Federation for state support of young Russian scientists MK-4020.2021.1.2, agreement No. 075-15-2021-453.

## References

1. Vorobev A.Y., Moskalensky A.E. Long-wavelength photoremovable protecting groups: On the way to *in vivo* application. *Comput Struct Biotechnol J.* 2020;18:27-34.

## **Curdlan/Chitosan/Ag NPs foams for healing chronic wounds**

Permyakova E.S.<sup>1\*</sup>, Shtansky D.V.<sup>1</sup>, Manakhov A.M.<sup>2</sup>, Solovieva A.O.<sup>2</sup>

<sup>1</sup> *National University of Science and Technology “MISiS”, Moscow, Russia*

<sup>2</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *permyakova.elizaveta@gmail.com*

**Key words:** superabsorbent dressings, diabetic wound regeneration, curdlan, chitosan, silver nanoparticles

*Motivation and Aim:* The healing of burns and wounds is one of the most common health problems. Metabolic and physiological disorders affect the normal process of skin reparation, resulting in ulcers, bedsores, and amputations. To enhance rate of healing process it is necessary to provide good oxygenation, sterility, and deletion excessive exudate.

*Methods and Algorithms:* We developed Ag-contained foams based on curdlan and chitosan polymers (CUR/CS/Ag). To obtain a foam-like structure mixture of polymers with Ag NPs was subjected to a lyophilization process. The structure and chemical composition of materials were studied by optic/scanning electron microscopy and XPS/FT-IR analysis respectively. The defects on the back of Mice with genetically determined type 2 diabetes mellitus were covered with either curdlan/chitosan or curdlan/chitosan/Ag dressings for 10 days. After that, the wounds were photographed until one of the wounds healed (24 days). The histology analysis of the wound after 24 days was also obtained.

*Results:* The wounds treated with CUR/CS/Ag showed complete healing after 24 days, which was accompanied by the formation of new vessels of different diameters and significant re-epithelialization. At the same time, CUR/CS-induced healing was only 20 %, and the untreated control showed an increase in the area of the wound by 20+ %.

*Conclusion:* The developed CUR/CS/Ag foams could meet the needs for clinical practice and may have future medical applications for wound care especially in diabetic patients.

*Acknowledgements:* This work was supported by the Russian Science Foundation (grant No. 18-75-10057). Part of this work (SEM analyzes) was carried out during the implementation of the strategic project, “Biomedical materials and bioengineering”, within the framework of the Strategic Academic Leadership Program “Priority 2030” at NUST “MISiS”.

## Role of HSP70 proteins in regulation of mitochondrial protein content in skeletal muscle

Popov D.V.<sup>1,2\*</sup>, Zgoda V.G.<sup>3</sup>, Kukushkina I.V.<sup>1,2</sup>, Gazizova G.R.<sup>4</sup>, Shagimardanova E.I.<sup>4</sup>, Gusev O.A.<sup>4</sup>, Lysenko E.A.<sup>1</sup>, Vinogradova O.L.<sup>1</sup>, Makhnovskii P.A.<sup>1</sup>

<sup>1</sup> Institute of Biomedical Problems, RAS, Moscow, Russia

<sup>2</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> V.N. Orekhovich Research Institute of Biomedical Chemistry, Moscow, Russia

<sup>4</sup> Institute of Fundamental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russia

\* danil-popov@yandex.ru

**Key words:** skeletal muscle, proteome, transcriptome

**Motivation and Aim:** At a normal level of physical activity, skeletal muscles intensively oxidize carbohydrates and fats, playing a key role in the regulation of carbohydrate and fat metabolism, and insulin sensitivity in the organism. A decrease in the content of mitochondria in skeletal muscles and their oxidative capacity is one of the factors in the development of metabolic disorders. Increased physical activity effectively stimulates mitochondrial biogenesis in muscles, thereby positively regulating carbohydrate and fat metabolism. We aimed to study the specificity of the regulation of mitochondrial biogenesis in skeletal induced by regular physical exercise and the role of HSP70 family proteins in the regulation of the content of mitochondrial proteins.

**Methods and Algorithms:** In a study with volunteers, the effects of 2-month physical training on the content of proteins (shotgun quantitative [with isobaric labeling] proteomic) in skeletal muscle (*m. vastus lateralis*) and their mRNA expression (RNA sequencing) were studied. In an experiment on *D. melanogaster* line *Hsp70-(Df(3R)Hsp70A, Df(3R)Hsp70B)*, we studied the effects of knockout of 6 genes (out of 9) of the HSP70 heat shock protein family on the performance (climbing speed) of flies and the content of mitochondrial proteins in the skeletal muscles of their limbs.

**Results:** In human skeletal muscle, regular physical training increased the content of ~250 proteins (out of 800 detected), including ~100 mitochondrial proteins. Interestingly, a training-induced increase in the content of some proteins (mainly mitochondrial) was not accompanied by an increase in their mRNA level. An analysis of the available databases showed that these proteins (especially mitochondrial proteins) and their mRNAs are highly abundant proteins and mRNAs with superior stability. Additionally, training causes an increase in the content of 20 chaperones (including those from the HSP70 family), proteins that play an important role in the regulation of proteostasis. Knockout of the *Hsp70* genes in *D. melanogaster* caused a decrease in the content of individual mitochondrial proteins in skeletal muscle and the vertical running speed of flies compare to the control line.

**Conclusion:** In skeletal muscle, the increase in the content of mitochondrial proteins induced by increased physical activity is not regulated by the increase in the corresponding mRNAs and, apparently, depends on the rate of mitochondrial protein synthesis and their degradation/stability. The latter, in part, may be associated with the functions of heat shock proteins of the HSP70 family.

**Acknowledgements:** The study was supported by the Russian Science Foundation (grant 21-15-00405).

# Integrated bioinformatics analysis of the gene structure of polygenic diseases to develop effective methods for their early prognosis

Prokofiev V.\*, Shevchenko A.V., Kononkov V.

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

\* [vf\\_prok@mail.ru](mailto:vf_prok@mail.ru)

**Key words:** polygenic diseases, cytokine genes, gene polymorphism, gene network, machine learning (ML), Cytoscape, multifactor dimensionality reduction (MDR)

*Motivation and Aim:* It is impossible to imagine modern medical genetics of polygenic human diseases without the widespread use of intelligent analysis and machine learning methods, in particular, such as simulation of stochastic processes, MDR analysis, bioinformatic analysis of complex biological networks based on graph theory, which allow replacing the original, largely stochastic, object with its image – a mathematical model with further study of this model using computer-implemented computational logic algorithms or heuristic [1–3]. In connection with the above, it is very relevant to conduct comparative clinical and genetic studies with the analysis of combined genotypes in combination with various variants of mathematical modeling, the results of which can become the basis for the development of fundamentally new ways to predict the development of socially significant human diseases, their early diagnosis and prevention.

*Methods and Algorithms:* The study of the features of the genetic structure of cytokines, angiogenesis and vascular remodeling factors in healthy individuals and patients with various forms of diseases – myocardial infarction, breast cancer, rheumatoid arthritis, female infertility of unknown origin and type 2 diabetes mellitus was carried out. 526 healthy and 1074 with various pathology people were examined. The study was conducted using the SNP methods, various variants of data mining and machine learning (simulation modeling, MDR method, bioinformatic analysis of gene networks) [4, 5].

Statistical processing of the results was carried out based on a methodological, including a comprehensive computer analysis of gene chains of various dimensions with an analysis of their structural and functional organization in terms of their possible pathogenetic role in the development of the particular polygenic disease [6]. This approach allowed us to identify highly informative gene ensembles associated with the development of each of the studied diseases.

*Results:* With the help of these genetic complexes and combined methods of mathematical modeling, bioinformatic matrices were created, on the basis of which, for each of the multifactorial diseases, prognostic algorithms with an optimal ratio of sensitivity and specificity indicators were developed, ensuring maximum accuracy of the method (up to 94 %) with a minimum set of prognostic predictors. It is shown that for the construction of prognostic models of polygenic diseases with the greatest predictive potential, it can be effective to use simulation modeling based on a computational algorithm embedded in the modified procedure of sequential recognition A. Wald. To improve the quality of predictive models, it is necessary to use a combined mathematical

modeling method that combines simulation modeling, MDR analysis and graphical visualization of gene networks using the Cytoscape platform.

*Conclusion:* The results of the conducted research can serve as a scientific basis for the development of effective technologies for the prediction, diagnosis and treatment of the polygenic diseases.

*Acknowledgements:* The study was carried out as part of the government assignment (budgetary funding within the framework of the comprehensive research topic No. 0324-2019-0046-C-02 “Development of molecular profiling methods and innovative technologies for prediction, early diagnosis, drug and cell therapy of socially significant human diseases of autoimmune, inflammatory and dysmetabolic nature”).

### *References*

1. Moore J.H., Folkert W.A., Scott M.W. Bioinformatics challenges for genome-wide association studies. *Bioinformatics Rev.* 2010;26:445-455.
2. Lvovs D., Favorova O.O., Favorov A.V. A polygenic approach to the study of polygenic diseases. *Acta Nat.* 2012;4:59-71.
3. Deo R.C. Machine Learning in Medicine. *Circulation.* 2015;132(17):1920-1930.
4. Lance W.H., Marylyn D.R., Moore J.H. Multifactor dimensionality reduction software for detecting gene–gene and gene–environment interactions. *Bioinformatics.* 2003;19(3):376-382.
5. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504.
6. Prokofiev V., Konenkov V., Koroleva E., Shevchenko A., Dergacheva T., Novikov A. The structure of the cytokine gene network in uterine fibroids. In: Proceedings 2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB), 6–10 July 2020. IEEE. 2020;261-264. doi: 10.1109/CSGB51356.2020.9214588.

## Topology of the associative cytokine gene network in primary open-angle glaucoma

Prokofiev V.<sup>1\*</sup>, Shevchenko A.V.<sup>1</sup>, Kononkov V.<sup>1</sup>, Trunov A.<sup>2</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk Branch of The academician S.N. Fyodorov Federal State Institution “Intersectoral Research and Technology Complex Eye Microsurgery”, Ministry of Health of the Russian Federation, Novosibirsk, Russia*

\* [vf\\_prok@mail.ru](mailto:vf_prok@mail.ru)

**Key words:** primary open-angle glaucoma, cytokine genes, gene polymorphism, gene network

*Motivation and Aim:* Primary open-angle glaucoma (POAG) is one of the leading causes of global irreversible blindness. Therefore, it remains a serious medical and social problem throughout the world [1]. There is an assumption that the predisposition to the development of glaucoma is due to polymorphism of the regulatory regions of coding genes that determine the individual level of production of cytokines, chemokines, and growth factors [2, 3]. The aim of our study was to conduct a comprehensive bioinformatics analysis of polymorphic regions of the cytokine network genes to identify the relationship between the features of the topology of this network with the development of POAG.

*Methods and Algorithms:* A genetic study was performed on 199 subjects, including 99 patients with POAG and 100 patients without glaucoma of similar gender and age as a comparison control group. Polymorphisms of the regulatory regions of 10 cytokine genes (*IL1B -31 C/T*, *IL4 -590 C/T*, *IL6 -174 C/G*, *IL8 -251 T/A*, *IL10 -592 C/A*, *IL10-1082 A/G*, *IL17A -197 G/A*, *MMP2-1306 C/T*, *MMP3-1171 5A/6A*, *MMP9-1562 C/T*) were examined. To graphically visualize the stable gene compositions of the cytokine network, we performed a computer simulation of the network interactions of various cytokine genotypes involved in the formation of susceptibility to POAG using the Cytoscape bioinformatics platform.

*Results:* Analysis of the topology of the associative gene network of cytokines allowed us to identify the main genes and main intergenic interactions that make the greatest contribution to the development of primary open-angle glaucoma. In our opinion, six polymorphisms (*MMP2-1306CT*, *IL17-197GG*, *IL1B-31TT*, *IL8-251TA*, *MMP9-1562CC* and *IL6-174GC*) can act as the main genes, which account for 80 % of all involved genes. These genes form the main nodes ("hubs") in the gene network, since they have the largest number of interactions with other genes. These polymorphisms form the main intergenic interactions, which account for 50 % of all interactions in the gene network.

*Conclusion:* The construction of gene networks of transcriptional regulation and their topological analysis makes it possible to build and study the structural and functional organization of gene-gene interactions in relation to the study of the pathogenesis of primary open-angle glaucoma for the subsequent development of approaches to personalized prevention and therapy.



*Acknowledgements:* The study was carried out as part of the government assignment (budgetary funding within the framework of the comprehensive research topic No. 0324-2019-0046-C-02 “Development of molecular profiling methods and innovative technologies for prediction, early diagnosis, drug and cell therapy of socially significant human diseases of autoimmune, inflammatory and dysmetabolic nature”).

### *References*

1. Tham Y.C., Li X., Wong T.Y., Quigley H.A., Aung T., Cheng C.Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090. doi: 10.1016/j.ophtha.2014.05.013.
2. Zhang Y.H., Xing Y.Q., Chen Z., Ma X.C., Lu Q. Association between interleukin-10 genetic polymorphisms and risk of primary open angle glaucoma in a Chinese Han population: a case-control study. *Intl J Ophthalmol*. 2019;12(10):1605-1611. doi: 10.18240/ijo.2019.10.13.
3. Shevchenko A.V., Prokof'ev V.F., Konenkov V.I., Chernykh V.V., Ermakova O.V., Trunov A.N. Analysis of IL1B (rs1143627), IL4 (rs2243250), IL6 (rs1800795), IL8 (rs4073), IL10 (rs1800896, rs1800872), IL17A (rs227593) cytokines genes polymorphism and its complex genotypes among Caucasian patients of Western Siberia with primary open-angle glaucoma. *Immunologiya*. 2021;42(3):211-221. doi: 10.33029/0206-4952-2021-42-3-211-221 (in Russian).

## The role of insulin receptor signaling in regulation of synaptic glutamatergic hippocampal neurotransmission

Proskura A.L.\*, Vechkapova S.O., Ratushnyak A.S.

*Federal Research Center of Information and Computational Technologies, Novosibirsk, Russia*

\* [annleop@mail.ru](mailto:annleop@mail.ru)

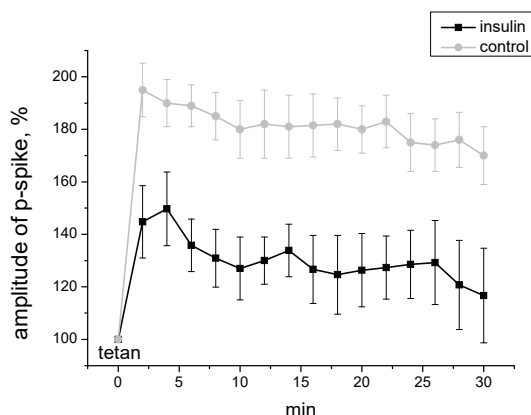
**Key words:** hippocampus, dendritic spine, synaptic plasticity, glutamate receptors, insulin

*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  $\alpha$ -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. It mediated of intermolecular rearrangements and form the functional regulatory networks [1]. In addition, the receptors of various hormones, in particular insulin, are present in the synaptic contacts zone.

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 [2]. Insulin has also been shown to regulate the endocytosis of synaptic AMPA receptors [3]. It is established that the activity of insulin receptors provides the maintenance of spine AMPAR density that probably mediates as basal neurotransmission as the long-term changens in the CA1 region of the hippocampus, at the same time the molecular mechanisms of these processes are described very superficially.

*Results:* The interactom of hippocampal dendritic spines was reconstructed (technology GeneNet was used (ROSPATENT No. 990006 от 15/02/1999)). The reception of the basic mediator signal – glutamat – is carried out by NMDAR macrocomplexes. A key event is the groundwork of second messengers pools.  $P4 \rightarrow PIP2$  (PtdIns(4,5)P2) catalyzed Phosphoinositide 5-kinase (PI5K),  $PIP2 \rightarrow PIP3$  (PtdIns(3,4,5)P3) mediated Phosphoinositide 3-kinase (PI3K). The AMPA subtype of glutamate receptors is subject to functionally distinct constitutive and regulated clathrin-dependent endocytosis, contributing to various forms of synaptic plasticity. It is established that insulin stimulated endocytosis of GluR2-AMPA receptors. This process has been accompanied disrupt interaction this subtype of receptors with synaptic scaffold proteins [3, 4].

To assess of insulin contribution to the processes of long-term potentiation induction, a series of experiments was carried out on hippocampal slices obtained from 2-month-old male ICR mice. Standard electrophysiological technique was used. Stimulation of Schaffer collaterals and recording of induced population spikes (p-spike) of pyramidal neurons in the CA1 field was performed via extracellular electrodes filled with physiological saline. A single stimulus inducing p-spike with amplitude  $\leq 50$  % of maximal amplitude was used for tetanization of the Schaffer collaterals. A stimulus of same intensity was used for analysis of responses after potentiation. Tetanization was induced by a single series of electric stimulation with a frequency of 100 Hz for 1 sec.



**Fig. 1.** Effects of insulin on long-term potentiation

Before tetanization, the slices were perfused with standard solution containing insulin (10 nM) for 30 min. The control slices were perfused the same time with the standard solution. 2 minutes after tetanization of the Schaffer collaterals, the amplitude of p-spike in the control group reached  $195.5 \pm 10.2$  % of the basal level, while in the group incubated with insulin, it was significantly lower than in the control group,  $144.8 \pm 13.8$  % of the basal level ( $p \leq 0.01$ ). By the 30th minute after tetanization, the amplitude of p-spike in the control group reached  $183 \pm 9.5$  %, and in the group incubated with insulin it was only  $116 \pm 15.1$  % ( $p \leq 0.001$ ). Insulin effects mediated of PIP3 accumulation, that provides the prolonged of PKC activity. PKC mediate GluR2-AMPA 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 [5].

**Conclusion:** Incubation of slices in solution with the addition of insulin leads to impaired development and maintenance of long-term potentiation. The insulin effects did not with NMDA receptor and implemented in the presence of ROS.

**Acknowledgements:** The work utilized the materials obtained in a basic research project of the Russian Academy of Sciences (project No. VI.35.2.6).

### References

1. Proskura A.L. et al. Synapse as a multi-component and multi-level information system. *Stud Comput Intell.* 2018;736:186-192.
2. Grillo C.A. et al. Hippocampal insulin resistance impairs spatial learning and synaptic plasticity. *Diabetes.* 2015;64(11):3927-3936.
3. 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.
4. Ahmadian G. et al. Tyrosine phosphorylation of GluR2 is required for insulin-stimulated AMPA receptor endocytosis and LTD. *EMBO J.* 2004;23(5):1040-1050.
5. Seo J.H. et al. 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.

## Effect of silver-containing sorbent on erythrocytes of erythrocyte mass during perfusion *in vitro*

Rachkovskaya L.N.<sup>1\*</sup>, Smagin A.A.<sup>1</sup>, Nimaev V.V.<sup>1</sup>, Rachkovsky E.E.<sup>1</sup>,  
Yastrebova E.S.<sup>2</sup>, Maltsev V.P.<sup>2</sup>, Letyagin A.Yu.<sup>1</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Voevodsky Institute of Chemical Kinetics and Combustion, SB RAS, Novosibirsk, Russia*

\* [noolit@niikel.ru](mailto:noolit@niikel.ru)

The effectiveness of the method of hemosorption detoxification of the body in various pathological conditions can be enhanced and corrected by using sorbents modified with various biologically active ingredients. The aim of the study was to study the effect of a silver-containing sorbent on the parameters of erythrocytes of erythrocyte mass under *in vitro* perfusion conditions. Sorbents based on porous aluminum oxide and polydimethylsiloxane with and without nanocluster silver were studied. The porous structure of sorbents and their adsorption activity were evaluated by standard methods. Perfusion of erythrocyte mass through columns filled with sorbents was carried out according to the standard procedure. Morphofunctional parameters of erythrocytes before and after perfusion were studied by scanning flow cytometry. The statistical significance of the differences between the groups was assessed using the Mann-Whitney test and the Student t-test at  $p < 0.05$  (Statistica 7 software package). Analysis of morphofunctional parameters of erythrocytes before and after perfusion showed that all indicators were within normal limits. However, the indicators of the ultimate extensibility of the erythrocyte membrane when using a silver-containing sorbent exceed 1.5 and 2 times the value compared to the indicators for a sorbent without silver and the initial erythrocyte mass respectively. Thus, the sorbent modified with silver does not have a traumatic effect on the morphofunctional parameters of erythrocytes during perfusion. The mechanisms of action affecting the indicators of the ultimate extensibility of the erythrocyte membrane after perfusion with the introduction of nanocluster silver into the sorbent require further research.

*Acknowledgements:* This study was supported by the budget project of the Research Institute of Clinical and Experimental Lymphology – Branch of the ICG SB RAS (FWNR-2022-0009).

# Uncovering the genes linking glucose variability with endothelial dysfunction in diabetes by the analysis of gene networks

Saik O.\* , Klimontov V.

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

\* saik@bionet.nsc.ru

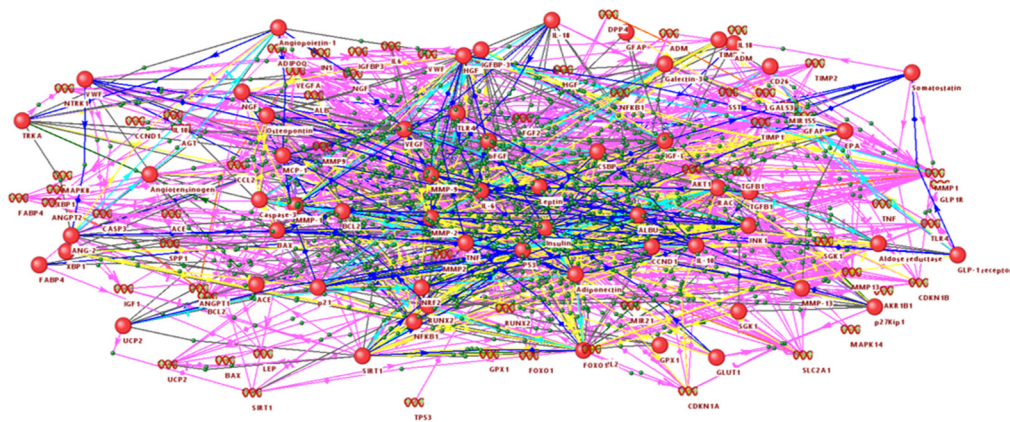
**Key words:** glucose variability, diabetes, endothelial dysfunction, gene networks, ANDSystem

**Background:** High glucose variability (GV) is recognized as a contributor for diabetic microvascular and macrovascular complications, including diabetic retinopathy, chronic kidney disease, diabetic neuropathy, and macrovascular disease [1–3]. The results of numerous studies have pointed to the role of endothelial dysfunction in the development of these complications [1, 4].

**The Aim:** to identify genes linking GV with endothelial dysfunction in diabetes.

**Methods:** The gene networks reconstruction and analysis were performed by the ANDSystem ([www-bionet.sccc.ru/and/cell/](http://www-bionet.sccc.ru/and/cell/)) [5, 6]. The statistical significance of the enrichment of gene networks by common genes was assessed according to hypergeometric distribution calculated by the “hypergeom.sf” function of Python “scipy.stats” module [7]. The analysis of biological processes overrepresentation for the list of common genes was performed by the web-service DAVID ([david.ncifcrf.gov/](http://david.ncifcrf.gov/)) [8]. The values of betweenness centrality were calculated by the “Analysis” function of ANDSystem. The Mann-Whitney U test was performed by the “mannwhitneyu” function of Python “scipy.stats” module [7].

**Results:** Based on the information from ANDSystem the gene networks of the GV and the endothelial dysfunction were reconstructed. These gene networks contained 151 and 682 genes/proteins respectively. Sixty genes were common for both analyzed networks (Fig. 1).



**Fig. 1.** Cluster of the gene network containing the 60 genes/proteins that are common for both the gene networks of the GV and the endothelial dysfunction

The probability of observing so many common genes for random reasons is very small ( $p$ -value  $< 10^{-47}$ ). The identified 60 common genes/proteins are involved in the activation, apoptosis, development, differentiation, migration, morphogenesis and proliferation of endothelial cells according to the results of the Gene Ontology enrichment analysis. It was found that in the GV gene network many genes associated with endothelial dysfunction have relatively high values of betweenness centrality. The average of betweenness centrality values for the common 60 genes involved in endothelial dysfunction and GV was 536, while for all GV genes it was only 270. The difference of the betweenness centrality values for these two gene sets is statistically significant

( $P$ -value =  $10^{-5}$ , Mann-Whitney U test). All the top ten genes with highest betweenness centrality values in the GV gene network (*INS*, *IL6*, *TP53*, *FOXO1*, *LEP*, *MAPK14*, *TNF*, *IL1B*, *IGF1*, and *VEGFA*) were also involved in the endothelial dysfunction.

**Conclusions:** Discovering the key genes involved both in the GV and endothelial dysfunction could shed light on understanding of the pathogenesis of diabetes complications and could support the future drug development.

**Acknowledgements:** This work was supported by the grant of Russian Science Foundation (20-15-00057).

### References

1. Klimontov V.V. et al. Glucose variability: How does it work? *Int. J. Mol. Sci.* 2021;22(15):7783.
2. 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.
3. 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.
4. Domingueti C.P. et al. Diabetes mellitus: the linkage between oxidative stress, inflammation, hypercoagulability and vascular complications. *J. Diabetes Complicat.* 2016;30(4):738-745.
5. Ivanisenko V.A. et al. ANDSytem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol.* 2015;9(Suppl 2):S2.
6. Ivanisenko V.A. et al. A new version of the ANDSytem 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):5-15.
7. Shukla X.U., Parmar D.J. Python—A comprehensive yet free programming language for statisticians. *J. Stat. Manag. Syst.* 2016;19(2):277-84.
8. Huang D.W. et al. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols.* 2009;4(1):44-57.

## 3D-2 heterogeneous breast cancer models

Savinkova M.M.<sup>1,2\*</sup>, Ermakov M.S.<sup>1</sup>, Novak D.D.<sup>1,2</sup>, Varlamov M.E.<sup>1,2</sup>,  
Richter V.A.<sup>1</sup>, Koval O.A.<sup>1</sup>, Nushtaeva A.A.<sup>1</sup>

<sup>1</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* m.savinkova@g.nsu.ru

**Key words:** breast cancer, 3D-models, cells technology, invasion, tumour cells, stromal cells

*Motivation and Aim:* Cell models of breast cancer provide an opportunity for the investigation of molecular mechanisms of tumour progression. Tumour spheroids (3D-models) are frequently used *in vitro* as a model in preclinical studies as we can observe the gradient of the cells receiving oxygen and nutrients. However, most modern 3D-models contain only one type of cells, while in order to be physiologically relevant the models must be heterogeneous. 3D-2 cellular model consists of tumour and stromal cells. Growing heterogeneous spheroids is of current importance today because such models are necessary for the development of new approaches to antitumour therapy and the study of the intercellular interaction in cancer progression.

The aims of the research are to construct heterogeneous 3D-2 cellular models of breast cancer and to analyse the proliferative and invasive ability of cells stimulated by 17 $\beta$ -estradiol and transforming growth factor beta (TGF- $\beta$ ).

*Methods and Algorithms:* MCF-7 (ER<sup>+</sup>/PR<sup>+</sup>/HER2<sup>-</sup>), MDA-MB-231 (ER<sup>-</sup>/PR<sup>-</sup>/HER2<sup>-</sup>) and SK-BR-3 (ER<sup>-</sup>/PR<sup>-</sup>/HER2<sup>+</sup>) were used as cancer components of 3D-2 spheroids for different imitating tumour types. We chose patient-derived cancer fibroblasts (BrC4f and BrC120f) and a normal fibroblast (BN120f) as stromal components of 3D-2 models. To visualise the relative position of cells in the spheroids, we obtained tumour cells (MCF-7, MDA-MB-231 and SK-BR-3) expressing red fluorescent protein (mKate) and fibroblasts (BrC4f, BrC120f and BN120f) expressing green fluorescent protein (eGFP). We used low-adhesive plastic and co-cultivated tumour cells and different types of fibroblasts to construct 3D-2 spheroids. Invasion potential of cells which form 3D-2 spheroids was tested on Matrigel<sup>TM</sup> substrate.

*Results:* When tumour and stromal cells were mixed, fibroblasts formed the inner frame, and tumour cells formed the outer layer of 3D-2 models. The largest spheroids were obtained from cell culture SK-BR-3. Forming a dense necrotic core and a clear boundary of the spheroid were observed. The smallest spheroids were obtained from MDA-MB-231 cell culture. 3D-models from MCF-7 cells were characterised by a smaller size of necrotic core and the absence of a clear border of the spheroids in comparison with spheroids obtained from SK-BR-3. When we co-cultivated tumour cells with fibroblasts, we found out that the spheres became denser and they got a more precise structure in comparison with the spheres obtained only from tumour cells.

17 $\beta$ -estradiol stimulated the proliferation of MDA-MB-231 cells during mono-culturing, which may be due to the expression of the ER $\alpha$ 36 splice variant in cells associated with an aggressive cell phenotype. 17 $\beta$ -estradiol influenced the self-organization of MCF-7 culture spheroids during mono- and co-cultivation with fibroblasts. This was reflected in an increase in their density, which was accompanied by self-organisation into

microtissue. TGF- $\beta$  stimulated cell proliferation of hormone-independent culture SK-BR-3 in 3D-2 spheroids, which may be connected to the implementation of TGF $\beta$ /HER2/EGFR signalling in them.

The analysis of the invasive potential of tumour cells revealed that co-cultivation with stromal cells stimulated cell invasion only for the MDA-MB-231 culture. The stimulation with TGF- $\beta$  reduced the invasion of cells from 3D-2 models, while the stimulation with 17 $\beta$ -estradiol increased the invasive potential.

*Conclusion:* As a result, the conditions for construction of 3D-2 heterogeneous spheroids of breast cancer capable of imitating tumour tissue were optimized. 3D-2 heterogeneous breast cancer model is an appropriate *in vitro* model for studying tumour-stroma interactions, invasion potential of tumour cells and research of cytotoxic effect of antitumour drugs.

*Acknowledgements:* The research was supported by the Russian Science Foundation Grant No. 20-74-10039.



## Milk exosomes – perspective agents for drug delivery

Sedykh S.<sup>1,2\*</sup>, Timofeeva A.<sup>1</sup>, Cherenko V.<sup>1,2</sup>, Paramonik A.<sup>1,2</sup>, Dome A.<sup>1</sup>, Antropov D.<sup>1</sup>, Stepanov G.<sup>1,2</sup>, Nevinsky G.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *sedyh@niboch.nsc.ru*

**Key words:** exosomes, milk, drug delivery, mRNA-vaccine, RBD, SARS-CoV-2

*Motivation and Aim:* Exosomes are natural vesicles with a 40–100 nm diameter. Exosomes have a high diagnostic potential and are universal and effective agents for the target delivery of drugs into cells. The advantages of exosomes over other delivery methods are their low cytotoxicity, size, low immunogenicity, and expression of specific surface markers. It has been shown that bovine and human milk-derived exosomes may successfully penetrate cells. Various mechanisms have been proposed to deliver biologically active substances using exosomes, both inside these vesicles and on their surface, for example, by functionalizing the surface of exosomes with peptides, transfection with commercial reagents, ultrasound, electroporation, liposomes and others.

*Methods and Algorithms:* In 2017, we were the first group to describe exosomes of horse milk. We analyzed the protein and nucleic acids content and assessed the prospects for the use of milk exosomes as anticancer drug delivery agents. Here, the exosome preparations were isolated from the horse, human, and cow milk by ultrafiltration and ultracentrifugation and further purified by gel filtration. Milk exosome preparations were loaded with chemically synthesized short RNAs and RBD-mRNA corresponding to the SARS-CoV-2 S-protein. The cytotoxicity of intact and loaded vesicles was determined. Laboratory mice of the CD1 line were immunized, and several cell cultures were transfected.

*Results:* Using RT-qPCR and ELISA, we have shown the production of antibodies against RBD and changes of corresponding genes in cell cultures.

*Conclusion:* The results will allow developing the new methods for delivering mRNA vaccines, therapeutic nucleic acids, and anticancer drugs.

*Acknowledgements:* The study was supported by a project of RSF 18-74-10055 to Sergey Sedykh.

## Metabolic markers of hypertension in ISIAH rats

Seryapina A.A.<sup>1\*</sup>, Redina O.E.<sup>1</sup>, Markel A.L.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *seryapina@bionet.nsc.ru*

Arterial hypertension, being a complex multifactorial disease, affects various organs and tissues, eventually altering their metabolism. Search for metabolic markers of the disease is considered to be essential for working out ways of diagnostics, treatment and prevention. The aim of our research was to identify metabolic markers for hypertensive disease. In our study, NMR technique was applied to detect concentrations of 50 low-molecular metabolites in the serum sampled from the 6-month-old ISIAH rats with inherited stress-dependent arterial hypertension. Control group was formed of normotensive WAG rats of the same age. Statistical analysis of the data (two-way ANOVA with Tukey post-hoc test) showed that ISIAH rats had significantly lower serum concentrations of citrate, creatine, cytidine, dimethylglycine, formic acid, glucose, glycine, mannose, methionine, terephthalic acid and threonine, as compared to WAG rats. These data may indicate the alterations in TCA cycle and  $\beta$ -oxidation of fatty acids, which are due to arterial hypertension pathogenesis. Decreased levels of glycine in hypertensive rats may alter many metabolic pathways, although, according to previous studies, glycine concentration may be restored through oral consumption, which results in lowering blood pressure and reversing some of the adverse effects of the disease, if administered at the early stages. In addition, multivariate statistical analysis was performed on the data obtained. Principal component analysis demonstrated that differences between ISIAH and WAG rats may be due to alterations in  $\beta$ -oxidation of fatty acids and regulation of arterial wall elasticity involving NO-synthase and nitric oxide. Therefore, metabolic profiling does provide important information for understanding molecular mechanisms underlying the pathogenesis of arterial hypertension in ISIAH rats.

*Acknowledgements:* This study was supported by RFBR (project No. 20-04-00119) and budget project No. FWNR-2022-0019.

## Dkk-1 level detection in ankylosing spondylitis by the aptamer-based colorimetric assay

Shatunova E.<sup>1\*</sup>, Kolpakov K.<sup>2</sup>, Kurochkina Y.<sup>2</sup>, Korolev M.<sup>2</sup>, Vorobyeva M.<sup>1</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [lizashatunova@yandex.ru](mailto:lizashatunova@yandex.ru)

**Key words:** ankylosing spondylitis, Bekhterev's disease, Dickkopf-1 like protein, enzyme-linked immunosorbent assay, DNA-aptamers

*Motivation and Aim:* The pathological process in ankylosing spondylitis (AS) includes bone tissue remodeling with possible bone loss and the development of secondary osteoporosis. A set of proteins regulate pathological bone metabolism in AS. Among them, the soluble serum protein Dkk-1, an antagonist of the Wnt signaling pathway, is of particular interest and importance. The lack of a generally accepted standard method for Dkk-1 detection is an obstacle in research of its role in AS structural progression. Oligonucleotide aptamers specific to certain biomarkers are very prospective for developing novel diagnostic assays due to their high affinity and the possibility of reproducible, scalable and cost-effective chemical synthesis. Stable properties of aptamers provide high reproducibility of the assays, which is especially important for longitudinal studies. In this work, we aimed to compare methods for measuring Dkk-1 in blood serum using conventional ELISA and the aptamer/antibody assay and to investigate the relationship between Dkk-1 levels and AS structural progression and secondary osteoporosis.

*Methods:* The study included 17 patients with AS (5 at the advanced stage and 12 at the late stage) and 8 healthy volunteers. Serum samples were diluted 10-fold prior to analyses. Colorimetric microplate ELISA of Dkk-1 in blood serum was performed using a commercial ELISA Kit. For the aptamer/antibody assay, the TD10 DNA aptamer [1] to human Dkk-1 was chemically synthesized, purified, and covalently immobilized in the wells of a transparent 96-well DNA-BIND plate. In all cases, anti-Dkk-1 antibody with horseradish peroxidase and chromogenic substrate were used for specific visualization of the analyte. The data from colorimetric analyses were preliminarily checked for normality and then the nonparametric Mann–Whitney U-test was used. The difference was considered statistically significant at  $p < 0.05$ .

*Results:* The results of colorimetric detection of the Dkk-1 level by ELISA aptamer-based assays showed good coincidence. The data indicate a significant increase in the level of serum Dkk-1 in patients with AS compared with healthy donors. Notably, the analysis by the aptamer/antibody assay fairly distinguished the samples from healthy volunteers and from AS patients without structural radiographic progression. In patients with syndesmophytes, the levels of serum Dkk-1 exceeded those for patients without syndesmophytes. At the same time, measuring Dkk-1 levels by both methods did not reveal significant differences between the groups of patients with or without osteoporosis.

*Conclusion:* Serum levels of Dkk-1 in AS patients differed significantly depending on the disease stage, but did not depend on the presence or absence of osteoporosis. Dkk-1 represents a potential serum marker of AS progression, capable of reflecting a trend towards the structural progression of the disease before the appearance of characteristic radiographic changes. The DNA aptamer-based test system allows for measuring the level of Dkk-1 in the blood serum of AS patients. In terms of diagnostic significance, this assay is highly competitive with conventional ELISA and exceeds the latter by a number of benefits.

*Acknowledgements:* This study was supported by the Russian state-funded projects for ICBFM SB RAS (grant No. 121031300042-1) and to RICEL – Branch of ICG SB RAS (grant No. 0324-2019-0046-C-02).

### *References*

1. Zhou Y. et al. Developing slow-off dickkopf-1 aptamers for early-diagnosis of hepatocellular carcinoma. *Talanta*. 2019;194:422-429.

# Computational pipeline and database for the epidemiological typing and antimicrobial resistance profiling of important bacterial pathogens

Shelenkov A., Mikhaylova Y., Akimkin V.

*Central Research Institute of Epidemiology of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Moscow, Russia*

\*fallandar@gmail.com

**Key words:** epidemiological surveillance, isolate typing database, whole genome sequencing, antimicrobial resistance, genomic markers, ESKAPE pathogens

*Motivation and Aim:* Monitoring the spread of particular clones of pathogenic bacteria and associated antimicrobial resistance determinants within a particular healthcare facility, country or across the world represents an important task for national and international public health institutions. During recent years, the antibiotic resistance within different species of nosocomial and community-acquired bacterial pathogens has increased to dangerous levels [1]. This problem cannot be solved without comprehensive investigations of resistance transmission mechanisms and global epidemiological surveillance. Currently, such a surveillance is often dependent on whole genome sequencing (NGS) of bacterial genomes and corresponding bioinformatics analysis pipelines [2]. Although the number of available genomic sequences for bacterial isolates increases very rapidly (more than 100,000 bacterial genomes are already available in NCBI Genbank, <https://www.ncbi.nlm.nih.gov/genome>), public genomic databases usually lack the essential epidemiological information for the isolates, for example, the results of multilocus sequence typing (MLST) or typing using other schemes, as well as the antimicrobial gene content. Thus, it is very difficult to use the information available in such databases for epidemiological surveillance of resistance spreading. The pipeline developed and the corresponding database under development aim to assist the users to retrieve such information from the genomes of important pathogenic bacteria.

*Methods and Algorithms:* Earlier we have developed a computational pipeline for seamless integration of various software tools for annotation of newly sequenced clinical bacterial isolates, including typing using various schemes [3]. In this work, we have added additional typing tools (cgMLST and hybrid profiles, SPA types, serovars etc) and performed an analysis of all genomes available in Genbank for important nosocomial pathogens belonging to ESKAPE group (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp). We used Resfinder 4.0 database for antimicrobial gene detection (<https://cge.cbs.dtu.dk/services/ResFinder/>) and MentaLiST (<https://github.com/WGS-TB/MentaLiST>, version 0.2.4) [4] to obtain the cgMLST profiles. The database of MLST-based sequence types (ST) for the isolates available in Genbank is currently available in local network of our institution and contains the information regarding the STs obtained and the assignment of them to known international or global clones of high risk.

**Results:** We have applied the computational tools developed to investigate the epidemiological characteristics of important nosocomial pathogens. Here we present the results for one of them, *Acinetobacter baumannii*, for which more than 12,000 genomes are available in Genbank. Currently, there are 10 known international clones of high risk for this important pathogen, and they are no longer specific to particular regions of the world and may co-exist even within a single hospital [5]. Most part of the isolates (about 70 %) was found to belong to well-established international clone (IC) 2, while IC1 genomes constituted only about 10 %. These data represents the skew of IC distribution towards well-known clones, while the epidemiological data report different clonal variability [1]. However, the researchers are usually interested in making comparison between the isolates obtained in particular location and the ones available in public databases to infer the possible ways of clone spreading and resistance transmission via, e.g., phylogenetic investigations. For this purpose, the data for comparatively rare occurring STs could be useful, and our database provides such data. At the same time, even the studies of ubiquitous clones like IC1 and IC2 can provide important epidemiological information since the antibiotic gene content cannot be always associated with particular clones, and resistance gene distribution could vary within particular clonal group.

**Conclusion:** We have developed a user-friendly computational pipeline for typing and annotation of pathogenic bacteria genomes, including antibiotic resistance prediction, and applied this pipeline and other tools available to collect the epidemiologically important data from public genome databases. The obtained data will be available online soon and will help in performing various genomic epidemiology investigations and prepare scientifically valid manuscripts.

**Acknowledgements:** The study was supported by the industry research program of Rospotrebnadzor (2021–2025) “Scientific support of epidemiological monitoring and public health measures of the Russian Federation territory” and Russian Federation Government Resolution No. 3116-r.

## References

1. Kuzmenkov A. et al. AMRmap: An interactive web platform for analysis of antimicrobial resistance surveillance data in Russia. *Front Microbiol.* 2021;12:620002.
2. Maljkovic B. et al. Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J Infect Dis.* 2020;221:S292-S307.
3. Shelenkov A. et al. Characterization of antimicrobial resistance, virulence profiling and analysis of whole-genome sequence of clinical *Klebsiella pneumoniae* isolates. *Antibiotics (Basel).* 2020;9(5):261.
4. Feijao P. et al. MentaLiST – A fast MLST caller for large MLST schemes. *Microbial Genomics.* 2018;4(2):146.
5. Shelenkov A. et al. Diversity of international high-risk clones of *Acinetobacter baumannii* revealed in a Russian multidisciplinary medical center during 2017–2019. *Antibiotics (Basel).* 2021;10(8):1009.

## Ag modification of PLC-COOH nanofibers provides mesenchymal cells proliferation

Sitnikova N.A.\*, Solovieva A.O., Manakhov A.M., Permyakova E.S.

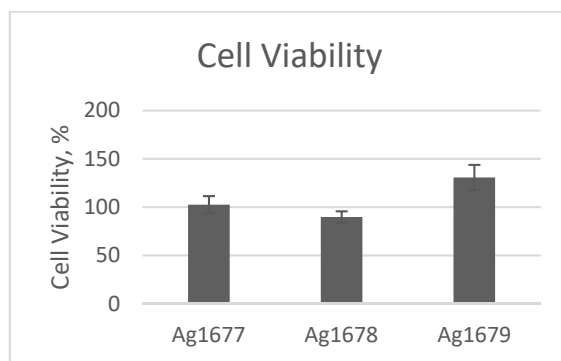
*Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *Sitnikovanat9@gmail.com*

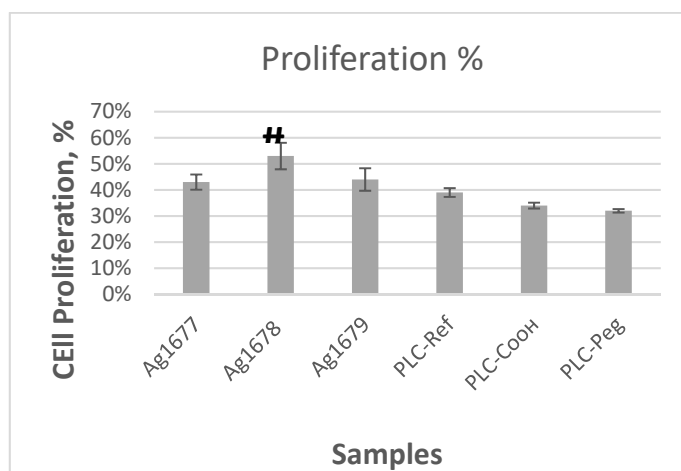
**Key words:** Nanofibers, Polycaprolacton, Ag, Cell proliferation

*Motivation and Aim:* Currently, wound healing materials with protective and regenerative properties are in great demand, the development of which is actively being carried out all over the world. To date, active research is underway on regenerative materials with an antibacterial effect. Silver is actively studied as an antibacterial component and is used in various creams, substances and nanomaterials for regenerative medicine [1, 2]. In addition to the antibacterial effect, silver has a regenerative effect, provided by stimulating the proliferation of cells involved in regeneration. Antimicrobial properties of silver have been known for a long time, but there is also cytotoxicity of high concentrations of silver [1]. Therefore, it is important to select the concentration and shape of silver depending on the goals. The ideal wound dressing should ensure that the wound remains moist, protected from infections, and has no toxic compounds. In present work, we obtained a series of polycaprolactone-based nanomaterials fabricated by electrospinning, modified with COOH groups and incorporated with silver ions (up to 0.6 at. %). By adjusting the magnetron current (0.3 A) and implanter voltage (5 kV), the deposition of TiO<sub>2</sub> and Ag<sup>+</sup> implantation into PCL/PEO nanofibers was optimized in order to achieve implantation of Ag<sup>+</sup> without damaging the nanofibrous structure of the PCL/PEO [3]. It was shown that the maximum amount of silver is released from the PCL-Ti0.3-Ag-5 kV sample, and in the first hours there is a rapid release, then a smoother one, and for 7 days it continues to be released slower [3]. The aim of this work is to determine the effect of the obtained samples on the viability and proliferative potential of mesenchymal stromal cells.

*Methods and Algorithms:* PCL nanofibers were created by the method described in the article [3]. Here we used PLC nanofibers with the following characteristic: A1678 (0.3 A \* 2 min + Ag 8 kV\*3 mA\*12 min) 180 mm, A1679 (0.3 A\*2 min + Ag 5 kV\*3 mA\*15 min) 180 mm, A1677 (0.5 A\*3 min + Ag 15 kV\*3 mA\*10 min) 120 mm. Membranes were cut in pieces with 0,5 diameter and used in set of experiments. MTT test was conducted to test cytotoxicity of used Ag concentrations. 5\*10<sup>3</sup> hMSC were seeded in 96-wells plate on plastic, Ag-PCL membranes samples were soaked in culture medium for 24 hours, then this medium was transferred to cells for 3 days incubation. Results were measured by optical density by used Multiscan FC spectrophotometer. To test proliferation rate human mesenchymal stromal cells were seeded on Ag-PCL membranes by 5\*10<sup>3</sup> cells per sample and incubated for 3 days in standard conditions (95 % humidity, 5 % CO<sub>2</sub>). To further evaluate total amount, cells were dyed with Hoechst reagent to visualized cells nuclei and with EdU Click iT-reagent to count number of proliferating cells. *Results:* Ag<sup>+</sup> ion release differed in samples depending on the initial Ag concentrations and was presented in the article [3]. Ag concentrations in all samples had no toxicity on human mesenchymal cells, even more Ag<sup>+</sup> ions released in culture medium from Ag1677 and Ag1679 resulted in increased the percentage of living cells in contrast with the control (Fig 1). Total cell count was significantly higher on Ag1678 samples in comparison to PLC-ref membranes. However, proliferation rate was higher on all samples in comparison to PLC-ref (Fig. 2).



**Fig. 1.** Cell viability after incubation in culture medium from Ag-PLC nanofibrous samples, soaked for 24 hours in relation to 100 % viability in control wells. Results presented as mean  $\pm$  std



**Fig. 2.** Proliferation rate of hMSC on Ag incorporated PLC nanofibers in comparison with PLC-ref, PLC-COOH and PLC-Peg modifications. Results presented as means  $\pm$  std. # - statistical significance ( $p < 0,05$ ). Ag1677 – (0.5A\*3 min + Ag 15kV\*3mA\*10 min) 120 mm, Ag1678 – (0.3A \* 2 min + Ag 8 kV\*3 mA\*12 min) 180 mm, Ag1679 – (0.3A\*2 min + Ag 5 kV\*3 mA\*15 min) 180 mm

**Conclusion:** Ag modification of PLC nanofibers membrane with the following parameters: A1678 (0.3A \* 2min + Ag 8kV\*3mA\*12min) induces proliferation of human mesenchymal cells, though other concentration has a tendency to promote proliferation too, that we can see by the results of MTT test. Besides that, fast Ag<sup>+</sup> ion release from PLC nanofibers (Ag1677) in the first 24 hours could've possibly reduce bacterial activity in wound dressings, but is low enough for a longer period of time and did not cause cytotoxicity. Obtained results allows us to speculate about that material application in biomedical field and in tissue engineering as it can provide effective protection not only from mechanical influence but also has antimicrobial effect because of Ag, which is important for chronic wounds and injuries with large area of damage, and can activate host cells proliferation.

**Acknowledgements:** The work has been conducted with the support of RSF No. 18-75-10057-II.

### References

1. Madhavan R.V. et al. *Tissue Engineering. Part A.* 2011;17(3-4):439-449.
2. Thomas R. et al. *Appl. Biochem. Biotechnol.* 2015;176(8):2213-2224.
3. Manakhov A.M. et al. *Molecules.* 2022;27(4):1333.



## The nearest and one-month effects of endovascular revascularization for patients with critical limb ischemia

Sobolevskaya E.V.<sup>1,2\*</sup>, Shumkov O.A.<sup>1</sup>, Smagin M.A.<sup>1</sup>, Hapaev R.S.<sup>1</sup>, Nimaev V.V.<sup>1</sup>

<sup>1</sup> Laboratory of Operative Lymphology and Lymphatic Detoxification of the Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> V. Zelman Institute for Medicine and Psychology – Novosibirsk State University, Novosibirsk, Russia

\* darren98@mail.ru

**Key words:** endovascular therapy, critical limb ischemia (CLI), peripheral arterial disease (PAD), transcutaneous oximetry

*Introduction:* The method of endovascular revascularization has a number of key advantages in the treatment of critical lower limb ischemia in comparison with other methods alongside with a rapid advance in angioplasty procedure every year. However, the drawbacks remain that affect the pace and quality of achieving technical and clinical success in percutaneous transluminal angioplasty. For instance, the effectiveness of the treatment of critical limb ischemia is influenced by the segment level in which the arterial revascularization was performed, as well as by the number of "open" arteries following the surgery. In particular, such treatment efficiency assessment is highly relevant for patients possessing lesions of two or more arteries, as well as suffering of purulent-necrotic lesions. It should be noted that the relevant results obtained in this study will be used to guide the development of personalized revascularization tactics.

*Materials and Methods:* The transcutaneous partial pressure of oxygen (TcPO<sub>2</sub>) was performed in 19 patients who were diagnosed with lesions of the vascular atherosclerosis by ultrasonic Dopplerography. In the structure of the severity of peripheral arterial disease (PAD), patients with necrotic and ulcerative foot defects predominated. In 14 patients, the course of obliterating atherosclerosis of lower limb arteries was aggravated by diabetes, in 1 patient was detected hyperglycemia. The age of the patients ranged from 54 to 82 years, averaging 65.95 ± 7.91 years. Mediana 66 [65;67]. The clinical characteristics of the patients are presented in Table 1.

**Table 1.** Clinical characteristics of patients (N = 19)

Characteristics	n	%
Male	14	73,68
Female	5	26,32
CLI manifestations:		
- foot ulcer	11	57,89
- Pain in rest	8	42,10
Comorbidities:		
- Coronary artery disease	5	26,31
- Arterial hypertension	19	100
- Renal insufficiency	9	47,37
- C2 (GFR < 60 mL/min/1.73 m <sup>2</sup> )	4	21,05
- C3a (GFR < 45 mL/min/1.73 m <sup>2</sup> )	5	26,32
Type 1 diabetes	2	10,53
Type 2 diabetes	11	57,89
Hyperglycemia	1	5,26
Including Diabetes Therapy		
- Only insulin	4	21,05
- Only tablet preparations	2	10,53
- Insulin + tablet preparations	9	47,37

According to examination results (ultrasound and peripheral angiography), all patients revealed occlusion or significant stenosis of the superficial femoral artery (SFA), popliteal

artery (PA), posterior and anterior tibial arteries (PTA and ATA). More than 47 % ( $n = 9$ ) of cases had multilevel lesions of the femoral-popliteal and popliteal-tibial segments. All 19 patients were diagnosed with both isolated hemodynamically significant arterial stenoses – 11 (57.8 %), and multiple – 8 (42.10 %). Level of isolated lesions: SFA – 3 (15.79 %); PA – 1 (5.26 %); PTA and ATA – 5 (26.3 %). Among the multiple lesions, there was a combination of stenosis at various levels of the great vessels of the lower extremities. TcPO<sub>2</sub> was conducted before X-ray endovascular revascularization, after surgery on day 1, and after 1 month also.

*Results:* Percutaneous transluminal angioplasty of EIA and SFA was performed in 3 patients (15.79 %), EIA only – 2 (10.53 %), EIA and CFA – 2 (10.53 %), SFA – 3 (15.79 %), SFA + PA – 1 (5.26 %), POP – 2 (10.53 %), only ATA – 5 (26.32 %), ATA + TPT + PTA – 1 (5.26 %), PTA – 1 (5.26 %). Assessment of the state of the microvasculature after the endovascular intervention was carried out by measuring the TcPO<sub>2</sub> in standard places [1]. The severity of ischemia was assessed according to the Wifi classification [2]. Thus, the values of the TcPO<sub>2</sub> are in the range of less than 30 mmHg caused the 3rd degree of ischemia, 30–39 mmHg – 2nd degree, 40–59 mmHg – 1st degree [2]. Before surgical treatment the TcPO<sub>2</sub> values were at a critical level with an average value of  $12.01 \pm 8.43$  mm Hg. On the 1st day after the operation, the indicator increased and averaged  $16.79 \pm 8.52$  mm Hg. Further, a month after the operation, an increase in pO<sub>2</sub> was noted and the average value for patients reached  $28.58 \pm 12.79$  mm Hg. It should be noted that a month after the operation, there was a warming of the skin of the lower limb, a decrease in the intensity of pain at rest in more than half of the cases. According to ultrasound data after endovascular therapy, the effectiveness of endovascular revascularization was 71.43 % (5 out of 7 cases) for EIA, 50.0 % (1 out of 2 cases) for CFA, 57.14 % (4 out of 7 cases) for SFA, on popliteal artery (POP) 0 % (2 cases), on ATA in 33.33 % (2 of 6 cases), on PTA in 50 % (1 of 2 cases), on peroneal artery (PA) in 100 % (1 of 1 cases), on tibio-peroneal trunk (TPT) in 0 % (1 case out of 1). The efficiency of endovascular revascularization in the iliac-femoral segment is higher than in the tibia-foot segment. This can be explained by the fact that the lesion of the tibia-foot segment is more typical for patients suffering of concomitant diabetes mellitus. Diabetes mellitus predominantly affects small-caliber vessels; multifocal arterial stenoses are also characteristic. In this regard, the number of repeated restenoses in this group of patients is much higher.

*Conclusion:* The increase in the transcutaneous partial pressure of oxygen (TcPO<sub>2</sub>) the next day after surgery occurs due to dilatation of the stenotic segment, while the increase in tension occurring in a month after surgery is well correlated with a clinical picture of the disease, namely, with symptoms such as decrease in the pain intensity at night as well as warming of the skin of the operated limb. According to the ultrasound test conducted before and after surgery, there is a greater degree of the iliofemoral segment dilatation on compared to the popliteal-tibia segment. Such difference occurs due to small dimension of target arteries and concomitant diabetes mellitus suffered by the patients, which predominantly affects smaller vessels.

*Acknowledgements:* The study is supported by the Ministry of Education and Science of Russia (theme No. FWNR-2022-0026).

## References

1. Terskov D.V., Shnayder N.A. Transcutaneous oxymethria is the method of critical ischemia diagnostics at patients with diabetic foot syndrome and chronic severe ischemia of lower limbs. *Bull Clinical Hosp* No. 51. 2010;3(8):56-61 (in Russian).
2. Mills J.L. et al. The society for vascular surgery lower extremity Threatened limb classification system: Risk stratification based on wound, ischemia, and foot infection (Wifi). *J Vasc Surg Elsevier BV*. 2014;59(1):220-234.

# The resistome of human-associated microbial communities, its mobility and temporal dynamics

Starikova E.<sup>1\*</sup>, Manolov A.<sup>1</sup>, Selezneva O.<sup>2</sup>, Zoruk P.<sup>2</sup>, Ilina E.<sup>1,2</sup>

<sup>1</sup> *Institute of Systems Biology and Medicine, Rospotrebnadzor, Moscow, Russia*

<sup>2</sup> *Research Institute of Physical-Chemical Medicine, FMBA, Moscow, Russia*

\**hed.robin@gmail.com*

**Key words:** resistome, antimicrobial resistance genes, microbiome

*Motivation and Aim:* Complex microbial communities associated with the human body such as intestinal and respiratory microbiomes contain many antibiotics resistance genes (ARGs). Some of them are thought to be mobile and able to spread from one species to another. Considering rampant antibiotics resistance reported worldwide, we find it necessary to assess resistomes of human-associated microbial communities. In this study, we aimed to characterize sets of ARGs present in oropharyngeal and fecal samples of 182 patients using a targeted gene sequencing panel based on 4937 antibiotics resistance genes.

*Methods and Algorithms:* We have used sequencing reads obtained from 94 oropharyngeal and 88 fecal samples collected in different time points during the antibiotics course. Those reads were sequenced with Illumina HiSeq platform based on targeted gene sequencing panel of 4937 antibiotics resistance genes. We have aligned sequencing reads to the reference MEGARes v.2.0. database using bwa mem tool and calculated the numbers of mapped reads with samtools. Gene coverage calculation and the visualization of the results were carried out in R language. We have also assembled reads into contigs using metaspades v.3.15.4. In order to identify possible mobile elements in the proximity of antibiotics resistance genes, we have used a set of 5970 insertion sequences (ISs) from the ISFinder database.

*Results:* In each sample, 113 to 1440 AR-associated genes were covered with at least one read. The most abundant and common genes identified in fecal samples were class C beta-lactamases, lincosamide nucleotidyltransferases and macrolide phosphotransferases, while chloramphenicol acetyltransferase and quinolone resistance peptides Qnr were abundant in oropharyngeal samples. Class A beta-lactamases and various efflux pump genes were abundant both in fecal and oropharyngeal samples. Interestingly, most samples demonstrated no obvious temporal dynamics during the antibiotics. While analyzing assembly contigs, we have identified sequences presumably belonging to insertion sequence (IS) elements from IS1, IS6, IS5, IS30 and IS605/IS200 families. Of those contigs, the ones with the highest coverage contained putative IS5 family sequences in the proximity of class A beta-lactamase genes. Another set of contigs with high coverage contained putative IS6 family sequences in the proximity of streptothricin acetyltransferase genes.

*Conclusion:* The use of targeted gene sequencing panels based on the sequences of known resistance determinants allows us to assess the resistomes of human-associated microbial communities and explore the changes it might undergo over time. In our study, we have identified resistome patterns that might be specific to intestinal and oropharyngeal microbial communities. We have also shown that the resistomes of the microbial communities analyzed are likely to maintain stable composition over time. We have also identified several associations of ARGs with putative mobile genetic elements that require further exploration.

*Acknowledgements:* This study was supported by Governmental order No. 122030900064-9.

## Optical control of intercellular communication between platelets during activation

Starodubtseva E.S.\*, Moskalensky A.E.

*Novosibirsk State University, Novosibirsk, Russia*

\* *e.starodubtseva@g.nsu.ru*

**Key words:** platelet activation, UV light, calcium signaling

*Motivation and Aim:* Platelet activation is considered a cornerstone of hemostasis and thrombosis. One of the important features of this process is the feedback mechanism which allows platelets to activate each other. Current research is dedicated to the local platelet activation and investigation of activation spreading patterns between cells. In this study we used a previously developed method of activation via UV light [1]. Our hypothesis is that platelet activation in regions not illuminated by UV light occurs due to a positive feedback mechanism by ADP released from already activated platelets. Therefore, it is expected that the closer cells to the light source the faster will be an activation and the fluorescence intensity changes correspondingly.

*Methods and Algorithms:* In this study the “caged” analog of adenosine-diphosphate (CADP) is used to initiate reaction by releasing free ADP into the platelet-rich plasma sample triggered by UV light.  $\text{Ca}^{2+}$ -sensitive fluorescent label Fluo-4 is used to monitor changes in cytoplasmic calcium which increases during activation. A simple algorithm for estimating the activation rate was developed and tested. The procedure of estimating includes division of the frame into even rectangles, which subsequently constitute a matrix where each element is processed separately. Processing relies on obtaining the dependence of mean intensity on time and linear approximation of the fluorescence kinetics after the UV flash.

*Results:* Received coefficients have the meaning of speed of intensity change and agrees well with our hypothesis. On the basis of this data we can conclude that the closer to the boundary of non-illuminated region platelets are the lesser their activation rate will be. It can be assumed that platelets that were activated by cleavage of CADP become a secondary source of ADP.

*Conclusion:* The results obtained in the experiment clearly show intercellular communication between platelets. Further research will focus on mathematical model formulation, improvement of sample preparation and modification of the current set up system.

*Acknowledgements:* The study was supported by the Russian Science Foundation (grant No. 18-15-00049).

### References

1. Spiryova D.V., Vorobev A.Yu., Klimontov V.V., Koroleva E.A., Moskalensky A.E. Optical uncaging of ADP reveals the early calcium dynamics in single, freely moving platelets. *Biomed Opt Express*. 2020;11:3319-3330.

## Generation of stable breast cancer cell lines with overexpression of $\alpha\beta 3$ integrin

Sukhinina E. \*, Pershina A.

Siberian State Medical University, Tomsk, Russia

Tomsk Polytechnic University, Tomsk, Russia

\*suhininaev@mail.ru

**Key words:** integrin, model cell line, CRISPR/Cas9, gain of function

**Motivation and Aim:** Integrins are transmembrane receptors that mediate adhesion of cells to extracellular matrix.  $\alpha\beta 3$  integrin is known to be involved in angiogenesis and metastasis, and is aberrantly expressed on the endothelium of tumor blood vessels and on the membrane of many types of tumor cells [1, 2]. This receptor is a promising target for anticancer therapy, especially due to their ability to recognize and bind to RGD motif. One of the approaches is the modification of drugs and diagnostic agents with RGD peptides for their targeted accumulation in tumors [3]. The development and testing of such drugs requires the availability of a model cell line with enhanced expression of  $\alpha\beta 3$ . The aim of the study was to establish cell lines of mouse breast carcinoma 4T1 and human breast adenocarcinoma MDA-MB-231 with stable overexpression of integrin  $\alpha\beta 3$ .

**Methods and Algorithms:** To induce endogenous gene expression we used CRISPR-Cas9 guide RNA-directed synergistic activation mediator (SAM) system [4]. Integration of transgenes encoding components of the SAM complex into the cell genome using lentiviral vectors ensures stable activation of target gene expression. We constructed 11 recombinant plasmids encoding various sgRNA to the promoter regions of the *H. sapiens* ITGAV and ITGB3 and *M. musculus* itgav and itgb3 genes. T293 cells were used for production of lentiviral vectors. First, MDA-MB231 and 4T1 cell were transduced with dCas-VP64 and MS2-p65-HSF1 encoding lentiviruses and selected for stable expression. Then the selected cell lines were transduced with a mixture of lentiviral vectors expressing sgRNAs. Next, the antibiotic-resistant clones were selected and monoclonal cell lines were obtained by the limiting dilution method. The expression of  $\alpha$  and  $\beta 3$  integrin subunits was quantified by qRT-PCR and presence of  $\alpha\beta 3$  integrin on the surface of modified cells was confirmed by flow cytometry with FITC labeled Anti-Integrin alpha V beta 3 antibody.

**Results:** We generated two genome-modified cell lines of human breast adenocarcinoma MDA-MB-231<sub>27</sub> and mouse breast carcinoma 4T1<sub>27</sub> with over expression of  $\alpha\beta 3$  integrin. According to qRT-PCR data, the expression of  $\alpha$  and  $\beta 3$  subunits was upregulated 16.5-fold and 9-fold in MDA-MB-231<sub>27</sub>, 4-fold and 1.5-fold in 4T1<sub>27</sub> cell line respectively.

**Conclusion:** Two stable breast cancer cell lines with overexpression of  $\alpha\beta 3$  integrin were generated for testing diagnostic and/or therapeutic efficiency of RGD-conjugates.

**Acknowledgements:** The study was supported by Siberian State Medical University.

### References

1. Seguin L., Desgrosellier J.S., Weis S.M., Cheresh D.A. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends Cell Biol.* 2015;25:234-40.
2. Hamidi H., Ivaska J. Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer.* 2018;18:533-548.
3. Danhier F., Le Breton A., Préat V. RGD-based strategies to target alpha(v) beta(3) integrin in cancer therapy and diagnosis. *Mol. Pharm.* 2012;9(11):2961-73.
4. Konermann S., Brigham M.D., Trevino A.E. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* 2015;517:583-588.

## An increase in the number of CAG repeats in the huntingtin gene enhances pathological ultrastructural aberrations in the cells and neurons

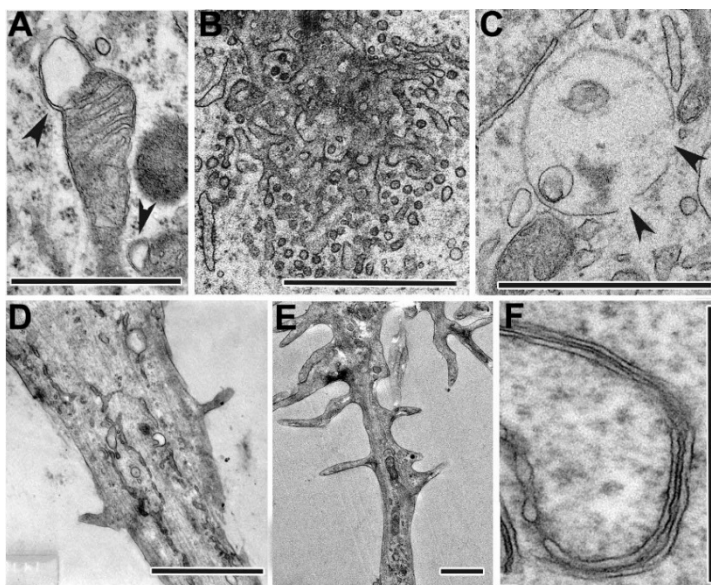
Suldina L.A., Morozova K.N., Malankhanova T.B., Kiseleva E.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *suldina@bionet.nsc.ru*

**Key words:** Huntington's disease, CAG repeats, ultrastructural aberrations

Huntington's disease is a hereditary neurodegenerative illness caused by an increased CAG repeat number ( $> 35$ ) in *HTT* resulting in elongation of the polyglutamine tract in the huntingtin protein and contributing to huntingtin aggregation. We compared previously documented ultrastructural aberrations in neurons and other cells featuring different numbers of CAG repeats: patient specific neurons (PSN) with 42–45 CAG repeats [1], neurons with transgenic insertion (TN) of 69 CAG repeats [2], and HEK293 cells with insertion of 150 CAG repeats [3]. Severe damage of mitochondria (Fig. 1, A), hypertrophy and vesiculation of the endoplasmic reticulum (see Fig. 1, B) and Golgi apparatus, disrupted autolysosomal membrane integrity (see Fig. 1, C) were noted in all studied mutant cells together with lower number of spines on dendrites in mutant neurons (see Fig. 1, D).



**Fig. 1.** Comparative analysis of morphological defects in mutant cells with an increased number of CAG repeats in the huntingtin gene. A – defective mitochondria with membrane disruption (arrowheads); B – ER membrane hypertrophy and vesiculation; C – breaks of autolysosome membrane integrity (arrowheads); D – dendrite with sparse spines; E – fragment of atypical neuron with a high frequency of spines distribution on hypertrophied dendrite; F – quadruple membranes of smooth ER in the cytoplasm HEK293 cell with insertion of 150 CAG repeats. Scale bar – 1  $\mu\text{m}$

Compared to the patient specific cells, in the cell of transgenic model were additionally observed greater neuronal heterogeneity at various stages of degradation, and appearance new atypical neurons with greater length and frequency of spines distribution on hypertrophied dendrites (see Fig. 1, E). In the HEK293 cells (largest number of repeats), numerous abnormal four-layer endoplasmic-reticulum membranes were identified for the first time in the cytoplasm (see Fig. 1, F). The study offers new information about the dynamics of progressive pathological changes in mutant cells and shows correlations with a greater CAG repeat number.

*Acknowledgements:* The work was supported by budget project FWNR-2022-0015.

### *References*

1. Nekrasov E.D., Vigont V.A., Klyushnikov S.A., Lebedeva O.S., Vassina E.M., Bogomazova A.N., ..., Kiselev S.L. Manifestation of Huntington's disease pathology in human induced pluripotent stem cell-derived neurons. *Mol Neurodegener.* 2016;11:27.
2. Malankhanova T., Suldina L., Grigor'eva E., Medvedev S., Minina J., Morozova K., ..., Malakhova A. A human induced pluripotent stem cell-derived isogenic model of Huntington's disease based on neuronal cells has several relevant phenotypic abnormalities. *J Pers Med.* 2020;10(4):215.
3. Morozova K.N., Suldina L.A., Malankhanova T.B., Grigor'eva E.V., Zakian S.M., Kiseleva E., Malakhova A.A. Introducing an expanded CAG tract into the huntingtin gene causes a wide spectrum of ultrastructural defects in cultured human cells. *PLoS One.* 2018;13(10):e0204735.

## Vaccination with dendritic cells pulsed with glioma killed by photodynamic therapy induces efficient antitumor immunity

Turubanova V.D.<sup>1,2\*</sup>, Savyuk M.O.<sup>1</sup>, Efimova I.<sup>2,3</sup>, Mishchenko T.A.<sup>1,2</sup>,  
Vedunova M.V.<sup>1</sup>, Krysko D.<sup>1,2,3</sup>

<sup>1</sup> National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

<sup>2</sup> Cell Death Investigation and Therapy (CDIT) Laboratory, Department of Human Structure and Repair, Ghent University, Ghent, Belgium

<sup>3</sup> Cancer Research Institute Ghent, Ghent, Belgium

\* vikaturu@mail.ru

**Key words:** immunogenic cell death, photodynamic therapy, dendritic cells, antitumor immunity

Currently, there are no optimal methods for treated oncological diseases that would ensure the complete recovery of the patient and guarantee the absence of relapse. The immunogenic cell death (ICD) model involves the activation of an antitumor immune response in response to the emission of damage-associated molecules (DAMPs) due to cell death [1]. Photodynamic therapy (PDT) is one of the cancer therapies that can cause ICD.

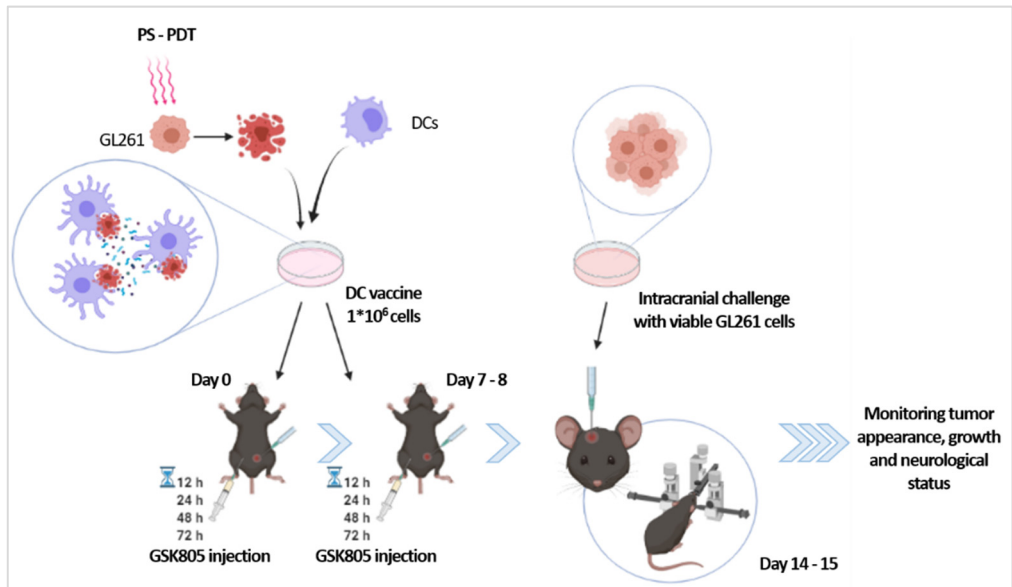
We have chosen a photosensitizer used in clinical protocols – Photosens (Niopic, Russia). The effectiveness of photosens (PS) has already been shown *in vitro* on the MCA205 line, and a vaccine based on cells induced by photosens-PDT protected mice from tumor development in a syngeneic model [2]. First, we found that, at doses of 20 J/cm<sup>2</sup>, they effectively induce regulated cell death in a mouse glioma cell line (GL261). The intracellular distribution of PS was studied using a laser scanning microscope. Tumor cells undergoing PS-PDT-induced cell death emit calreticulin, HMGB1, and ATP, and these are efficiently taken up by bone marrow dendritic cells, which then mature, become activated, and produce IL-6.

Using dying cancer cells induced by photosens-PDT, we demonstrate the efficient vaccination potential of ICD *in vivo*. In the model, activated dendritic cells have become a powerful stimulus for the activation of T-cell populations.

We assessed the survival rate of animals, analyzed the neurological status of the animals during the experiment, visualized the tumor in the brain by magnetic resonance imaging, assessed the size of developing tumors.

The ICD-based DC vaccine demonstrated high efficacy in subcutaneous and orthotopic mice models. By using RNA sequencing to characterize the transcriptional program of ICD based DC vaccine, we identify a high expression of 1357 genes and by assessing their functional relevance in immune responses we observed a robust induction of Tgfb-3, IL-6, IL-23a implicating Th17 immunity. We found that the increased intratumor prevalence of Th17 cells linked genetic signatures is associated with good patient prognosis on The Cancer Genome Atlas low grade glioma cohort. The efficacy of ICD based DC vaccines was significantly reduced by antagonist of ROR $\gamma$ t inhibiting the development of Th17 response.





Thus, this indicates the immunogenicity of the death of tumor cells after photodynamic exposure and is the basis for a therapeutic protocol for the using of a dendritic vaccine.

*Acknowledgements:* The study was supported by a grant from Russian Science Foundation (RSF, project No. 22-25-00716, <https://rscf.ru/project/22-25-00716/>).

### References

1. Krysko D.V., Garg A.D., Kaczmarek A., Krysko O., Agostinis P., Vandenabeele P. Immunogenic cell death and DAMPs in cancer therapy. *Nat Rev Cancer*. 2012;12(12):860-75.
2. Turubanova V.D., Balalaeva I.V., Mishchenko T.A. et al. Immunogenic cell death induced by a new photodynamic therapy based on photosens and photodithazine. *J Immunother Cancer*. 2019;7:350. doi: 10.1186/s40425-019-0826-3.

# Construction of comprehensive amino acid networks from Molecular Dynamics trajectories and tumor mutational profile of human sodium transporter NaPi2b

Vlasenkova R.<sup>1\*</sup>, Kozlova A.<sup>1</sup>, Akhberova N.<sup>1</sup>, Kiyamova R.<sup>1</sup>, Bogdanov M.<sup>1,2</sup>

<sup>1</sup> *Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

<sup>2</sup> *McGovern Medical School, The University of Texas Health Science Center, Houston, USA*

\* *r.mukhamadeeva@yandex.ru*

**Key words:** NaPi2b, amino acid networks (AAN), cancer

*Motivation and Aim:* The active roles of individual amino acids across the proteome is determined by many factors including nearby amino acids, secondary structure, posttranslational modifications and interacting ligands. Moreover, chemical properties of individual amino acids are utilized in drastically different contexts to generate required functions: for example, cysteines can form either disulfide or hydrogen bonds or stay as free thiols while the charged amino acids can make or not use of their charge. Each tumor is characterized by a unique profile of mutated genes. If tumor mutational profile of the proteins is available, we can match these experimental readouts and capture the three-dimensional impact of intra-protein amino acids short and long-range interactions on the development of oncological diseases in the accordance with predicted altered protein activity or stability. In addition, the received information will allow us to develop more effective personalized therapeutic agents taking into account the individual mutational status of each patient. Comparison of the AAN (amino acid networks) with different molecular characteristics-using Molecular Dynamics allows us to explore the contribution of certain amino acids to structural stability and conformational mobility of proteins. We tackled recently the mutation profiles of patients and identified 111 functionally significant mutations of the NaPi2b protein in tumor samples [1]. By mapping AANs within the tumor mutational landscape of the protein, we examined mutational consequences and scored deleterious cancer-related impact of a substitutions on the protein structure of NaPi2b.

*Methods and Algorithms:* The NaPi2b structure was predicted ab initio and thermodynamically stabilized by the method of Molecular Dynamics. Disulfide bonds (aa 303-328, aa 328-350) were introduced by controlled Molecular Dynamics. Molecular Dynamics simulation was carried out using the NAMD program, Molecular Dynamics parameters were obtained using the bio3d R package and the VMD software.

Considering the entire complex of amino acid parameters in the construction of networks, a new algorithm for determining the formula of the node weight has been developed. Weighted graphs were constructed, the weight in which was determined by one of the following likelihood parameters: residence time for amino acid residue staying continuously in forbidden regions, residence time for amino acid residues interacted with water, RMSF (Root Mean Square Fluctuation), the average conservation of positions of each amino acid.

An unweighted graph was also constructed as a control graph. The parameters of the nodes of the weighted graphs were compared with similar parameters of the nodes of the

control unweighted graph using the Kruskal-Wallis test with the pairwise Wilcoxon test adjusted for the multiple testing. The coefficient for each amino acid parameter was calculated based on the significance level obtained during the comparison. The coefficients and values of the parameters determine the weight formula.

To compare AANs adjacency matrices were obtained considering the obtained weights. Assessment of interactions clustering in NaPi2b structure is performed based on adjacency matrices using proximity propagation. The similarity score was calculated for each cluster. The similarity score of the graphs was calculated as the median of the similarity scores of clusters.

The set of mutations found in tumor samples were mapped on the obtained graphs. The procedures for determining the node weight formula, the graph comparison, and the mutation mapping were fully automated.

*Results:* The developed robust clustering algorithm was utilized to conduct a comparative analysis of the AANs of NaPi2b with the absence and the presence of potential disulfide bonds (303–328, 328–350). The AANs of NaPi2b were mostly similar to each other, but the considerable difference was found between the control and the AAN with disulfide bond 303–328.

Additionally, we identified the amino acid residues from tumor mutational dataset and had the most impact on the structure of NaPi2b. Critical amino acid residues are located in one of the transmembrane domains (217ALA, 218THR, 221ASP, 224ASN), in the largest extracellular domain (297SER, 299VAL), in one of the two pseudosymmetrically located loops (163SER) comprising serine rich QSSS functional motifs within protein transport core, and in the N-domain (90LYS). Mutations in these sites were discovered in samples of oncological diseases of stomach, lung, bowel, liver and brain.

*Conclusion:* The algorithm for determining the node weight formula, the graph comparison and the mutation mapping were developed and automated. And the comparative analysis of the AANs of NaPi2b with the absence and the presence of potential disulfide bonds were carried out, which results in the list of amino acid residues which were found altered in tumor samples and had the most impact on the structure of NaPi2b protein.

*Acknowledgements:* This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by Russian Science Foundation support (grant No. 20-14-00166).

## References

1. Vlasenkova R. et al. Characterization of *SLC34A2* as a potential prognostic marker of oncological diseases. *Biomolecules*. 2021;11(12):1878. doi: 10.3390/biom11121878.

## CDH13 and it's signaling partners in the functioning of the endothelium cells

Yamileva K.I.<sup>1\*</sup>, Mirzaeva S.<sup>1,2</sup>, Arbatsky M.<sup>1</sup>, Sysoeva V.Yu.<sup>1</sup>, Klimovich P.S.<sup>1,2</sup>, Semina E.V.<sup>1,2</sup>, Rubina K.A.<sup>1</sup>

<sup>1</sup> Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia

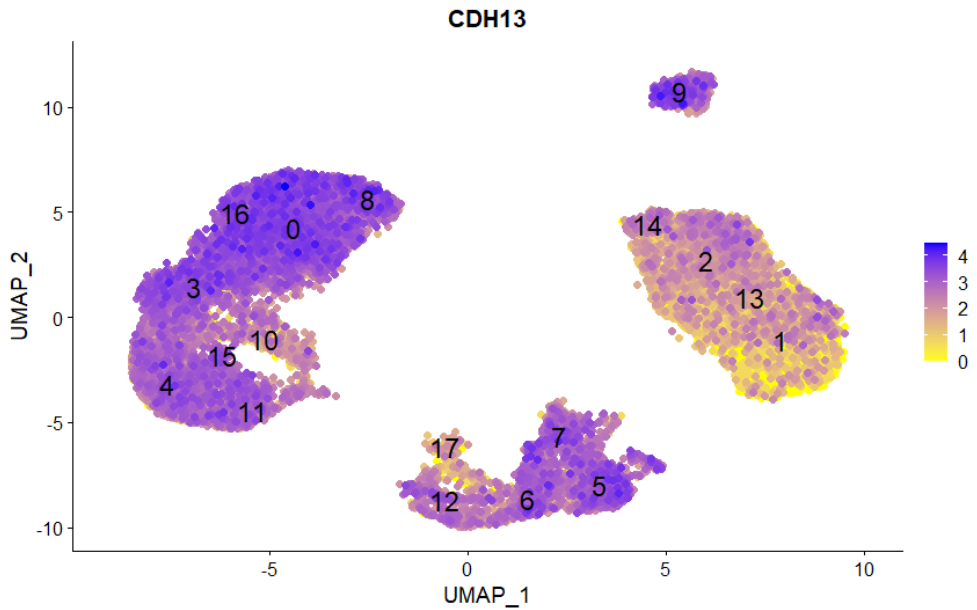
<sup>2</sup> Institute of Experimental Cardiology, National Cardiology Research Center of the Ministry of Health of the Russian Federation, Moscow, Russia

\*yamkamilla@gmail.com

**Key words:** CDH13, T-cadherin, STRING.ORG bioinformatics tool, endothelium, lung

*Motivation and Aim:* CDH13 (known as T-cadherin) is an important participant in such cellular processes as the formation of intercellular contacts, the dynamic regulation of morphogenetic processes in embryos and the maintenance of the integrity of adult body tissues. Cadherins function as membrane receptors mediating signals from the outside, activating small GTPases and the beta-catenin/Wnt pathway and leading to dynamic reorganization of the cytoskeleton and phenotype changes. T-cadherin is abundantly expressed on the lumen surface of endothelial cells, but not at the sites of intercellular contact; increased expression is observed in such pathological conditions as atherosclerotic lesions of the human aorta and in endothelial cells of the vascular network of the tumor. However, the exact signaling partners and adapter proteins for CDH13 have yet to be determined.

*Methods and Algorithms:* This work is based on the existing reference atlas of endothelial cells in the normal lung (Fig. 1), obtained as a result of human RNA sequencing and subsequent iterative clustering with analysis of differential expression of endothelial cells (GSE164829, "Lung and primary pulmonary endothelial single cell RNA seq data" Single Cell RNAseq of Whole Lung Dissociates from control lungs and primary pulmonary endothelial cells) [1]. Using Illumina HiSeq 4000 Six lungs' atlas of scRNAseq datasets were reanalyzed and annotated to identify >15 000 vascular EC cells from 73 individuals. Using the CreateSeuratObject function, an object was created from the source files. Data normalization was carried out using the SCTransform function. After sequentially launching the RunPCA, RunUMAP, FindNeighbors, and FindClusters functions, cluster images were obtained using the DimPlot function. Differential expression analysis of EC revealed signatures corresponding to endothelial lineage, including panendothelial, panvascular, and subpopulation-specific marker gene sets. To find and identify cellular targets and signaling partners of T-cadherin cascades, using FindAllMarkers we analyzed 17 clusters of endothelial cells with different expression of T-cadherin using the STRING.ORG bioinformatics tool by subclustering the protein pool into functional units and tagging biological processes using keywords: "endothelium", "vascular", "vessel" and "blood".



**Fig. 1.** Atlas of endothelial cells in the normal lung

**Results:** According to preliminary results, targets common to all endothelial cell clusters involved in the formation and regulation of the lung endothelium in humans and interacting with CDH13 were identified. Among them are nine proteins: CTNNB1 (Catenin beta-1, or beta-catenin  $\beta$ -catenin), CDH5 (Cadherin 5 or VE-cadherin), CDH2 (Cadherin 2 or N-cadherin), CAV1 (caveolin 1), DNMT1 (DNA (cytosine-5)-methyltransferase 1), BRCA1 (BRest Cancer gene 1), SPFR1 (Secreted Frizzled Related protein 1), JUP (Junction Plakoglobin), ADIPOR2 (Adiponectin Receptor 2), which through protein-protein interaction could potentially influence endothelial functions such as positive regulation of endothelial cell differentiation, regulation of vasculature development, blood vessel development and morphogenesis.

**Conclusion:** Thus, in order to understand the molecular mechanisms and interactions of proteins involved in the physiological and pathological conditions of the human lung endothelium, a bioinformatics analysis was performed to search for new signaling pathways and proteins critical for the functioning of the endothelium that interact with CDH13, as a result of which 9 new targets were found.

**Acknowledgements:** This research was supported by the Russian Foundation of Basic Research (Project No. 20-24-60029).

### References

1. Schupp J.C. et al. Integrated single-cell atlas of endothelial cells of the human lung. *Circulation*. 2021;144(4):286-302. doi: 10.1161/CIRCULATIONAHA.120.052318.

## Renal pathologies development during experimental opisthorchiasis

Zaparina O.<sup>1\*</sup>, Kapushchak Y.<sup>1</sup>, Pakharukova M.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* zp.oksana.93@gmail.com

**Key words:** Opisthorchiasis, *Opisthorchis felineus*, systemic diseases, kidney failure, heart failure

**Motivation and Aim:** Motivation and Aim: Opisthorchiasis caused by trematode *Opisthorchis felineus* (Rivolta, 1884) is the most common helminthiasis in Western Siberia. This disease is accompanied by serious liver lesions. Currently, there is indirect evidence that opisthorchiasis, like other parasitic infections, can cause damage to organs that are not related to the its direct localization and migration. Based on this, we suggested that chronic opisthorchiasis can cause structural and functional disorders of the kidneys and heart due to the components contained in the excretory-secretory product in combination with mechanical damage to the epithelium of the bile ducts by helminths [1, 2]. The aim was comprehensive study of renal and heart failure markers in experimental opisthorchiasis renal and heart failure markers in experimental opisthorchiasis.

**Methods and Algorithms:** Experimental opisthorchiasis *in vivo* on golden hamsters *M. auratus* (1 month to 1.5 years). Renal and heart pathological changes, including cortical tubular casts, casts in medullar tubules, mesangial matrix expansion and cardiomyocytes structure were assessed using semi- quantitative histological analysis. To assess the functionality of the kidneys, biochemical parameters in the blood serum and urine were measured.

**Results:** Renal tubules casts accumulation and mesangial matrix expansion are linear with the during of infection. In addition, it was found that infection with *O. felineus* was accompanied by an increase in the concentration of creatinine and KIM-1 in the blood serum; protein, KIM-1 and 8-OH-dG in the urine of experimental animals. Biochemical markers of metabolic disorders were also found: blood glucose concentration, cholesterol and triglycerides was increased. Pathomorphological changes in the kidneys correlated with changes in biochemical parameters. INo significant structural and functional changes of the heart during the opisthorchiasis inection in hamsters were revealed.

**Conclusion:** The mechanisms underlying renal pathology in opisthorchiasis remain unclear, we can conclude that *O. felineus* infection causes gradual progression of glomerulopathy accompanied by tubulopathy.

**Acknowledgements:** The study is supported by the Russian Science Foundation (No. 18-15-00098).

### References

1. Gouveia M.J., Pakharukova M.Y., Laha T., Sripa B., Maksimova G.A., Rinaldi G., Brindley P.J., Mordvinov V.A., Amaro T., Santos L.L., Correia da Costa J.M., Vale N. Infection with *Opisthorchis felineus* induces intraepithelial neoplasia of the biliary tract in a rodent model. *Carcinogenesis*. 2017;38(9):929-937.
2. Pakharukova M.Y., Zaparina O.G., Kapushchak Y.K., Baginskaya N.V., Mordvinov V.A. *Opisthorchis felineus* infection provokes time-dependent accumulation of oxidative hepatobiliary lesions in the injured hamster liver. *PLoS One*. 2019;14(5):e0216757.

# Natural selection against insufficiency in sarcomeric proteins reduces the risks of cardiomyopathy in humans versus chimpanzees as conventionally the nearest wild congeners, for whom it is the most common cause of spontaneous mortality

Zolotareva K.<sup>1\*</sup>, Chadaeva I.<sup>1</sup>, Khandayev B.<sup>1</sup>, Bogomolov A.<sup>1</sup>, Oshchepkov D.<sup>1</sup>, Nazarenko M.<sup>2</sup>, Ponomarenko M.<sup>1</sup>, Kolchanov N.<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*

\* ka125699ri@yandex.ru

**Key words:** human, gene, promoter, TATA-box, SNP, expression alteration, disease, candidate SNP-marker, bioinformatics model, *in silico* prediction, *in vitro* verification

*Motivation and Aim:* Cardiomyopathy is the most common cause of spontaneous death in chimpanzees, who is our nearest wild congeners [1].

*Methods and Algorithms:* Using our previously published Web service SNP\_TATA\_Comparator, we analyzed single-nucleotide polymorphisms (SNPs) within 70 bp proximal promoters of eight human sarcomeric protein genes, as shown in Figure.

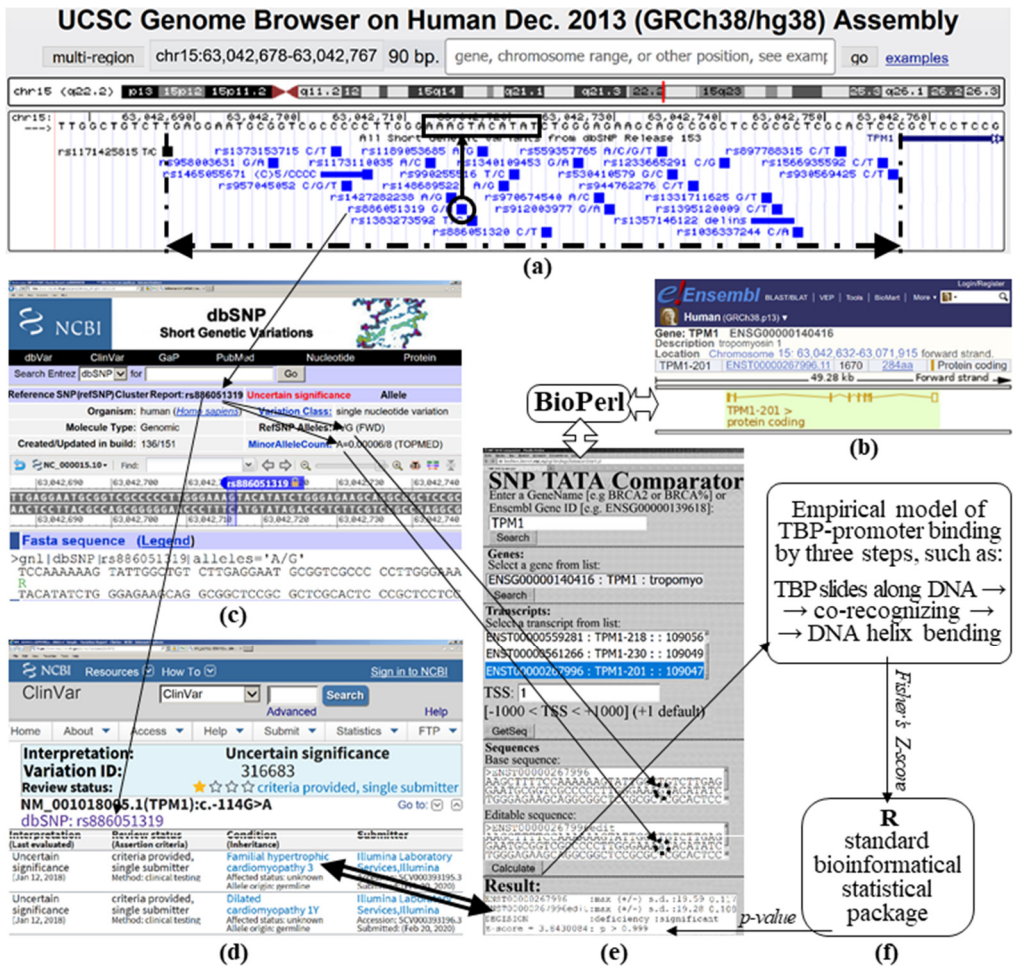
*Results:* After analyzing 637 such SNPs, we identified 92 and 54 candidate SNP markers for worsened and relieved respectively, hypertrophic cardiomyopathy (HC), as susceptibility to heart failure. Among them, 84 and 62 candidate SNP markers overexpressing and underexpressing respectively, these genes as natural selection against the human sarcomeric protein deficit, being resistance to heart failure. This simultaneity of susceptibility and resistance to heart failure means disruptive natural selection of the considered human genes as if humans could exposed self-domestication that is debatable. We tested this using public 5591 differentially expressed gene (DEGs) of domestic versus wild animals. In the domestic animals, the overexpressed sarcomeric protein DEGs dominated the underexpressed ones (29 vs 15 respectively) just as natural selection against the human sarcomeric protein deficiency. Amounts of the domestic animal DEGs corresponding to relieved HC condition in humans surpassed those in wild animals (8 vs 1).

*Conclusion:* Thus, natural selection against the sarcomeric proteins deficit reduces the risks of cardiomyopathy in humans versus chimpanzees as the nearest wild congeners, for whom it is the most common cause of spontaneous death.

*Acknowledgements:* We thank the Multi-Access Center “Bioinformatics” for the use of computational resources supported by Russian government project FWNR-2022-0020.

## References

1. Laurence H., Kumar S., Owston M.A., Lanford R.E., Hubbard G.B., Dick E.J.Jr. Natural mortality and cause of death analysis of the captive chimpanzee (*Pan troglodytes*): a 35-year review. *J Med Primatol.* 2017;46(3):106-115. doi: 10.1111/jmp.12267.



**Fig. 1.** An example of the results SNP\_TATA\_Comparator in the case of the clinically proven hypertrophic cardiomyopathy-related SNP marker (rs886051319) increasing risks of this disease via TPM1 downregulation: a) The UCSC Genome Browser visualizes a 70 bp promoter (double-headed dash-and-dot arrow) of the human TPM1 gene where there is only one TBP-binding site (framed) and SNP rs886051319 with an arrow pointing to it (circle); (b) The Ensembl database. (c) The dbSNP database. (d) The ClinVar database. (e) Our prediction of TPM1 deficiency (textbox “Result”: row 3) caused by SNP rs886051319 (dotted circles) by means of its both alleles (i.e., ancestral and minor ones as input data within two textboxes “Basic sequence” and “Editable sequence,” respectively). Double-headed open arrows depict how our Web service SNP\_TATA\_Comparator retrieved the ancestral DNA sequence of the human TPM1 promoter from the Ensembl database using the Bioperl toolkit. (f) Our bioinformatics model of the TBP–promoter binding in three steps, Double-headed double arrow, a consistency between the clinical data on rs886051319 as a clinically proven SNP marker of both familial hypertrophic cardiomyopathy 3, dilated cardiomyopathy 1Y, which are within the ClinVar database on one hand and the service SNP\_TATA\_Comparator result that the minor allele of this SNP can cause TPM1 underexpression linked clinically with hypertrophic cardiomyopathy on another hand



# 9 Symposium “Mathematics, bioinformatics and systems computational biology of COVID-19”



- 9.1 Section “Mathematical immunology” [865](#)
- 9.2 Section “Mathematical epidemiology” [894](#)

## Detection of recombination events in coronavirus genomes from the subgenus sarbecovirus

Arefieva N.<sup>1\*</sup>, Bukin Y.S.<sup>1,3</sup>, Dzhioev Y.P.<sup>2</sup>, Zlobin V.I.<sup>2</sup>

<sup>1</sup> Irkutsk State University, Irkutsk, Russia

<sup>2</sup> Limnological Institute, SB RAS, Irkutsk, Russia

<sup>3</sup> Irkutsk State Medical University, Irkutsk, Russia

\*arefieva.n4@gmail.com

**Key words:** Recombination events, Sarbecoviruses, SARS-CoV-2

*Motivation and Aim:* The pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) poses a great threat to public health. Genetic recombination is one of the leading factors in the variability and evolution of viruses. Detection of recombination events in coronavirus genomes will provide new data on the possible origin and variability of SARS-CoV-2.

*Methods and Algorithms:* The complete genome sequences of the subgenus Sarbecovirus were downloaded from the NCBI GenBank database. The whole genomes of different lines of Sars-Cov-2 were downloaded from the GISAID database. The final sample contained 321 genomes. The annotation of open reading frames in genomes was performed using VADR (Viral Annotation DefineR) program. Multiple sequence alignment of complete genomes was performed in MAFFT v.7. The protein-coding genes 1ab, S, E, M, N present in all genomes present in all genomes were separated from whole-genome sequences. Each protein-coding gene was aligned according to the codon position using RevTrans 2.0. The aligned protein-coding parts were assembled into a fasta file. To detect potential recombination events in complete genomes, 9 algorithms included in RDP (Recombination Detection Program) v.5.5 software package were used (RDP, Genconv, Bootscan, Maxchi, Chimaera, SiSscan, PhylPro, LARD, 3Seq). The coordinates of the found recombination positions in the complete genomes were translated into coordinates for the coding part. The R script was written to verify the results and to separate recombination events from adaptive evolution and other evolutionary processes.

*Results:* Using different algorithms, 239 potential recombination sites were found. Results with insignificant p-value obtained by more than three algorithms were discarded. As a result, the number of recombination events decreased to 177. Most of the potential recombination events are located in the spike (S) protein gene, which is more susceptible to variability than other structural genes. The recombination events were detected between SARS-CoV-2, SAR-Cov-1, Bat coronavirus and Pangolin coronavirus strains.

# Genetic traits of the delta variant in the Russian Federation and its spread during the epidemic caused by the SARS-CoV-2 virus

Ayginin A.A.<sup>1\*</sup>, Vinogradov K.S.<sup>1</sup>, Bulusheva I.A.<sup>2</sup>, Matsvay A.D.<sup>1</sup>, Mikhaylov I.M.<sup>1</sup>, Luparev A.R.<sup>1</sup>, Polyakova V.A.<sup>1</sup>, Shipulin G.A.<sup>1</sup>

<sup>1</sup> *Centre for Strategic Planning and Management of Biomedical Health Risks of the Federal Medical Biological Agency, Moscow, Russia*

<sup>2</sup> *Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

\* *AAyginin@cspmz.ru*

**Key words:** COVID-19, SARS-CoV-2, delta, mutations, G1048T, C27527T

*Motivation and Aim:* The COVID-19 epidemic caused by the SARS-CoV-2 virus has affected most countries globally and is widely studied by scientists worldwide. The first cases of infection with this virus were identified in December 2019 in China in Wuhan Province [1]. Over the past two years, the virus has repeatedly mutated, producing many variants that have spread worldwide. Evolution of the delta variant of coronavirus may occur independently in different countries and have specific characteristics. This study describes the characteristic genetic variants of the delta variant of coronavirus in Russia.

*Methods and Algorithms:* To analyze the genetic variants of the delta variant of coronavirus, both samples from the GISAID database (database version dated July 31, 2021) and samples collected by the Centre for Strategic Planning of FMBA laboratory in various regions of Russia were used. For the samples sequenced by the Centre for Strategic Planning of FMBA laboratory, the virus genomes were assembled in the following stages: trimming of adapters and low-quality bases; primers trimming; mapping of reads to the MN908947 reference genome; read filtering; SNV/indel calling; consensus building. For samples obtained from the GISAID database, filtration was conducted according to the following algorithm: sequences were aligned to the reference genome, the first and last 100 nucleotides were removed, sequences were filtered by length (at least 29,000 bp) and the number of uncertain bases (not exceeding 3000 N). The resulting genomes were clustered using the cd-hit [2] program, which reduced the database size for further analysis. Phylogenetic tree was constructed using the RAxML software [3].

*Results:* Russian isolates of the delta variant of the virus were found in 4 different branches of the phylogenetic tree. At the same time, most (95 %) of the Russian samples of the delta variant formed a separate branch, defined by G1048T and C27527T substitutions. Unlike other countries, where the proportion of G1048T mutation among delta strain samples did not exceed 7 % on average at the beginning of the spread of the delta variant, the corresponding value in Russia increased up to 100 % in early May 2021. The same result was observed for the C27527T substitution. By the end of July 2021, the proportion of G1048T+C27527T substitution in the Russian samples was 93 % while this number in the worldwide samples did not exceed 1 %. One of the countries in which these two mutations were also observed shortly before Russia could become a source of import of the delta strain to the country. Assuming that the candidates for the import of the variant to Russia are Italy, Germany, Israel, Singapore, Japan, and Switzerland.

*Conclusion:* The first double mutant sample found in Russia was obtained on April 19, 2021. The tree constructed using the RAxML program suggests that the G1048T mutation appeared first, after which the G27527T substitution appeared inside this cluster. However, this conclusion requires an in-depth study to confirm. These mutations may affect coronavirus transmissivity. The G1048T mutation (NSP2:K81N) is located in the NSP2 protein, and its function has not been properly studied to date. The protein participates in viral replication and forms a protease complex required to cleave the ORF1ab polyprotein [4]. Notably, the NSP2 protein is considered a potential target for COVID-19 medical therapy; therefore, mutations arising in it may affect the effectiveness of the medicines being developed. The C27527T mutation (ORF7A:P45L) is located in the ORF7a protein. This protein interacts with the immune cells and may affect the strength of the immune response. Recently, Zhou et al. proved that ORF7a SARS-CoV-2 binds to CD14+ monocytes [5]. The data obtained in this study prove that the ORF7a protein is a modulating factor for immune cells binding to the virus particles and an inflammatory response, indicating the possibility of its use as a therapeutic target for the COVID-19 treatment. Based on the analysis of emerging mutations, tracing the evolution of this variant in Russia is possible. Additionally, data on existing mutations and their combinations may be used to determine the sources of the spread of coronavirus in different countries.

### *References*

1. Wu F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. DOI 10.1038/s41586-020-2008-3.
2. Li W., Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659. doi: 10.1093/bioinformatics/btl158.
3. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinforma Oxf Engl*. 2014;30(9):1312-1313. doi: 10.1093/bioinformatics/btu033.
4. Raj R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem Biophys Rep*. 2021;25:100847. doi: 10.1016/j.bbrep.2020.100847.
5. Zhou Z. et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience*. 2021;24(3):102187. DOI 10.1016/j.isci.2021.102187.

## Epigenetics behind SARS COV2

Basu A.

*Department of Biochemistry, Gurudas College, Affiliated to University of Calcutta, Calcutta, India  
basuanamikaami@gmailcom*

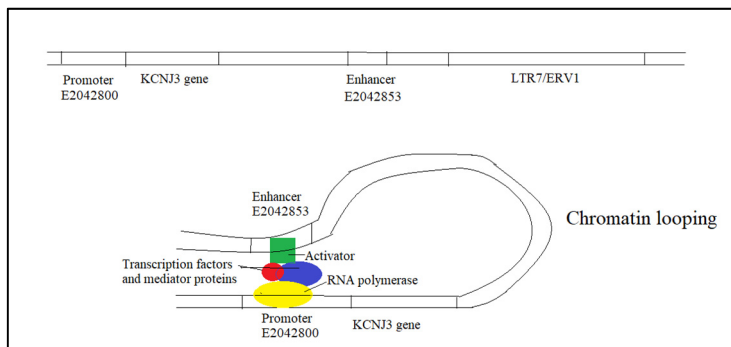
**Key words:** Human endogenous retroviruses (HERV) sequences, distal enhancer sequence, hypokalemia, *KCNJ3* gene, SARS CoV2

*Motivation and Aim:* There is a 100 % sequence similarity between SARS CoV2 and human genome for seventy-one nucleotide sequence. In human genome this seventy-one-nucleotide sequence, is located from region (154,867,731–154,867,802) of chromosome 2 with no gap identity. This region is a part of HERV-H Chr\_2q24LTR repeat region (154,867,204–157,867,802) with sequence ID: (MT992309.1). Human endogenous retroviruses (HERV) sequences, contain long terminal repeats regions (LTRs). These endogenous retroviruses (ERVs) can control the transcription of neighboring genes which are related with some diseases e.g., cancer etc. Long terminal repeats (LTRs) can function as either promoters or enhancers that may regulate adjacent human genes. In this study, the influence of this similar seventy-one nucleotide sequence, present in LTR, on neighboring human host gene expression, is analyzed.

*Methods:* From different databases of Human endogenous retroviruses (HERV) sequences, the detail characteristics of HERV-int elements, present in chromosome 2, have been searched. The Candidate Cis-regulatory Elements (cCREs) present in human genome for all cell types, have been identified for the specific ERV element. cCREs are the subset of representative DNase hypersensitive sites across ENCODE and Roadmap Epigenomics samples that are supported by either modifications (H3Kme3 and H3K27ac) or CTCF-binding data. From SCREEN database, the nearest protein-coding genes are investigated in the nearby location, for that chromosome region of cCRE. The data from RAMPAGE (RNA annotation and Mapping of Promoters for Analysis of Gene Expression) allows the annotation and the expression of gene at transcription start sites with nucleotide precision. The specific association of HERV/LTRs and transcription factors are frequently observed in human genome. Identification of TF binding sites (TFBSs) on the specific HERVH-int/LTR and DNase I hypersensitive sites (DNSs) for this specific chromosomal region are important for epigenetic control of genes within the chromosome. From JASPAR CORE database, several transcription factor binding sites are available in the regulatory region of genome. Chromosome folding data from in situ Hi-C and Micro-C XL experiments on H1-hESC (embryonic stem cells) and HFFc6 (foreskin fibroblasts) cell lines can be obtained from the 4D Nucleome Data Portal. With these data the number interactions are estimated between regions of genome. Finally, the effect of cis regulatory elements and control of gene expression on neighboring gene in SARS CoV2 patients has been elucidated.

*Results:* The search for HERV-int element, ERV type in human endogenous retrovirus database (HERVd) (<http://herv.img.cas.cz>) shows 1,337 entities. Among them, ERV element, ERV\_2041653 (LTR7, HERVH-int) is present in Chr 2 (154,867,229–154,874,560) region. By searching from different databases viz, HERVd, ENCODE etc., the details of ERV\_2041653 have been characterized. The presence of a candidate cis-regulatory element as distal enhancer sequence (154,865,721–154,866,057) has been recognized in this region. This enhancer sequence is present outside the 2kb of a transcription start site, so it is considered as distal enhancer sequence. The DNase activity and histone protein modification scores for this enhancer sequence are collected from SCREEN database. With the help of dbHERV-REs, the regulatory elements associated with HERV/LTR can be recognized.

Thus, TF binding sites (TFBSs) and DNase I hypersensitive sites, along with regulatory elements present in this retrotransposon are identified for LTR7. High scores for DNase activity and acetylation in histone, reveals that this distal enhancer sequence plays important role in epigenetic control of nearby gene expression in human genome. *KCNJ3* (potassium inwardly rectifying channel subfamily J member 3) is a protein coding gene is present in the nearby location (154,698,748–154,856,459) of this distal enhance sequence (ID E2042853). The gene product of this gene helps to transport  $K^+$  inside the cell. In case of tested COVID risk variants, two genetic variants are detected within the gene sequence of *KCNJ3* gene. Since hypokalemia is detected in SARS CoV2 infected patients, so there may be a correlation between occurrence of hypokalemia in COVID-19 patients and activation of *KCNJ3* gene during the viral infection. During transcription of *KCNJ3* gene, epigenetic control of this gene expression occurs through acetylation in histone protein present in nucleosome complex. Thus the packing in this complex becomes loose. This results in facilitating more transcription factors binding in the promoter region of the gene. Moreover that, the presence of DNSs on this chromosome region helps to access DNase I enzyme, results in the remodeled state of DNA for TF binding. In the heatmap of chromatin folding data, a high score is obtained between *KCNJ3* gene and distal enhancer region. This enhancer region is present in the downstream region of the gene sequence in chromosome. This high score between two regions suggests that they are probably located in the close proximity in chromosomal three-dimensional space. Both *KCNJ3* gene and enhancer sequence comes in close proximity by chromatin looping method as shown in Fig. 1.



**Fig. 1.** Chromosome 2 remodeling causing hypokalemia in COVID-19 patients

In the distal enhancer region (E2042853), STAT6 has been selected, from JASPAR CORE 2022 database. Signal transducer and activator of transcription 6 (STAT6) can act as activator molecule during epigenetic control of *KCNJ3* gene expression due to presence of the downstream enhancer (E2042853). The target prediction study for transcription factor reveals that the *KCNJ3* gene is one of the target genes of the STAT6 transcription factor. The binding of activator protein STAT6, with enhancer sequence, facilitates the RNA polymerase binding in the promoter region of *KCNJ3* gene. Thus, the rate of gene expression of *KCNJ3* gene increases, which rises the  $K^+$  concentration inside the cells, resulting in hypokalemia in blood serum of COVID-19 patients.

**Conclusion:** Due to presence of similar seventy-one nucleotide sequence, the incorporation of viral genome in human genome can occur. It facilitates the regulation of neighboring protein coding gene expression. A single nucleotide polymorphism (SNP) is found in the binding site of STAT6 TF. The altered allelic frequency from cytosine to thymine, causes inefficient TF binding in the promoter region of this *KCNJ3* gene. This SNP in human genome, controls the gene expression of *KCNJ3* gene and finally the frequency of occurrence of hypokalemia in COVID-19 patient, may be observed.

# Mathematical modelling of antiviral immune responses to SARS-CoV-2 infection

Bocharov G.<sup>1,2\*</sup>, Grebennikov D.<sup>1,2</sup>, Savinkov R.<sup>1,2</sup>

<sup>1</sup> *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

<sup>2</sup> *Institute of Computer Science and Mathematical Modelling, Sechenov University, Moscow, Russia*

\* *g.bocharov@inm.ras.ru*

**Key words:** SARS-CoV-2 infection, immune-physiological responses, mathematical model, sensitivity of dynamics, multi-modal therapies

*Motivation and Aim:* Infections of humans with SARS-CoV-2 known as COVID-19 disease represents a global biomedical problem. The course, severity and outcome of COVID-19 disease are characterized by enormous variability related to the specific kinetics of virus spreading, which depend on viral entry sites and tissue tropism, and the dynamics of immune reactions determined by the genetic make-up of the host and the spatial organization of immune system [1]. Consideration of the ‘numbers game’ and the ‘immune geography’ in quantitative terms is required for characterization of the host protection and its limits. A systems biology approach combining experimental/clinical data from murine- and human coronavirus infections in conjunction with mathematical modeling has been extensively advanced during the last few years. Systematic analysis of the advances and limitations of the so far developed mathematical models of SARS-CoV-2 infection is critical for implementation of the data-based mathematically-driven exploration of the pathogenicity mechanisms and for predicting the dynamics and severity of COVID-19.

*Methods and Algorithms:* We develop a multiscale multiphysics approach [2] to describe, analyze and predict the in-host COVID-19 dynamics. The model integrates the description of virus-host interaction at a single cell-, organ/tissue cell populations- and the whole organism levels. The respective mathematical modules describe the virus life-cycle, immune-physiological responses of the host, the immune cell motility and systemic transport processes. These are represented by a combination of deterministic and stochastic frameworks.

*Results:* The sensitivity of coronavirus disease severity to the dose of infection and the virus growth rate, and the limits of protection provided by the type I interferon (IFN) system that acts at the front line of the virus-host interaction are predicted [2]. Mathematical model of intracellular replication of SARS-CoV-2 has been developed and calibrated [3]. It was used to identify the stages of virus replication with a significant effect on the reduction of viral progeny produced by the infected cell. The stochastic version of the model predicted a strong inhibitory effects of IFN on viral progeny and a much lower sensitivity of the produced SARS-CoV-2 to the ACE2 binding affinity [4]. The dynamics, severity and outcome of COVID-19 in relation to the parameters of the immune-physiological reactions has been examined in quantitative terms using the calibrated mathematical model of SARS-CoV-2 infection.

*Conclusion:* The presented results identify novel targets for multi-modal therapies combining the antiviral drugs, reinvigoration of cellular immune responses and control of harmful of inflammatory responses. The models provide an interface between the

infection phenotypes and epidemiological studies by quantifying the dynamics of viral load which determines the infectivity of individuals at various stages of SARS-CoV-2 infection.

*Acknowledgements:* The study is supported by the Russian Foundation for Basic Research (grants 20-04-60157; 20-01-00352) and the Russian Science Foundation (grant 18-11-00171).

### *References*

1. Bocharov G. et al. Mathematical immunology: from phenomenological to multiphysics modelling. *Russ J Numer Anal Math Modelling*. 2020;35(4):203-213.
2. Bocharov G et al. Numbers Game and Immune Geography as Determinants of Coronavirus Pathogenicity. *Front Cell Infect Microbiol*. 2020;10:559209.
3. Grebennikov D. et al. Intracellular Life Cycle Kinetics of SARS-CoV-2 Predicted Using Mathematical Modelling. *Viruses*. 2021;13(9):1735.
4. Sazonov I. et al. Sensitivity of SARS-CoV-2 Life Cycle to IFN Effects and ACE2 Binding Unveiled with a Stochastic Model. *Viruses*. 2022;14:403.



# Genomic variation underlies the severity of COVID-19 clinical manifestation

Das R.<sup>1\*</sup>, Shenoy P.U.<sup>1</sup>, Bhavya B.<sup>1,2</sup>

<sup>1</sup> *Yenepoya Reserch Centre, Yenepoya (Deemed to be University), Karnataka, India*

<sup>2</sup> *Department of Bioinformatics, Manipal School of Life Science, Manipal Academy of Higher Education (MAHE), Karnataka, India*

\* *das.ranajit@gmail.com*

**Key words:** COVID-19 Host Genetics, Ancestral variation in COVID-19, Individual variation in COVID-19

*Motivation and Aim:* COVID-19 is characterized by a wide spectrum of clinical phenotypes ranging from complete absence of perceivable symptoms to severe disease manifestations. COVID-19 susceptibility, severity and recovery has demonstrated individual as well as population specific variations. Several studies have attributed this variation to the hosts' genetic make-up. As a follow-up to our previous study [1], here we performed an Exome-Wide Association Study (EWAS) employing a novel and unpublished whole exome data from University of Sienna, Italy.

*Methods:* The dataset consisted of 2,960 COVID-19 patients of various degrees of disease presentations. We restricted our analysis to 2,692 Europeans present in the dataset by performing principal component analysis with individuals from 1000 genome project. We performed GWAS in two cohorts. In cohort-1, the non-hospitalized patients were employed as controls and all hospitalized patients were employed as cases, and in cohort-2, the non-hospitalized patients and patients who did not require any respiratory support were employed as controls and the rest were employed as cases. We further performed ancestry analysis using ADMIXTURE v1.3.

*Results:* We identified 1501 and 362 genetic variants that were significantly distinct (Multiple-testing corrected  $P < 0.001$ ) between cases and controls in cohort-1 and cohort-2 respectively. Among these variants, 117 were common to both cohorts, which were found to be associated with pathways governing host immunity, such as interferon gamma signaling, toll like receptor cascade, antigen processing, cross presentation of soluble antigens, G-protein coupled receptor pathways (GPR) and COVID-19 comorbidities, such as lipid metabolism. We identified three novel loci in X chromosome, one of which overlaps with our previous study and is associated with extra-cellular matrix organization and is likely linked to the pulmonary fibrosis among severe COVID-19 patients. We further identified novel SNVs in chrX including one at ACE2, the receptor of SARS-CoV-2. Further, hospitalized patients revealed significantly larger Central European ancestry ( $p$ -value = 0.03) and highly significantly larger Western European ancestry ( $p$ -value = 0.002) compared to their non-hospitalized counterparts.

*Conclusion:* Overall, our study indicated the likely association between polymorphisms in chrX and the severity of COVID-19 disease presentation and potential crucial role of GPR signaling in severe disease outcomes. Our study further revealed the hospitalized patients have higher Central and Western European ancestry fractions compared to their non-hospitalized counterparts.

## References

1. Upadhyai P. et al. Genomic and Ancestral Variation Underlies the Severity of COVID-19 Clinical Manifestation in Individuals of European Descent. *Life*. 2021;11(9):921.

## New features of the PCA-Seq method in time series analysis (on the example of COVID-19)

Efimov V.M.<sup>1,2,3,5\*</sup>, Polunin D.A.<sup>2</sup>, Kovaleva V.Y.<sup>3</sup>, Efimov K.V.<sup>4</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

<sup>4</sup> *Higher School of Economics, Moscow, Russia*

<sup>5</sup> *Tomsk State University, Tomsk, Russia*

\* *efimov@bionet.nsc.ru*

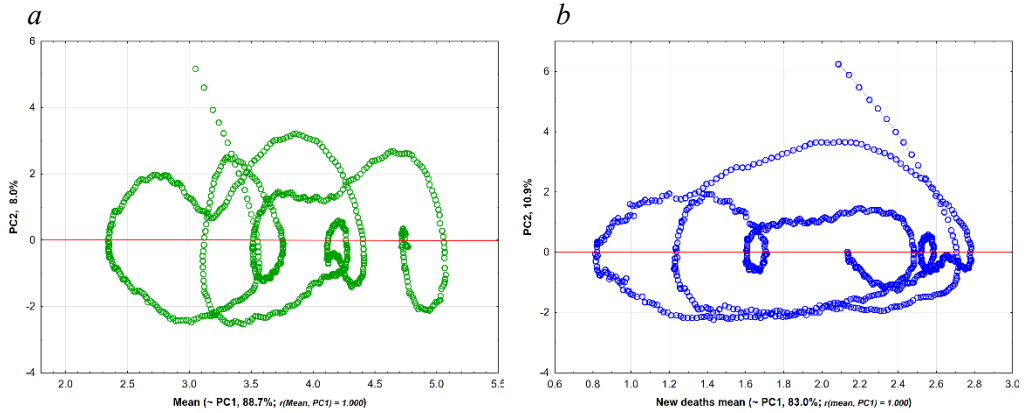
**Key words:** Phase portraits, dynamic system, SSA

*Motivation and Aim:* A feature of the mass reproduction of living organisms (viruses, bacteria, lemmings, rabbits, humans, etc.) is the "explosive" exponential growth of the population at the initial stage. This occurs when population growth is proportional to the abundance already present. At the beginning of the coronavirus pandemic, the number of cases increases exponentially, and the number of deaths, obviously a more important indicator, is a relatively small proportion of cases and is late by about two weeks. And although the causal relationship of these processes is beyond doubt and, it would seem, the second indicator should be easily predicted two weeks in advance, the statistical correlation between them turns out to be rather weak. Quarantine, vaccination and re-infection contribute to the official statistics. There is a need for more powerful methods for analyzing such data. A 1D time series is traditionally represented as the sum of a trend, cyclical components, and noise. However, the typical behavior of Covid-pandemic indicators does not correspond to this perception: an almost imperceptible and difficult to fix beginning; exponential growth; plateau – fluctuations around a certain level; decline, slow or fast. Cyclicity manifests itself in the form of waves of different scale and duration with intervals of different lengths between them. On a logarithmic scale, a deviation from a straight line means that there are obstacles to the reproduction of the virus. The first obstacle is quarantine, the next is vaccination. The PCA-Seq method (a generalization of SSA) is suitable for the analysis of such processes [1]. It made it possible to obtain similar phase portraits of the dynamics of morbidity and mortality from coronavirus in Italy (the first country in which a pandemic began in Europe).

*Methods and Algorithms:* Time series for daily new cases and new deaths from Covid-19 in Italy obtained from the website [https://github.com/CSSEGISandData/COVID-19/blob/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/). After Log transformation, the PCA-Seq method (N = 823; Lag = 50) was applied to them [1, 2]. This method is a generalization of the well-known SSA (singular spectral analysis) [3], which is also used in research on Covid-19 [4]. PCA-Seq, unlike SSA, allows you to do any transformations of trajectory matrix rows and calculate any other Euclidean distances between the rows [5].

*Results:* The figure shows the phase portraits of the first two principal components of the dynamics of new cases of morbidity and mortality. Despite the orthogonality, the second PC is a derivative of the first one (this follows from the loadings of the PC2). This is a well-known property of the principal components of time series. In turn, the correlation coefficient of PC1 with the mean is practically equal to one. This means that when PC2

$> 0$  (above the red line) the average morbidity (or mortality) increases, otherwise it falls. Rotation is always clockwise. Taking into account the inertia of the process, this circumstance can be confidently used for forecasting at a qualitative level if the trajectory is sufficiently far from zero. For example, both morbidity and mortality should increase at the moment, so the trajectory definitely intends to cross the red line. But it is difficult to predict whether it will be a small loop or a large circle.



**Fig. 1.** Phase portraits of first two Covid-19 dynamic principal components: *a)* daily new cases; *b)* daily new deaths. In both cases PC1 dynamic is equivalent Mean one

**Conclusion:** As the figure shows, the behavior of the trajectories is similar, as it should be, since one process is generated by another. On the other hand, this behavior appears to be complex enough to hopefully be described by simple dynamic models. Somewhat reassuringly, something like differential equations emerges as a rather unexpected consequence of straight-forward statistical processing.

**Acknowledgements:** The study is supported by the Kurchatov Genomic Centre of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

### References

1. Efimov V.M., Efimov K.V., Kovaleva V.Y. Principal component analysis and its generalizations for any type of sequence (PCA-Seq). *Vavilov J Genet Breed.* 2019;23(8):1032-1036.
2. Efimov V.M., Polunin D.A., Kovaleva V.Y., Efimov K.V. The PCA-seq method applied to analyze of the dynamics of COVID-19 epidemic indicators. *J Phys Conf Ser.* 2021;1715(1):012025.
3. Golyandina N. Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. *Wiley Interdiscip Rev Comput Statist.* 2020;12(4):e1487.
4. Kalantari M. Forecasting COVID-19 pandemic using optimal singular spectrum analysis. *Chaos Solit Fractals.* 2021;142:110547.
5. Efimov V.M., Efimov K.V., Polunin D.A., Kovaleva V.Y. New possibilities of the PCA-Seq method in the analysis of time series (on the example of solar activity). *J Phys Conf Ser.* 2021;2099(1):012034.

# Interferon system response to virus infection

Gainova I.

*Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

*gajnova@math.nsc.ru*

**Key words:** interferon system, virus infection, human immunodeficiency virus type 1 (HIV-1)

*Motivation and Aim:* The interferon system is a key factor in the reactions of innate immunity, which implements the mechanisms for protecting the body from virus infection. Interferon is activated in response to the entry of viruses into the body and prevents them from replicating in the host cell and further spreading to the neighboring ones. The aim of the work is to study a mathematical model that describes the interaction of two processes: the process of HIV-1 replication in the cell [1, 2] and the response of the interferon system to the penetration of the virus into the cell.

*Methods and Results:* The mathematical model describing the stages of the interferon systems response to the HIV-1 infection in terms of ordinary differential equations has been considered. The system consists of 39 equations representing the following stages of the HIV-1 life cycle: entry, reverse transcription, integration, transcription, translation, assembly, budding and maturation new virions and the process of the interferon response to the HIV-1 penetration into the CD4+ T-cell. Numerical experiments have been done to verify and calibrate the model as well as to analyze the sensitivity of the model.

*Conclusion:* As a result of the numerical analysis of the model sensitivity, the stages of the interferon response and the model parameters that have greatest impact on the efficiency of producing the viral particles having the infected cell have been determined.

*Acknowledgements:* The study was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (FWNF-2022-0015) and with partial financial support by the Russian Foundation for Basic Research (20-01-00352).

## References

1. Shcherbatova O., Grebennikov D., Sazonov I., Meyerhans A., Bocharov G. Modeling of the HIV-1 life cycle in productively infected cells to predict novel therapeutic targets. *Pathogens*. 2020;9(4):255. DOI 10.3390/pathogens9040255.
2. Likhoshvai V.A., Khlebodarova T.M., Bazhan S.I., Gainova I.A., Chereshev V.A., Bocharov G.A. Mathematical model of the Tat-rev regulation of HIV-1 replication in an activated cell predicts the existence of oscillatory dynamics in the synthesis of viral components. *BMC Genomics*. 2014;15(Suppl 12):S1.

## Big Data in creating a clinical decision support system

Gaysler E.<sup>1\*</sup>, Demenkov P.<sup>1,2</sup>, Ivanisenko V.<sup>1,2</sup>, Ivanisenko T.<sup>2</sup>

<sup>1</sup> Zelman Institute of Medicine and Psychology, Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* Evgeniy.gaysler@mail.ru

**Key words:** CDSS, association of symptom with disease, ternary logic

**Motivation and Aim:** CDSS (Clinical decision support system) – “... Link the results of clinical decisions with data available for a particular patient, influencing the choice of a medical decision for more effective medical care ...” (Robert Hayward, Center for Health Evidence). The relevance of the creation of CDSS is due to the general trend in the development of digital medicine, aimed at reducing the number of medical errors and optimizing the work of medical institutions. It is believed that the development and implementation of CDSS belong to the most important directions in the development of artificial intelligence in medicine [1]. There are two main varieties of CDSS - based on “scientific knowledge” and “non-knowledgeable”, based on the results of processing the collected statistical data, most often in the form of a neural network. There is an approach to the formation of reasoning by precedents (retrospective data on individual clinical cases are used) [2].

**Methods and Algorithms:** A hybrid approach is proposed: on the one hand, the knowledge contained in the largest library of peer-reviewed medical journals PubMed is used (28 million articles were used), on the other hand, these data (in our case, the number of mutual mentions of a particular symptom with various diseases) are statistically processed to build an information base (712 symptoms and 584 diseases, syndromes, and disease states). “Knowledge” CDSS use inference mechanisms that take the form of IF-THEN rules. In medical practice, situations often arise (the symptom is not manifested) that do not fit into the usual Boolean logic, so the ternary logic (YES; UNKNOWN; NO) is used in the inference mechanism. The directed selection algorithm is implemented as a question-answer system, and the results of the answers are used to assess the degree of proximity of diseases to the specified set of symptoms (their presence or absence). The solutions at the top of the list of diseases are recombined and the next question is found based on the proximity score. The process can be stopped at any time. The result is a list of diseases with an indication of the likelihood of matching the indicated symptoms. **Results:** Testing in a medical institution showed that the probability of a match between the diagnosis established by experienced doctors and the disease from the first 5 noted by the system is more than 80 %.

**Conclusion:** The chosen approach to the creation of CDSS has shown its potential effectiveness. Extended testing of the system by employees of the Zelman Institute of Medicine and Psychology NSU will clarify the operation of the inference mechanism. Further development of the system involves the creation of a base of precedents (based on the case histories of many patients), which will allow us to confirm or refute the hypotheses of choosing a diagnosis based on the work of the CDSS.

### References

1. Moja L. et al. *Am J Public Health*. 2014;104(12):e12-22.
2. Tanveer Syeda-Mahmood plenary talk: The role of machine learning in clinical decision support. *SPIE Newsroom*. 2015.

## Finding possible inhibitors for SARS-CoV-2 proteins using virtual ligand screening

Gorelov S.V.<sup>1,3\*</sup>, Shvetsov A.V.<sup>1,3</sup>, Egorov V.V.<sup>1,2</sup>, Isaev-Ivanov V.V.<sup>1</sup>

<sup>1</sup> Saint Petersburg Nuclear Physics Institute named by B.P. Konstantinov, NRC "Kurchatov Institute", Gatchina, Russia

<sup>2</sup> Smorodintsev Research Institute of Influenza, St. Petersburg, Russia

<sup>3</sup> St. Petersburg Polytechnic University of Peter the Great, St. Petersburg, Russia

\* Gorelov\_sv@npni.nrcki.ru

**Key words:** nsp, molecular docking, molecular dynamics, virtual screening, structure, energy minimization, active center, box, AutoDock Vina, GROMACS

*Motivation and Aim:* SARS-CoV-2 is an enveloped, single positive-stranded RNA belonging to the genus Betacoronavirus, a virus with large viral RNA genomes that causes COVID-19 disease. It was first registered in 2019. Although SARS-CoV-2 belongs to the group of betacoronaviruses, it differs from MERS-CoV and SARS-CoV. So far, it has already affected more than 210 million people worldwide. Existing virtual screening and computer simulation methods allow for research that can detect new potential inhibitors of the main protease of the SARS virus, thereby identifying substances with possible therapeutic properties against COVID-19. Coronavirus replication cycle begins with translation of ORF1a and ORF1b after penetration and release of the genome. First, two large replicase polyproteins (pp1a and pp1ab) are synthesized. They are then cleaved into 16 non-structural proteins (nsps), including nsp12, viral RNA-dependent RNA polymerase (RdRp), papain-like protease (nsp3), and 3C-like protease (nsp5). Viral unstructured proteins engage the host cell's membrane structure to assemble into replication and transcription complexes that are involved in the synthesis of the RNA minus strand. Then subgenomic RNAs are synthesized and structural and accessory proteins are expressed. The pathogenesis of severe acute respiratory syndrome coronavirus SARS-CoV-2 is an important issue for the treatment and prevention of SARS, and the current state of the problem under study allows finding new potential inhibitors [1, 2]. *Methods and Algorithms:* This work is devoted to the virtual screening of ligands for Sars-CoV-2 proteins, in particular, for unstructured proteins nsp5 and nsp12. For this, the methods of molecular docking and molecular dynamics were used. Molecular docking was calculated using the AutoDock Vina program, while the GROMACS universal software package was used for molecular dynamics and energy minimization. Substances from the ZINC15 database in the amount of 5901 units served as ligands for molecular docking. *Results:* Using the methods of molecular docking and molecular modeling, a virtual screening of replicase ligands of SARS-CoV-2 polyprotein 1ab was performed. Carrying out test runs showed that the AutoDock Vina software for molecular docking and virtual screening really has high accuracy and speed of calculations. Software algorithms were created for visualizing search areas and carrying out stream docking with subsequent sorting of the results obtained, which can be used in further work. Structured lists of substances have been obtained that can be used for future calculations of binding constants in solution with recombinant proteins. Since the ligands from the ZINC15 database, with small

differences in structure, had an insignificant spread in affinity values, it can be assumed that docking was applied correctly. Using the method of molecular dynamics, the possible binding of the nsp7 protein to the NILYRSLAETR peptide sequence was analyzed. The natural nsp7 protein was found to have at least 6 potential binding sites for the putative ligand. The nearest amino acid residues that are supposedly involved in intermolecular interactions have been characterized. The process of attaching a peptide sequence to nsp7 was also visualized. It was found that NILYRSLAETRs are indeed able to form fibril-like structures among themselves. Conclusion: A workflow algorithm was developed for calculating molecular docking for any number of ligands with almost any receptor protein, followed by structuring the output data. A similar algorithm was used to calculate molecular docking for structured proteins nsp5 and nsp12 in different regions. For each ligand, 20 stacking conformations with their corresponding affinity values were obtained at the output. Of these, the 3 highest affinity values were selected and, having calculated their arithmetic mean, sorted in descending order with their corresponding ligands. This made it possible to create a list of the 25 most optimal ligands for each receptor. Using the data obtained from the molecular dynamics procedure using the GROMACS universal modeling software package, it was possible to visualize the trajectory of 10 molecules of fragments of the sequence of feline, canine and porcine coronaviruses and the unstructured nsp7 protein. The trajectory of protein movement was analyzed, due to which some potential binding sites of nsp7 with molecules of peptide sequences NILYRSLAETR in the amount of 6 pieces were identified.

### *References*

1. Wang W., Zhou Z., Xiao X. et al. *Cell Mol Immunol*. 2021 Apr;18(4):945-953.
2. Mariano G., Farthing R.J., Lale-Farjat S.L.M., Bergeron J.R.C. *Front Mol Biosci*. 2020 Dec;17(7):605236.

## Features of the interaction of human piRNAs with the SARS-CoV-2 coronavirus genome

Ivashchenko A.T.<sup>1\*</sup>, Pyrkova A.Yu.<sup>1,2</sup>, Akimniyazova A.N.<sup>2</sup>, Rakhmetullina A.K.<sup>2</sup>, Niyazova T.K.<sup>2</sup>, Yurikova O.Yu.<sup>2</sup>

<sup>1</sup> Center for Bioinformatics and Nanomedicine, Almaty, Kazakhstan

<sup>2</sup> al-Farabi Kazakh National University, Almaty, Kazakhstan

\* a.iavashchenko@gmail.com

**Key words:** piRNA, gRNA, SARS-CoV-2, genome, biomedicine, bioinformatics

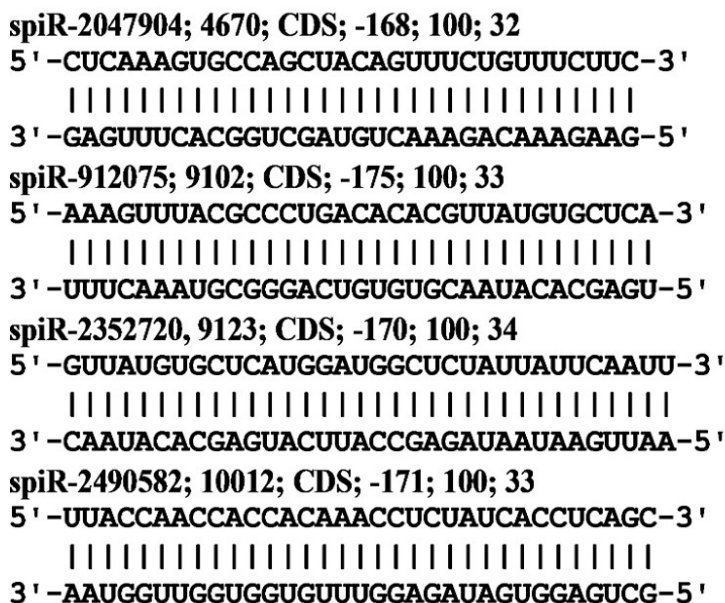
*Motivation and Aim:* The prolonged COVID-19 pandemic, with numerous human casualties, requires a rapid search for remedies against various strains of SARS-CoV-2. We investigated the interaction of over eight million molecules of piRNAs (PIWI-interacting RNA) with the genomic RNA (gRNA) SARS-CoV-2 (NC\_045512.2). Completely complementary binding of piRNAs to mRNAs of several human protein-coding genes has been shown, indicating the biological function of piRNAs as regulators of gene expression.

*Methods and Algorithms:* The nucleotide (nt) sequences of SARS-CoV-2 genome was taken from NCBI. The nucleotide sequences of 200000 piRNAs were taken from Wang et al [1]. The piRNA binding sites (BSs) in the mRNAs of genes were predicted using the MirTarget program [2]. This program defines the following features of piRNA binding to mRNA: a) the initiation of piRNA binding from the first nucleotide of the mRNAs; c) the schemes of nucleotide interactions between piRNAs and mRNA; d) the free energy ( $\Delta G$ , kJ/mol) of the interaction between piRNA and the mRNA; e) the ratio  $\Delta G/\Delta G_m$  (%) is determined for each site ( $\Delta G_m$  equals the free energy of piRNA binding with its fully complementary nucleotide sequence).

*Results:* More than 200 piRNAs can inhibit protein synthesis and simultaneously inhibit viral gRNA replication. Regions of gRNA with overlapping nucleotide binding sites (BSs) of many piRNAs, which we called clusters, were revealed. A 28 nt cluster of 70 piRNAs BSs was identified in the SARS-CoV-2 gRNA region encoding the LETIQITIS oligopeptide of the ORF1ab protein. The second cluster of 39 piRNAs BSs of 31 nt lengths in the gRNA region encodes the RATLQAIASEF oligopeptide of the ORF1ab protein. The third cluster of 24 piRNA BSs of 31 nt length encodes the oligopeptide PKLQSSQAWQP of the ORF1ab protein. Twelve piRNAs were identified that strongly interact with the gRNA single sites of the virus from 4670 to 29024 nt. These endogenous piRNAs can be used in higher concentrations to suppress coronavirus multiplication. Synthetic piRNAs (spiRNAs) based on these endogenous piRNAs can be used to rapidly suppress coronavirus multiplication by replacing non-canonical nucleotide pairs with canonical pairs.



## The schemes of spiRNAs interaction with CDS gRNA SARS-CoV-2 from 4670 nt to 29475 nt



The spiRNAs will inhibit the coronavirus multiplication even more strongly. These spiRNAs and selected endogenous piRNAs have little effect on human 17494 protein-coding genes, indicating a low probability of side effects. We assume that endogenous piRNAs can protect humans from mass infection with coronaviruses. The spiRNA and piRNA kits are proposed to suppress SARS-CoV-2 strains. The piRNA and spiRNA selection methodology created for the control of SARS-CoV-2 can be used to control all strains of this coronavirus.

*Acknowledgements:* The study is supported by Center for Bioinformatics and Nanomedicine, Almaty, Kazakhstan.

### References

1. Wang J., et al. piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.* 2019;47:175-180. doi: 10.1093/nar/gky1043.
2. Ivashchenko A. et al. MiR-3960 Binding Sites with mRNA of Human Genes. *Bioinformatics.* 2014;10:423-427. doi:10.6026/97320630010423.

## Stationary and non-stationary mutagenesis in the SARS-CoV-2 genome

Kobalo N.<sup>1\*</sup>, Kulikov A.<sup>1</sup>, Titov I.<sup>2</sup>

<sup>1</sup> *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *rerf2010rerf@yandex.ru*

**Key words:** *SARS-CoV-2, stationarity, mutations*

The pandemic of the coronavirus infection COVID-19, which began at the end of 2019 and caused by the SARS-CoV-2 virus, has led to unprecedented consequences in the world. By the end of March 2022, in the world there were 450 million infected and 6 million died directly from infection [1]. SARS-CoV-2 is known to be a betacoronavirus, so it shares many properties with other viruses in this family, including mutation patterns and conserved fragments [2]. Since the beginning of the spread of COVID-19, the issues of research the SARS-CoV-2 virus have been to discover its evolutionary variability, active points or standard mutation patterns, including those common with other betacoronaviruses.

This report examines the stationarity of mutations present in the SARS-CoV-2 genome, comparing both located in conserved RNA fragments of the genome and in coding regions. For this, a phylogenetic tree provided by the UShER project was used [3]. Hot spots of genome mutations, stationary and non-stationary mutations, their relationship with conservative and other functional fragments of the genome were characterized.

*Acknowledgements:* The work of IT was supported by the budget project No. FWNR-2022-0020.

### *References*

1. Roser M., Ritchie H., Ortiz-Ospina E., Hasell J. Coronavirus Pandemic (COVID-19). 2020. Published online at
2. Rangan R., Zheludev I.N., Hagey R.J. et al. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA*. 2020;26(8):937-959. doi: 10.1261/rna.076141.120.
3. Turakhia Y., Thornlow B., Hinrichs A.S. et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat Genet* 2021;53:809-816. doi: 10.1038/s41588-021-00862-7.

# Mathematical modeling of SARS-CoV-2 infection process and virus spreading in the human body considering B and T cell-mediated immune responses

Miroshnichenko M.I.<sup>1,2\*</sup>, Akberdin I.R.<sup>1,2,3</sup>, Afonyushkin V.N.<sup>1,4,5</sup>, Kolpakov F.A.<sup>2,3,6</sup>

<sup>1</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>2</sup> *BIOSOFT.RU, LLC, Novosibirsk, Russia*

<sup>3</sup> *Sirius University, Sochi, Russia*

<sup>4</sup> *Siberian Federal Scientific Center of Agro-BioTechnologies, RAS, Novosibirsk, Russia*

<sup>5</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

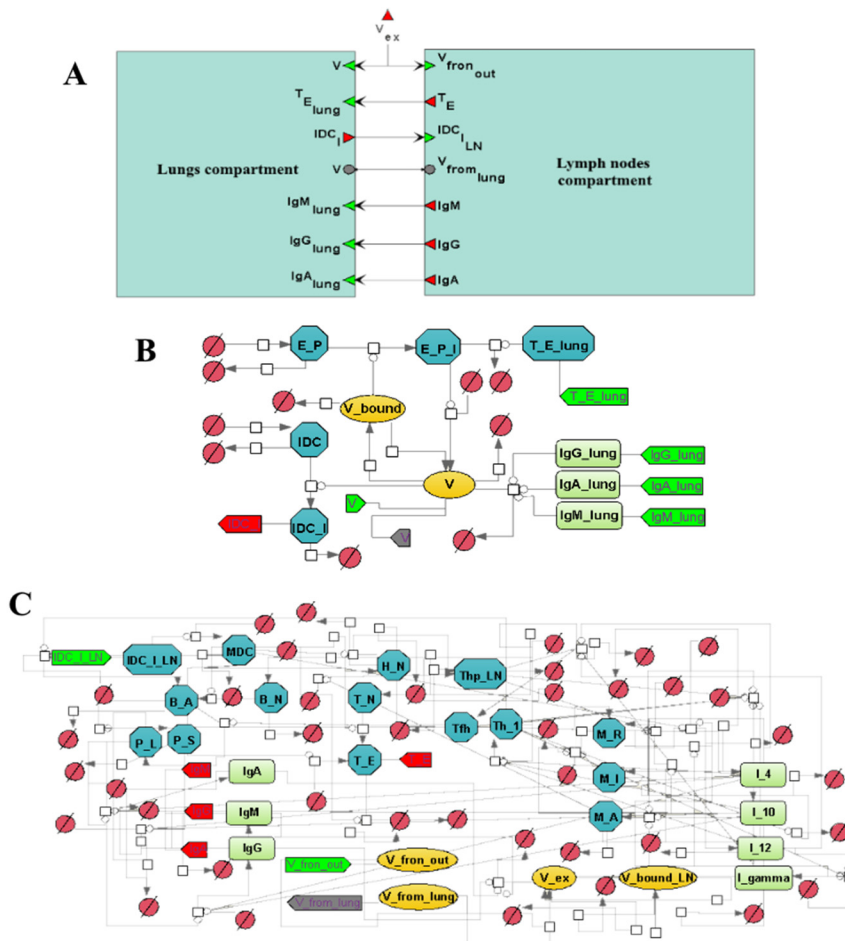
<sup>6</sup> *FRC for Information and Computing Technologies, Novosibirsk, Russia*

\* *m.miroshnichenko2@g.nsu.ru*

**Key words:** coronavirus, SARS-CoV-2, COVID-19, mathematical model, BioUML

In spite of the progress reached in understanding of COVID-19, prediction's fidelity of the development and outcome of the disease still remain deficient. To better understand the progression of COVID-19 patients' clinical picture, molecular mechanisms of interactions between human cells and virus particles taking into account B and T cell-mediated immune responses, a construction of multi-compartmental model considering the specificity of the virus interplay with host cells, its entry and replication, transport throughout tissues and organs, along with formation and development of local and global immune responses to the viral infection, is pivotally required.

We have developed the model, which is based on previously published and tested models of human immune response to Mycobacterium tuberculosis [1] and adaptive B and T cell-mediated immune responses to influenza A virus [2] using a modular modelling approach implemented in open-source software platform BioUML [3]. It incorporates two compartments: lungs and lymph nodes and includes functioning and dynamic of virus, macrophages (resting, activated, infected), dendritic cells (immature and mature), cytotoxic and helper T cells (naïve, Th precursors, Th1 and Tfh), B cells and antibodies (IgA, IgM, IgG), cytokines (IL-4, IL-10, IL-12, IFN- $\gamma$ ) and epithelial cells (infected and uninfected). The model was fitted to the published data on SARS-CoV-2 and it is able to reproduce typical disease progression scenarios, dynamics of T and B cells, and levels of both antibodies and cytokines production as well as the viral load (Fig. 1).



**Fig. 1.** Model diagrams in BioUML **A)** general composite model; **B)** lungs compartment; **C)** lymph nodes (LN) compartment. E\_P – epithelial cells, E\_P\_I – infected epithelial cells, IDC – immature dendritic cells, MDC – mature dendritic cells, V – virions, Ig – immunoglobulins (G, A, M), B\_N – naïve B cells, B\_A – activated B cells, P\_L – long-lived plasma cells, P\_S – short-lived plasma cells, T\_N – naïve CD8+ T cells, T\_E – CD8+ T cells, H\_N – naïve CD4+ T cells, Thp – precursor helper T cells, Th\_1 – type 1 helper T cells, Tfh - follicular helper T cells, M\_R – resting macrophages, M\_I – infected macrophages, M\_A – activated macrophages, I – interleukins (4, 10, 12), I\_gamma – interferon gamma

*Acknowledgements:* This work was supported by the RFBR project No. 20-04-60355.

### References

1. Marino S., Kirschner D.E. *J Theor Biol.* 2004;227(4):463-486.
2. Lee H.Y. et al. *J Virol.* 2009;83(14):7151-7165.
3. Kolpakov F. et al. *Nucleic Acids Res.* 2019;47(1):225-233.

# The potential usability of Oxford Nanopore Technology for revealing of SARS-CoV-2 dual infection and intra-host viral variability in clinical samples

Penkin L.N.\*, Korneenko E.V., Samoilov A.E., Manolov A.I., Speranskaya A.S.

*Research Institute for Systems Biology and Medicine of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Moscow, Russia*

\* [leopold.valerjanovitch@yandex.ru](mailto:leopold.valerjanovitch@yandex.ru)

**Key words:** SARS-CoV-2, Oxford Nanopore, heterogeneity, intra-host variability, double infection

*Motivation and Aim:* Due to the emergence of the epidemic of COVID-19 caused by SARS-CoV-2 it has become necessary to monitor the spread of different variants of the virus. Since its appearance, the virus has continued to evolve, giving rise to many variants with increased transmissibility and ability to avoid the immune system. The patient's organism presents the reservoir for the new strains due to ongoing viral evolution. In addition, the facts of Covid-19 double infection (the case when two strains are simultaneously detected in the patient [1]) and their impact on the course of the disease has not been sufficiently studied) has been described repeatedly. Therefore, it is important not only to be able to identify the strains of SARS-CoV-2, but also to identify the variability in samples and determine its nature as well. The most informative method of determining SARS-CoV-2 variants is high-throughput sequencing. Currently, a large amount of data of the SARS-CoV-2 virus genome has been obtained using Oxford Nanopore Technology (ONT). But, according to the public opinion, this technology produces a lot of mistakes. Therefore, this fact raises the question, if it's possible to distinguish the experimental artifacts from the true intra-host viral variability. True heterogeneity refers to positions that result from the presence of several virus variants in the sample, but are not the result of contamination or errors produced by polymerases, reverse transcriptases or sequencing errors.

The aim of our study is to estimate the usability of Oxford Nanopore data for detection of intra-host variability and double infection of SARS-CoV-2.

*Methods and Algorithms:* Samples for sequencing (oropharyngeal swabs) were obtained from 32 regions of the Russian Federation from patients positive for SARS-Cov-2 and collected from January 19, 2022 to March 30, 2022. RNA was extracted by *NA magnet SARS-CoV-2 (LYTECH)*, the whole genome sequencing was performed by the Oxford Nanopore platform (PromethION, GridION) using Midnight RT PCR Expansion Kit (EXP-MRT001, ONT) and Rapid Barcoding kit 96 (SQK-RBK-110-96, ONT). Consensus sequences were obtained using *epi2me-labs/wf-artic* utilities. Detection of the virus lineage was performed using the Pangolin algorithm. Variable positions were detected using Longshot software - is a variant calling tool for diploid genomes using long error prone reads such as Pacific Biosciences (PacBio) SMRT and Oxford Nanopore Technologies (ONT).

*Results:* The whole genome sequencing and detection of viral strain were performed for 6916 samples. The distribution of variants was as follows: Omicron subvariant BA.1 – 5310, Omicron subvariant BA.2 – 1028, Delta – 280, None – 257. The search for SNV in the samples was performed in several stages. At first, we used the *epi2me-labs/wf-*

artic algorithm with default settings (sample coverage normalization threshold = 200). As a result, the usage of the default settings revealed an artificial distribution of the frequency of occurrence of SNV in the samples (anomalous "clustering"). By increasing the maximum coverage threshold to 5000, this artifact was eliminated. Then, we identified samples in which reliable SNVs were present. The positions in which the minor variant was represented by at least in 35–50 % of the reads were taken as reliably defined SNVs. As a result, in 3,366 samples at least one valid SNV was detected. At the same time, 95 % of these samples contained 1 to 10 SNV (N = 3183). In the remaining 5 % (N = 35) of samples, between 11 and 153 SNV were found and these samples were further analyzed to determine if the nature and position of the SNV could be explained by the fact that these samples could potentially contain two different strains of the virus. The samples in our study relate to the period with the following distribution of SARS-CoV-2 variants: B.1.1.617.2 (Delta), BA.1 (Omicron), BA.2 (stealth-Omicron). Therefore, a list of mutations in the S-gene for Delta/BA.1, Delta/BA.2, and BA.1/BA.2 variants was constructed. In 4 samples, the degeneracy and SNV position corresponded to a mixture of Delta and Omicron strains. Additional analysis suggests that this variability is not the result of intra-laboratory contamination. Probably, the patients from whom these samples were obtained were simultaneously infected with two strains of SARS-CoV-2.

*Conclusion:* Our analyses revealed that 48.6 % of samples contained at least one SNV with low frequency variants in 35–50 % of allele frequency. We suppose the Oxford Nanopore sequencing data could be used for investigation of intra-host variation and double infection of SARS-CoV-2 at least for samples with high viral load of both strains.

*Acknowledgements:* This work was supported by Rospotrebnadzor service state task No. 122030900051-9.

### *References*

1. Samoilov A.E. et al. Case report: change of dominant strain during dual SARS-CoV-2 infection. *BMC Infect Dis.* 2021;21(1):959.

# Numerical treatment for Volterra integro-differential equations with time lag using $\mathfrak{S}$ -method

Rihan F.A., Mohamed O.

*Department of Mathematical Sciences, College of Science, UAE University, Al-Ain, UAE  
frihan@uaeu.ac.ae, 700039076@uaeu.ac.ae*

**Key words:**  $\mathfrak{S}$ -method, Volterra delay integro-differential equation, stability, time-lag

*Abstract:* Volterra delay integro-differential equations are indispensable tools in modeling a variety of biological and physical phenomena. The  $\mathfrak{S}$ -method is extended to solve Volterra delay integro-differential equations. We discuss the efficiency and stability of the method. The efficacy of the proposed method is illustrated by constructing some numerical examples.

## *References*

1. Rihan F.A., Delay Differential Equations and Applications to Biology. Springer, 2021.
2. Rihan F.A., Doha E.H., Hassan M.I., Kamel N.M. Numerical treatments for Volterra delay integro-differential equations. *Comput Methods Appl Math.* 2009;9(3):292-318.

# Large scale clustering in structural and evolutionary analysis of SARS-CoV-2 proteins

Mitić N.<sup>1\*</sup>, Pavlović-Lažetić G.<sup>1</sup>, Beljanski M.<sup>2</sup>, Malkov S.<sup>1</sup>, Maljković M.<sup>1</sup>,  
Stojanović B.<sup>1</sup>, Veljković A.<sup>1</sup>, Kapunac S.<sup>1</sup>

<sup>1</sup>Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

<sup>2</sup>Institute of General and Physical Chemistry, BioLab, Belgrade, Serbia

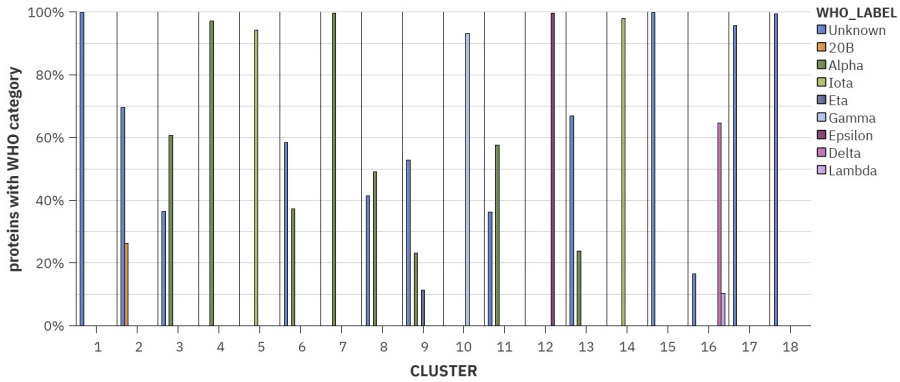
\*nenad@matf.bg.ac.rs

**Key words:** SARS-CoV-2, Relative Synonymous Codon Usage, proteins, WHO Annotation, clustering

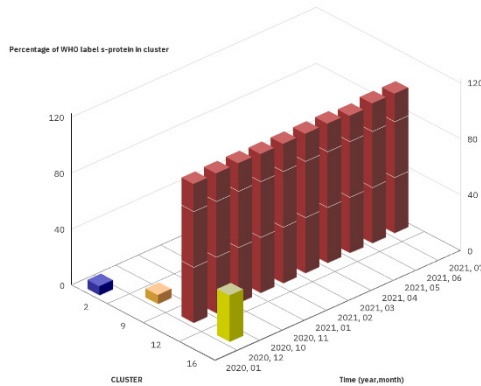
*Motivation and Aim:* In order to understand SARS-CoV-2 origin, evolution and interaction with host's cells, various aspects of viral genome structure and function are under investigation. Codon usage (CU) frequency of viral proteins as well as non-silent mutations are of special interest, since they may contribute to changing virus characteristics. Previous analyses have shown that rare codons often occur in large clusters within protein coding sequences. In the case of SARS-CoV-2, previous codon usage analyses show an antagonistic codon usage pattern (i.e., use of rare codons) reducing translation speed, but increasing its precision, and yielding accurate and correctly folded viral proteins [1]. At that end, clustering of protein sequences are investigated based on Relative Synonymous Codon Usage (RSCU) as well as edit distances of amino acid (AAc) sequences providing for both characterizing (identifying) specific protein groups (types) and temporal evolution of proteins within groups (types). The paper will present results of such analysis.

*Materials, Methods and Algorithms:* A dataset of 423425 complete isolate nucleotide sequences have been extracted from <https://www.ncbi.nlm.nih.gov/sars-cov-2> on August 25, 2021. After cleaning process, remains 347962 isolates with 225.934 unique (2.366.031 total) SARS-CoV-2 protein coding nucleotide sequences, as well as the corresponding AAc sequences. Consistency check has been performed between the two based on standard genetic code table (transl\_table 1). For all the proteins (141926) for which world-health-organization (WHO) SARS-CoV-2 annotation exists, submission date and protein sequence metadata are supplied and RSCU has been calculated for measuring CU bias in different proteins of different protein classes. Then different algorithms (including TwoStep clustering in SPSS Modeler program [2], hierarchical clustering in Cluto [3] and Python Scikit-learn library) were applied for k-clustering proteins based on RSCU, for  $k = 2,40$ . Proteins of the most heterogeneous protein type – Surface glycoprotein (S-protein)– have been further clustered based on RSCU, for each WHO label and year/month date. Furthermore, for all the protein sequences, edit distances of AAc sequences for each pair of proteins have been calculated. Then different algorithms (e.g., spectral clustering) were applied for clustering proteins in each protein group. *Results:* Fig. 1. presents S-protein 18-clustering based on RSCU and labeled by WHO annotations. Clustering results are quite correct with silhouette of 0.47. The figure presents all the groups with quantities higher 10 %. Clustering performed by the SPSS Modeler program. Fig. 2 (for Epsilon WHO label) is representative of a set of figures presenting specific WHO groups on year/month scale when S proteins are clustered by the SPSS modeler into 18 clusters.





**Fig. 1.** S-protein 18-clustering based on RSCU and labeled by WHO annotations



**Fig. 2.** Epsilon WHO group on year/month scale when S proteins are clustered by the SPSS modeler into 18 clusters

Fig. 2 is a representative of a set of figures presenting all the groups with quantities higher 5 %. Specific WHO groups mostly dominate in specific clusters in all the time periods (for example, as is the case with the Epsilon group dominating in the cluster 12). Spectral clustering of different types of proteins, based on AAc distances, give quite similar results regarding WHO labels, when applied to S protein (clusters rather homogenous), while less homogenous but still representative for other types of proteins. Hierarchical clustering of all the proteins for  $k = 2,40$  produces highly homogenous clusters regarding protein types. Specifically, for  $k = 12$  (the number of different protein types), each type is predominately represented by its specific cluster.

*Conclusion:* Since all the SARS-CoV-2 Orfs cluster in relatively homogenous clusters (according to WHO isolate classification), i.e., WHO-specifically annotated isolates make most of each cluster, this new approach may be used for annotation/prediction of strains that isolates belong.

### References

1. Mogro E.G., Bottero D., Lozano M.J. Analysis of SARS-CoV-2 synonymous codon usage evolution throughout the COVID-19 pandemic. *Virology*. 2022;568:56-71.
2. IBM SPSS Modeler 18.2 Algorithms Guide, <https://www.ibm.com/support/pages/spss-modeler-182-documentation>
3. CLUTO – Software for Clustering High-Dim. Datasets: <http://glaros.dtc.umn.edu/gkhome/views/cluto>

# Revealing the molecular basis of interactions of COVID-19 with hyperglycemia and diabetic complications based on the bioinformatics analysis of the gene networks

Saik O. \*, Klimontov V.

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

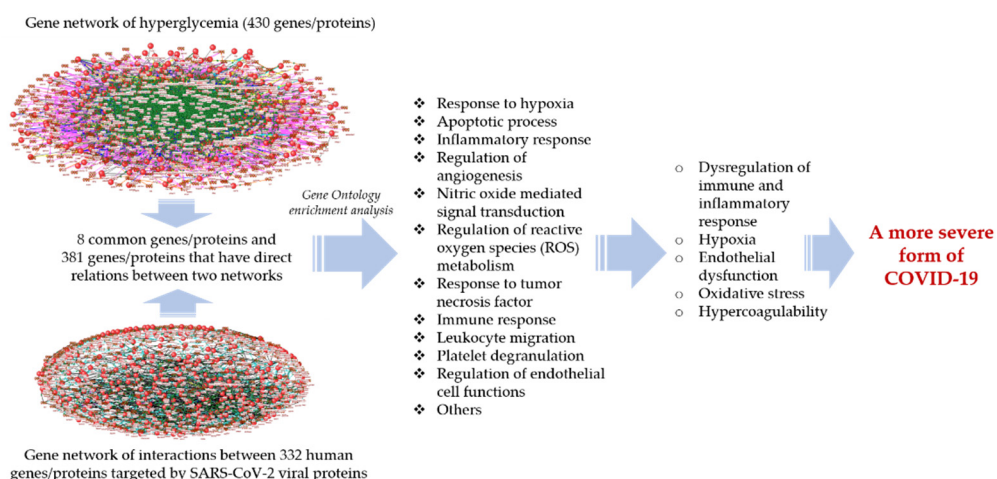
\* saik@bionet.nsc.ru

**Key words:** COVID-19, SARS-CoV-2, diabetes, hyperglycemia, gene networks, ANDSytem

**Motivation and Aim:** People with diabetes are more likely to have severe COVID-19 compared to general population. Moreover, diabetes and COVID-19 demonstrate a certain parallelism in the mechanisms and organ damage [1]. We applied bioinformatics analysis of associative molecular networks to identify key molecules and pathophysiological processes that determine SARS-CoV-2-induced disorders in patients with diabetes.

**Methods:** The text-mining based approaches of ANDSytem [2, 3] were used in this study for reconstruction and analysis of gene networks. The gene network associated with hyperglycemia was reconstructed in the study [4]. The list of 332 SARS-CoV-2-targeted proteins was extracted from the publication [5]. The Gene Ontology enrichment analysis was performed by DAVID online tool [6]. The networks of vascular diabetic complications were obtained from works [4, 7].

**Results:** Firstly, we have reconstructed and matched the network associated with hyperglycemia and the network of SARS-CoV-2-targeted proteins. These networks included 430/332 genes/proteins and 46855/1664 interactions respectively (Figure).



**Fig. 1.** Interconnections between the hyperglycemia gene network and the network of SARS-CoV-2-targeted proteins are involved in a number of biological processes which disruption leads to some pathological states that contribute to the development of a more severe form of COVID-19 in patients with diabetes

Eight hyperglycemia-related genes/proteins (DNMT1, FBN1, GDF15, GPX1, HMOX1, IDE, PLAT, and RHOA) were also targeted by SARS-CoV-2. Besides, 381 molecules were directly connected between two networks by protein-protein or regulatory connections. These molecules, according to Gene Ontology enrichment analysis, are involved in response to hypoxia, reactive oxygen species metabolism, immune and inflammatory response, regulation of angiogenesis, platelet degranulation, and other processes (Figure).

A number of genes/proteins targeted by SARS-CoV-2 (BRD2, COMT, DNMT1, ERP44, F2RL1, GDF15, GPX1, HDAC2, HMOX1, HYOU1, IDE, LOX, NUTF2, PCNT, PLAT, RAB10, RHOA, SCARB1, and SELENOS) were found in the networks of insulin resistance and vascular diabetic complications.

*Conclusions:* The results expand the understanding of the molecular basis of diabetes and COVID-19 comorbidity.

*Acknowledgements:* This work was supported by the grant of Russian Science Foundation (20-15-00057).

### References

1. Hartmann-Boyce J. et al. Risks of and from SARS-CoV-2 infection and COVID-19 in people with diabetes: a systematic review of reviews. *Diabetes Care*. 2021;44(12):2790-811.
2. Ivanisenko V.A. et al. ANDSsystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol*. 2015;9(2):1.
3. Ivanisenko V.A. et al. A new version of the ANDSsystem 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(1):5-15.
4. 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.
5. Gordon D.E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020;583(7816):459-468.
6. Huang D.W. et al. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols*. 2009;4(1):44-57.
7. 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.

# The characterization of complete genome of MERS-related coronavirus from bat (*Pipistrellus nathusii*) from Moscow region, prediction and analysis their Spike glycoprotein with DPP4 receptors of the different mammalian species

Speranskaya A.S.<sup>1,2\*</sup>, Samoilov A.E.<sup>1</sup>, Korneenko E.V.<sup>1</sup>, Khabudaev K.V.<sup>1</sup>, Artyushin I.V.<sup>2</sup>, Yusefovich A.P.<sup>2</sup>, Dolgova A.S.<sup>3</sup>, Dedkov V.G.<sup>3</sup>

<sup>1</sup> *Research Institute for Systems Biology and Medicine of the on Consumers' Rights Protection and Human Well-Being Surveillance, Moscow, Russia*

<sup>2</sup> *Lomonosov Moscow State University, Biological Department, Moscow, Russia*

<sup>3</sup> *Saint-Petersburg Pasteur Institute of the Federal Service on Consumers' Rights Protection and Human Well-Being Surveillance, St. Petersburg, Russia*

\* *hanna.s.939@gmail.com*

**Key words:** Bat-CoV, MERS-related viruses, *Pipistrellus nathusii*, bats, hedgehogs, humans, camels, DPP4, Spike

*Motivation and Aim:* The majority of infectious diseases that were discovered during the last few decades are actually zoonosis. The ongoing pandemic disease COVID-19 is an infection caused by the severe acute respiratory syndrome coronavirus SARS-CoV-2. The animal reservoir of SARS-CoV-2 is unknown, but actively discussed as originating from bats, due to reports of SARS-CoV-2-related viruses in Asian Rhinolophus [1]. Previous severe human outbreak caused by coronaviruses was middle East respiratory syndrome (MERS). This disease has a high fatality rate of up to 35 %. MERS was first identified in Saudi Arabia in 2012. Presently 27 countries have reported cases; the largest outbreaks were in Saudi Arabia, United Arab Emirates and the Republic of Korea. The origin of the MERS virus is not understood yet but supposedly it naturally circulated in bats and passed to humans through dromedary camels. Analysis of the virome of bats that are distributed in different geographical regions is an actual approach for identifying the new species of viruses that potentially cause human infection. Currently, a large number of MERS-related coronaviruses from bats of China, Asia, Near East, some countries of Europe and Africa are described. But on the territory of the Russian Federation the MERS-related coronaviruses of bats were not found earlier. We have been screening bat colonies in the Central Region of Russian Federation during the last few years. Some time ago we found the novel MERS-related betacoronavirus in the whild bat *Pipistrellus nathusii*, which was caught near Moscow in 2015. We sequenced and assembled the complete genome of this novel virus. The aim of this work – to better understand the viral adaptability to host as well as various other potential hosts and to estimate the possibility of viral transmission to mammalian species which could contact with humans using computer analysis methods like phylogenetic and phylogeographic analysis, comparative nucleic acid and protein motif analysis and protein-protein docking.

*Methods and Algorithms:* The whole genome sequencing of novel MERS-related Betacoronavirus from *Pipistrellus nathusii* was performed using Illumina sequencing. The complete genome was assembled. We described its genome organization features,

relationship with known coronaviruses, estimated the possibility of viral-host coevolution. Genome annotation was performed by Geneious 7.1.9 and manually edited. For phylogenetic analysis, the independent sampling of available GenBank records for complete genomes, S-, N- and RdRp-encoding sequences was performed, using “Merbecovirus” keywords as the primary filter. The alignments were constructed using MAFFT v7 and phylogenetic analyses were performed using online W-IQ-TREE (ModelFinder + tree reconstruction + ultrafast bootstrap of 1000 replicates). The three-dimensional structure of the S-protein of novel merbecovirus as well as DPP4 receptors for *Myotis brandtii* (EPQ03439.1), *Pipistrellus kuhlii* (KAF6353216.1), *Erinaceus europaeus* (XP\_016043930.1), *Felis catus* (NP\_001009838.1) and *Mus musculus* (NP\_034204.1) was modeled on the SWISS-MODEL server. The RBD 360-610 a.a. of Spike-protein and all DPP4 sites of the studied organisms were used to calculate the docking using the HDock server and analyzed in PyMol.

**Results:** The novel virus was named MOW-BatCoV strain 15-22 and deposited in the GenBank database under accession numbers ON325306. The genome organization of MOW-BatCoV/15-22 virus was found identical to that of other known MERS-related CoV viruses and encompassing the open reading frames (ORFs) in the following order: ORF1ab-spike(S)-ORF3-ORF4a-ORF4b-ORF5-envelope(E)-membrane(M)-nucleocapsid (N)-ORF8b. The trees constructed using complete genomes, N- and RdRp-sequences show the three distinct phylogenetic clades: I - consists of viruses from hedgehogs, II - viruses from the bat only, III - viruses of the bat, humans and camels. The novel MOW-BatCoV/15-22 falls into Clade III and formed a distinct subclade closely related to human/camel’s viruses together with MERS-related coronaviruses of bats *Hypsugo savii* and *Pipistrellus khuli* (from Italy) and from *Neoromicia capensis* (from South Africa). Unexpectedly, the results of a phylogenetic analysis of the Spike-genes showed differing results, namely: the closest similarity of two viruses from bats - MOW-BatCoV 15-22 and *Neoromicia/5038* - with coronaviruses from *Erinaceus* (hedgehogs) was determined. Using the computer molecular docking we analyzed the binding of Spike glycoprotein vs DPP4 receptors of different mammals. The highest binding was predicted in the interaction of MOW-BatCoV Spike-protein and DPP4 of the bat *M. brandtii* (docking score -320.15). Among the other organisms, the highest binding was predicted for the hedgehog, *E. europaeus* (docking score -294.51). The docking results were consistent with the results of the phylogenetic analysis performed. The recent investigation showed MERS-related coronaviruses in hedgehogs from some countries of Europe as well as in China, suggesting that hedgehogs may represent a wild reservoir of novel betacoronavirus, subgenus Merbecovirus in wild hedgehogs [2]. Hedgehogs are animals with cute faces and rising popularity as pets. Besides, humans contact them in the wild. We suggest that hedgehogs may be potential intermediate hosts for viruses between bats and humans.

**Acknowledgements:** The zoology, molecular virology and bioinformatician works were supported by Grant RFBR No. 20-04-60561. The molecular docking works were supported by the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing state task No. 122030900051-9.

## References

1. Temmam S., Vongphayloth K., Baquero E. et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature*. 2022;604:330-336. doi: 10.1038/s41586-022-04532-4.
2. Pomorska-Mól M., Ruzkowski J.J., Gogulski M. et al. First detection of Hedgehog coronavirus 1 in Poland. *Sci Rep*. 2022;12:2386. doi: 10.1038/s41598-022-06432-z.

## 9.2 Section “Mathematical epidemiology”



# Dynamics of stochastic delay differential models for COVID-19: case study in the UAE

Alsakaji H.J. \*, Rihan F.A.

United Arab Emirates University, Al-Ain, UAE

\* heba.sakaji@uaeu.ac.ae

**Key words:** Coronavirus, mathematical modelling, stochastic threshold, stationary distribution, time delays

*Motivation and Aim:* Public health science is increasingly focusing on understanding how COVID-19 spreads among humans. For the dynamics of COVID-19, we propose stochastic epidemic models, with time delays; in the first model we consider Susceptible-Infected-Asymptomatic-Quarantined- Recovered (SIAQR) [1, 3]. While in the second model we extend SEIR epidemic model with vaccination. For both models the stationary ergodic distribution of positive solution is examined, in which the solution fluctuates around the equilibrium of the deterministic case, causing the disease to persist stochastically. Theoretically, it is possible to attain infection free status (extinction) in some situations, in which diseases die out and with a probability of one [2, 4]. The numerical simulations and fit to real observations prove the effectiveness of the theoretical results. Combining stochastic perturbations with time delays enhances the dynamics of the model, and white noise intensity is an important part of the treatment of infectious diseases [5].

*Results:* COVID-19 can be reduced in severity, by prior vaccination (immunization) or the timely use of specific antiviral agents. Vaccination is then operated by reducing the pool of susceptible individuals, and when this is reduced sufficiently, an infectious disease cannot spread within the population. It is not, of course, necessary to vaccinate everyone to prevent an epidemic: immunizing someone not only protects that person but confers some protection to the population in general.

*Conclusion:* The effect of environmental factors on the spread of COVID-19 is of experimental and theoretical importance in understanding the disease dynamics. Stochastic epidemic model, with time-delays, gives an extra degree of authenticity in comparison with its comparing deterministic show. It is consistent with the physical sensitivity and fluctuation of the real observations.

*Acknowledgements:* The study is supported by the United Arab Emirates University.

## References

1. Rihan F.A., Alsakaji H.J. Dynamics of a stochastic delay differential model for COVID-19 infection with asymptomatic infected and interacting people: Case study in the UAE. *Results Phy.* 2021;28(2021):104658.
2. Alsakaji H.J. Stochastic delay differential equations with applications in ecology and epidemics. PhD Thesis UAE University. 2020.
3. Kundu S., Alsakaji H.J., Rihan F.A., Maitra S., Upadhyay R. K. Investigating the dynamics of a delayed stage-structured epidemic model saturated incidence and treatment functions. *Eur Phys J Plus.* 2022;137(1):1-23.
4. Rihan F.A. Delay differential equations and applications to biology. Springer. 2021.
5. Rihan F.A., Alsakaji H.J., Rajivganthi C. Stochastic SIRC epidemic model with time-delay for COVID-19. *Adv Differ Equ.* 2020;1(2020):1-20.

# Variant approach to genus identification of coronavirus genome

Chaley M.<sup>1\*</sup>, Kutyrkin V.<sup>2</sup>

<sup>1</sup> *Institute of Mathematical Problems of Biology, RAS – Branch of Keldysh IAM RAS, Pushchino, Russia*

<sup>2</sup> *Moscow State Technical University n.a. N.E. Bauman, Moscow, Russia*

\* *chaley@phystech.edu*

**Key words:** coronavirus genome, ORF1ab, S-gene, M-gene, N-gene, statistical analysis, variant approach, coronavirus genus identification

A new variant approach to identification of coronavirus genus basing on modified statistic which earlier showed efficiency for flavivirus species recognition is proposed. This statistic is founded on codon frequency distribution in the genes of viral genome. ORF1ab encoding non-structural proteins and the genes of coronavirus structural proteins (S, M and N) are under consideration. The variant approach uses both combinations of the genes and singular genes that are named variants. Six variants are considered in the work. In weighting length of each gene in a variant, joint distribution of codon frequency is calculated. Genus of coronavirus is determined in the result of analysis of all the variants considered. The approach was developed with the help of learning samples of prototype variants for the four genus of the coronaviruses:  $\alpha$ -CoV,  $\beta$ -CoV,  $\delta$ -CoV and  $\gamma$ -CoV. Results of genus identification for certain genome is represented by variant string  $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5, v_6)$ , where  $v_i \in \{\alpha, \beta, \delta, \gamma\}$  is the result of coronavirus determining on the  $i$ th ( $i = \overline{1, 6}$ ) combination of the genes. If at least four out of the six components point at the same genus of coronavirus, then such the genus is recognized as the result of identification. According to this rule reliability of genus identification for more than the 3000 of coronavirus genomes from the GenBank was of 97 % [1]. In the majority of events the components of variant string are represented by the same character. However, the cases are frequent enough when a few components in the string are different from the others. So, the variant approach reveals a mosaicism in determining genus of the coronavirus genome. It seems likely that phenomenon of mosaicism is due to high genetic variability and propensity to homologous recombinations in interpopulation interactions of the coronaviruses and their natural reservoirs.

## References

1. Chaley M., Kutyrkin V. Coronavirus genus recognition based on prototype virus variants. *Math Biol Bioinf.* 2022;17(1):10-27. (in Russian)



# Combining machine learning and non-linear dynamics modeling to understand COVID-19 risk factors

Djordjevic M.<sup>1\*</sup>, Salom I.<sup>2</sup>, Djordjevic M.<sup>2</sup>, Sofija M.<sup>1</sup>, Rodic A.<sup>1</sup>, Milicevic M.<sup>1</sup>, Tumbas M.<sup>1</sup>, Zigic D.<sup>2</sup>

<sup>1</sup> *Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia*

<sup>2</sup> *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*

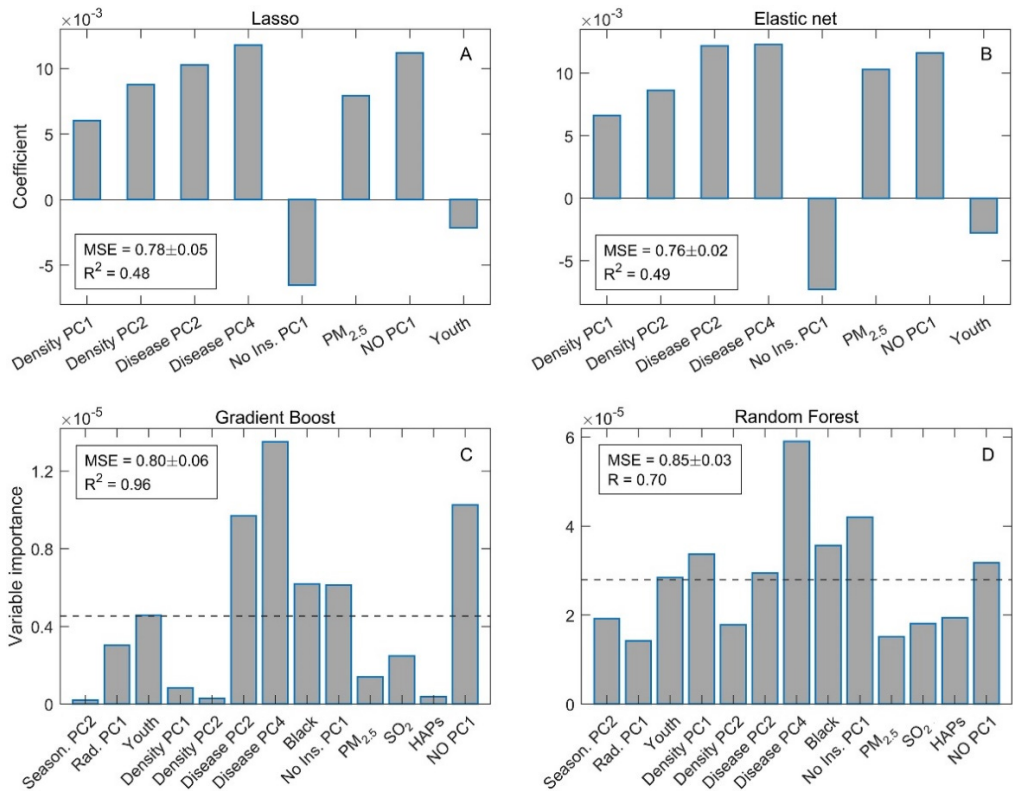
\* [dmarko@bio.bg.ac.rs](mailto:dmarko@bio.bg.ac.rs)

**Key words:** infectious diseases, dynamical epidemics models, COVID-19 transmissibility, COVID-19 severity, feature selection, Random Forest, Gradient Boost

*Motivation and Aim:* Many studies have been investigating the relationship between COVID-19 transmissibility and severity in population and different sociodemographic, health-related, and environmental factors. Confirmed case counts (infected and fatalities) were commonly used as a measure of disease transmissibility and severity. However, a major limitation of these studies is that they do not adequately account for complex disease dynamics, influenced by, e.g., significant differences in control measures, testing policies, and points on the epidemic curve associated with different geographic regions. Another difficulty is appropriately controlling the effects of multiple potentially important factors, due to both their mutual correlations, limited dataset, and potential non-linearities and interactions between these factors. To address these difficulties, we use epidemiological modeling to i) Estimate the basic reproduction number ( $R_0$ ) as a robust measure of the disease severity directly from observed infection count data [1], ii) Propose a disease severity measure independent of the disease transmission dynamics (i.e., the basic reproduction number) that has a direct mechanistic interpretation. The derived measure corresponds to the ratio of population-averaged mortality and recovery rates ( $m/r$ ) [2]. We use these two measures to assess demographic, medical, meteorological, and environmental factors associated with the disease severity. For this, we use machine learning regressions that employ different feature selection methods [1–3]. To reduce dimensionality without complicating the variable interpretation and partially decorrelate the variable set, we employ Principal Component Analysis on subsets of mutually related (and highly correlated) predictors. Finally, we investigate how the results change when the data are analyzed on a smaller special resolution (specifically, when going from USA states to USA counties).

*Results:* We analyze COVID-19 transmissibility [1] and severity [2, 3] on case count data from US states, where we assemble many potentially relevant variables. We use both linear regressions with regularization and feature selection (Lasso and Elastic Net) and non-parametric methods based on ensembles of weak-learners (Random Forest and Gradient Boost). Through these substantially different approaches, we robustly obtain that  $PM_{2.5}$  is a major predictor of  $R_0$  in USA states, while factors such as other pollutants, prosperity measures, population density, chronic disease levels, and possibly racial composition are also significant [1]. For the disease severity, without using prior knowledge from clinical studies, we recover significant predictors known to be risk factors, i.e., age, chronic diseases, and racial factors. Additionally, we identify long-term pollution exposure and population density as not widely recognized (though for the

pollution previously hypothesized) predictors of the disease severity (see Fig. 1 below) [2]. Analysis at the county level also identifies race (proportion of the Black population in particular) as a major factor strongly associated with the disease severity [3].



**Fig. 1.** Machine learning analysis of COVID-19 severity risk factors. A) Lasso and B) Elastic Net, C) Gradient Boosting, and D) Random Forest regressions. For A) and B), the height of bars corresponds to the value of the regression coefficients. For C) and D), the height of bars corresponds to estimated predictor importance. Variable names are indicated on the horizontal axes. Figure adapted from Ref. [2]

**Conclusion:** The measures of the disease progression inferred or derived from the epidemics dynamical modeling proved useful for inferring risk determinants of COVID-19. When applied to data through machine learning approaches, they lead to robust and plausible results that aid understanding COVID-19 risks. Overall, we find that a combination of non-linear dynamics and machine learning is advantageous in analyzing the disease progression, which may also be applied to potential future disease outbursts.

**Acknowledgements:** This study is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

## References

1. Djordjevic M., Rodic A., Salom I., Zigic D., Milicevic O., Ilic B., Djordjevic M. A systems biology approach to COVID-19 progression in a population. *Adv Protein Chem Struct Biol.* 2021;127:291-314.
2. Djordjevic M., Djordjevic M., Ilic B., Stojku S., Salom I. Understanding Infection Progression under Strong Control Measures through Universal COVID-19 Growth Signatures. *Global Challenges.* 2021;5(5):2000101. doi: 10.1002/gch2.20200010.
3. Tumbas M., Markovic M., Salom I., Djordjevic M. A large scale machine learning study of demographic factors contributing to COVID-19 severity, submitted. 2021.

# Analytical and numerical study of infection progression under social distancing measures

Ilic B.<sup>1\*</sup>, Djordjevic M.<sup>1</sup>, Djordjevic M.<sup>2</sup>, Salom I.<sup>1</sup>, Stojku S.<sup>1</sup>

<sup>1</sup> *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*

<sup>2</sup> *Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia*

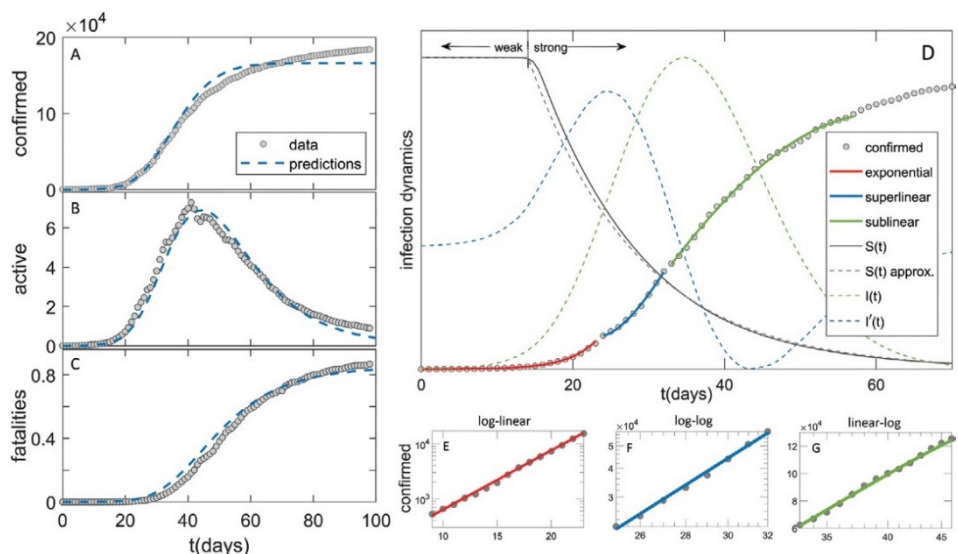
\* *bojanab@ipb.ac.rs*

**Key words:** COVID-19, infection outburst, social distancing measures, epidemiological compartmental models, systems computational biology, SEIR

*Motivation and Aim:* The unprecedented worldwide social distancing measures are introduced to respond to the COVID-19 pandemic outburst. While epidemiological compartmental models have conveyed extensive studies on interventions such as quarantine or vaccination, the effects of social distancing on infection outbursts are poorly understood, and even when considered, they have been addressed only numerically. To this end, we develop an analytically tractable model, which considers the gradual introduction of social distancing and optimally utilizes the vast of publicly available data.

*Methods and Algorithms:* Building upon standard SEIR epidemiological compartmental models, we develop a realistic and analytically solvable SPEIRD (Susceptible-Protected-Exposed-Infected-Recovered-Detected) model [1, 2] that accounts for the effect of social measures by introducing a new compartment – Protected. In particular, the model implies solving highly nontrivial inhomogeneous differential equations. Moreover, an approach well known to theoretical physics (and more recently to quantitative biology) is applied, where one searches for common dynamical features, regardless of other differences.

*Results:* The model delivers closed-form mathematical expressions for the time dependence of infected (I) [2], detected cases (D) [3], and fatalities (F) [3]. These main infection progression data (I(t), D(t), F(t)) are well reproduced (for a majority of COVID-19 hotspots) through joint analytical and numerical analysis, capturing empirically observed COVID-19 growth signatures of detected cases, that is its three distinct growth signatures (exponential, superlinear, and sublinear regime). The growth signatures and associated scaling laws are utilized [2] to pinpoint regions where analytical derivations are most effective for i) assessing the nearly constant value of the scaling exponent in the superlinear regime; ii) understanding the relationship between the duration of this regime and strength of social distancing ( $\alpha$ ); iii) recognizing changes in the reproduction number from infection outburst to its extinguishing; vi) constraining the main parameter quantifying the effect of social distancing ( $\alpha$ ), by analyzing the time duration of the sublinear regime. The combination of  $\alpha$  and the time of social measures introduction [3] defines a new parameter – protection time, essential for public policy decision making. For instance, our study not only suggests that rigorous measures can be often substituted by more relaxed ones imposed at earlier times, but provides a direct analytical expression to quantify this balance. We also provide simple analytical expressions [3] to estimate the peak's timing, the epidemic wave's tipping points, and the maximum of detected cases per day. Additionally, our model is applied to infer key infection parameters [2], such as case fatality, infected fatality, and attack rates.



**Fig. 1.** Comparison between the model (dashed blue curves) and the data (grey circles) in the case of Germany, during the first pandemic wave, for A) detected, B) active cases, and C) fatalities. D) exponential, superlinear, and sublinear fit to detected case data is presented. Arrows indicate the regions with a small and large magnitude of social distancing. The dashed green and blue curve corresponds to the number of infected and its first derivative (whose maxima is  $I(t)$  inflection points) respectively. The detected case counts in the three regimes are displayed on E) log-linear, F) log-log, and G) linear-log scale. Figure adapted from [2]

**Conclusion:** The strength of our analytically tractable model is its ability to qualitatively and quantitatively explain universal growth signatures of infection progression; to provide simple quantitative relations between the model variables and epidemiological observables; to ensure highly constrained infection parameters inference; and to shed light on cause-effect connections that underlie the epidemiological dynamics under social distancing measures, which are obscured in more standard, i.e., purely numerical, approaches. Due to its general relevance, it is applicable to the early stages of any new pandemics when pharmaceutical treatments and vaccines are unavailable and well-timed and appropriately chosen social distancing measures are of the utmost importance. The model does not consider loosening the measures, as was the case during the first COVID-19 wave in many countries. As an outlook, it can be done in our model through a reversible transition from the protected back to the susceptible compartment. Thus, the analytical expressions would be much more cumbersome (or unachievable), though exploring the model numerically in such a case would be highly relevant, which is our future goal.

**Acknowledgements:** This study is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

## References

1. Djordjevic M., Rodic A., Salom I., Zigic D., Milicevic O., Ilic B., Djordjevic M. A systems biology approach to COVID-19 progression in a population. *Adv Protein Chem Struct Biol.* 2021;127:291-314.
2. Djordjevic M., Djordjevic M., Ilic B., Stojku S., Salom I. Understanding Infection Progression under Strong Control Measures through Universal COVID-19 Growth Signatures. *Global Challenges.* 2021;5(5):2000101. doi: 10.1002/gch2.20200010.
3. Ilic B., Salom I., Djordjevic M., Djordjevic M. An Analytical Framework for Understanding Infection Progression Under Social Mitigation Measures. (Submitted to *Nonlinear Dynamics*). 2022. doi: 10.21203/rs.3.rs-1331002/v1.

## Behavior changes of epidemic control in condition of sufficient and limited resources

Kaminskiy G.<sup>1\*</sup>, Semenova D.<sup>1</sup>, Zatsepin O.<sup>2</sup>, Lebedev S.<sup>2</sup>

<sup>1</sup> National Medical Research Center of Phthisiopulmonology and Infectious Diseases of the Ministry of Health of the Russian Federation, Moscow, Russia

<sup>2</sup> Russian Federal Nuclear Center – Zababakhin All-Russia Research Institute of Technical Physics, Snezhinsk, Russia

\* gregkaminski.gk@gmail.com

**Key words:** epidemic process, epidemic control, system behaviour, bifurcation diagram, sufficient and limited resources

*Motivation and Aim:* Epidemic process is the propagation of the pathogen over humans. There are two points of influence on the development of the epidemic process:

1. First, on susceptible individuals by vaccination and pre-contact prophylaxis.
2. Second, on infected individuals by testing, diagnostics, isolation, effective cure and tracing contacts with isolation and diagnostics again.

Influence efficacy on the epidemic process varies. In some circumstances the resources are sufficient, vaccination and isolation are performed with a maximum high rate. In other circumstances the resources are limited. There are limited number of vaccination points of medical workers engaged in vaccination. There can be also limited number of medical and outreach workers with the function to find people with the symptoms of disease. The number of tests can be also limited. If there is a difference in behavior between two systems, up to now it is unknown.

*Methods and Algorithms:* We compared two variants of maximum deliberately simplified system (MDSS)

### MDSS

$$\begin{aligned} X' &= -R\alpha XY + \mu - \mu X - \lambda X^{p_1} \\ Y' &= R\alpha XY - \beta Y - \mu Y - \delta Y^{p_2} \end{aligned}$$

### System1 $p_1 = p_2 = 1$

$$\begin{aligned} X' &= -R\alpha XY + \mu - \mu X - \lambda X \\ Y' &= R\alpha XY - \beta Y - \mu Y - \delta Y \end{aligned}$$

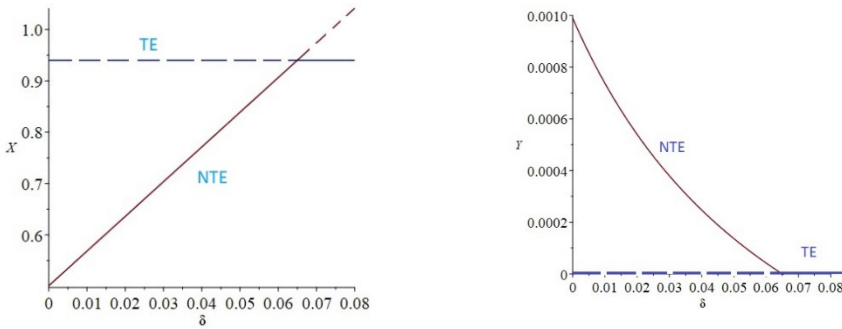
### System2 $p_1 = p_2 = 0$

$$\begin{aligned} X' &= -R\alpha XY + \mu - \mu X - \lambda \\ Y' &= R\alpha XY - \beta Y - \mu Y - \delta \end{aligned}$$

where  $X$  is the proportion of susceptible individuals,  $Y$  is the proportion of infected (uncontrolled sources of infection),  $R$  is the contact rate,  $\alpha$  is the intensity of infectiousness,  $\beta$  is the intensity of recovery,  $\mu$  is the intensity of the natural movement of the population (birth and death rates, inflow and outflow),  $\delta$  is the intensity of detection and taking for cure of sources of infection,  $p_2$  is value of  $\delta$  resources sufficiency,  $\lambda$  is the intensity of vaccination,  $p_1$  is value of  $\lambda$  resources sufficiency. Parameter values as of acute infection:  $\alpha = \beta = 0.074$  (1/day),  $R = 2$ ,  $\mu = 0.000157$  (1/day),  $\lambda = 0.00001$  (individuals/population).

*Results:* In system 1 (proportionate, sufficient resources) the trace ( $tr$ ) of the Jacoby matrix ( $Jac$ ) of the nontrivial equilibrium (NTE) is always below zero, that means that the system is either steady state or unstable with the hard loss of stability while travelling to trivial equilibrium (TE). In figures dash is unstable equilibrium, solid is stable equilibrium.

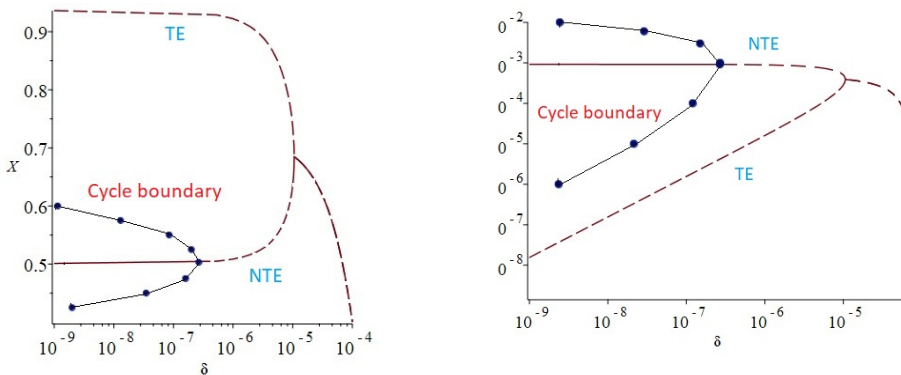
$$tr(Jac(System1)) = -\frac{\mu R \alpha}{\mu + \beta + \delta}$$



**Fig. 1.** Bifurcation diagram of system 1 showing control evolution in direction enhancing force  $\delta$ : X – left, Y – right

In system 2 (unproportionate, insufficient resources) the trace of the Jacoby matrix can be either above or below zero, that completely changes the system behavior. With the growth of  $\delta$  system passes the unstable cycle (1), mild loss of stability (with oscillations) (2), hard loss of stability (without oscillations) (3).

$$\begin{aligned}
 tr(Jac(System2)) = & - \left( ((-\mu + \delta + \lambda))R^2(\mu - \delta - \lambda)^2\alpha^2 - 2R\mu(\mu + \delta - \lambda)(\beta + \mu)\alpha + \mu^2(\beta + \mu)^2 \right)^{\frac{1}{2}} - \mu^3 \\
 & + (R\alpha - \beta - 3\delta + \lambda)\mu^2 + \left( -2R(\delta + \lambda)\alpha - 5\left(\delta - \frac{\lambda}{5}\right)\beta \right)\mu + R(\delta + \lambda)^2\alpha - 2\beta^2\delta \Big) R\alpha / ((\beta \\
 & + \mu)(-\mu^2 + (R\alpha - \beta)\mu - R(\delta + \lambda)\alpha \\
 & - \left( R^2(\mu - \delta - \lambda)^2\alpha^2 - 2R\mu(\mu + \delta - \lambda)(\beta + \mu + \mu^2(\beta + \mu)^2)^{\frac{1}{2}} \right)
 \end{aligned}$$



**Fig. 2.** Bifurcation diagram of unproportionate system 2 showing control evolution in direction enhancing force  $\delta$ : X – left, Y – right

Control aims satisfies only hard loss of stability.

*Conclusion:* In sufficient control resources the system is plane and the hard loss of stability occurs after reaching the threshold. In limited resources the system behavior becomes complex and big oscillations may precause bifurcation to the end of the epidemic process.

### References

1. Arnold V.I. Catastrophe Theory. Berlin, Heidelberg: Springer-Verlag, 1992.

# A Delay Differential Equation approach to model the COVID-19 pandemic

Kiselev I.N.<sup>1,2,3\*</sup>, Akberdin I.R.<sup>1,3,4</sup>, Kolpakov F.A.<sup>1,2,3</sup>

<sup>1</sup> BIOSOFT.RU, LLC, Novosibirsk, Russia

<sup>2</sup> FRC for Information and Computational Technologies, Novosibirsk, Russia

<sup>3</sup> Sirius University, Sochi, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

\* [axec@systemsbiology.ru](mailto:axec@systemsbiology.ru)

**Key words:** SARS-COV-2, epidemiology, mathematical modeling, SEIR, delay differential equations, BioUML platform

*Motivation and Aim:* SEIR (Susceptible – Exposed – Infected – Recovered) approach is a classic modeling method that has frequently been applied to the study of infectious disease epidemiology. However, in the vast majority of SEIR models and models derived from them transitions from one population group to another are described using the mass-action law which assumes population homogeneity. That causes some methodological limitations or even drawbacks, particularly inability to reproduce observable dynamics of key characteristics of an infection such as, for example, the incubation period and progression of the disease's symptoms which require considering different time scales as well as probabilities of different disease trajectories.

*Methods and Algorithms:* In this study, we propose an alternative approach to simulate the epidemic dynamics that is based on two concepts:

1. System of differential equations with time delays in the next form (e.g. for incubation period):

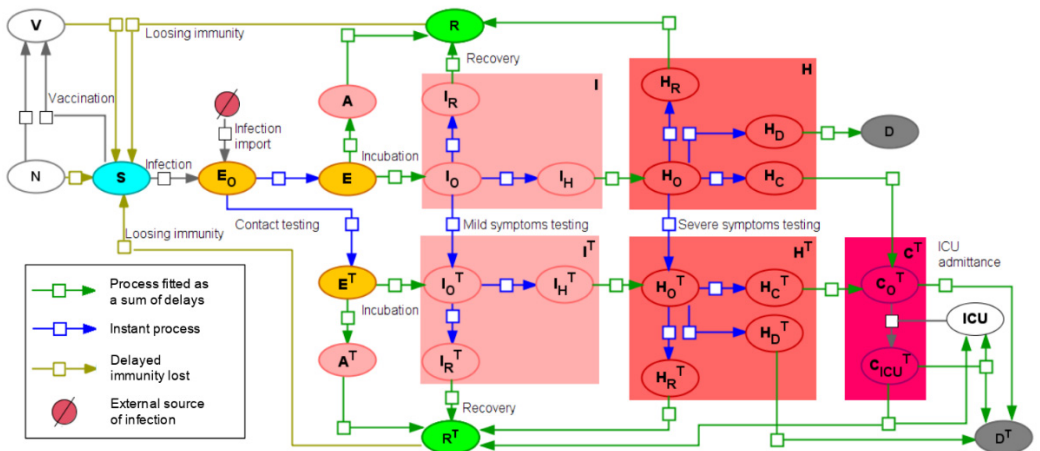
$$\frac{dI}{dt} = \frac{-dE}{dt}(t) = \sum_{i=1}^n k_i \cdot E(t - \Delta_i),$$

where  $n, \Delta_1, \dots, \Delta_n, k_1, \dots, k_n$  - parameters estimated to precisely reproduce a duration of a certain transition like incubation period, recovery period, etc.

2. Instant transitions to replicate correct fractions of patients with particular symptoms. For example, some of patients with mild symptoms will progress to worse symptoms, while the second part of patients will recover. Both these transitions will then be performed with different speed. The issue arises from the state when larger subgroup of patients should undergo slower transition with smaller derivative. Proposed solution is that patients with mild symptoms are instantly divided into two subgroups: “mild symptoms who will recover” and “mild symptoms who will get worse”.

*Results:* The suggested modeling approach is fundamental and can be applied to the study of many infectious diseases epidemiology. We used it to create SEIR-like model in BioUML platform [1] with instant and delayed transitions. The Stringency index [2] was used as a generalized characteristic of the non-pharmaceutical government interventions implemented in corresponding countries to contain the virus spread. The model also includes: availability of intensive care units for critically ill, testing regime, vaccination, loosing of immunity through time and appearance of new strains. The

overall scheme of the model is presented on Fig. 1. The model was calibrated and adapted to COVID-19 pandemic in Germany and France.



**Fig. 1.** Overall SEIR-like model with instant and delayed processes. Colored boxes represent compartments divided into subgroups with different disease paths

*Conclusion:* The developed approach allows for accurate modeling of transitions in the model (e.g. incubation period), while the transition parameters may be estimated separately of the rest of the model. Thus, the number of model parameters is decreased and their epidemiological interpretation becomes more clear (i.e. “fraction of patients with serious symptoms”). The developed models are available through the web interface of BioUML software at <https://gitlab.sirius-web.org/covid-19/dde-epidemiology-model>. Models are available both through visual representation in the platform and in Jupyter notebooks which enable users to reproduce simulation results and figures presented in this study.

*Acknowledgements:* The study was supported financially by the RFBR (No. 20-04-60355 “Development of a multi-scale immuno-epidemiological mathematical model COVID-19, taking into account the impact on the economy of the region and the scenarios of authority actions”).

### References

1. Kolpakov F., Akberdin I., Kashapov T., Kiselev L., Kolmykov S., Kondrakhin Y., Kutumova E., Mandrik N., Pintus S., Ryabova A., Sharipov R. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. *Nucleic Acids Res.* 2019;47(W1):W225-W233.
2. Hale T., Angrist N., Goldszmidt R., Kira B., Petherick A., Phillips T., Webster S., Cameron-Blake E., Hallas L., Majumdar S., Tatlow H. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav.* 2021;5(4):529-538.



## Markov model of hospitalized patients with COVID-19

Kisselevskaya-Babinina V.<sup>1\*</sup>, Romanyukha A.<sup>2</sup>, Sannikova T.<sup>2</sup>

<sup>1</sup> N.V. Sklifosovsky Research Institute for Emergency Medicine of the Moscow Healthcare Department, Moscow, Russia

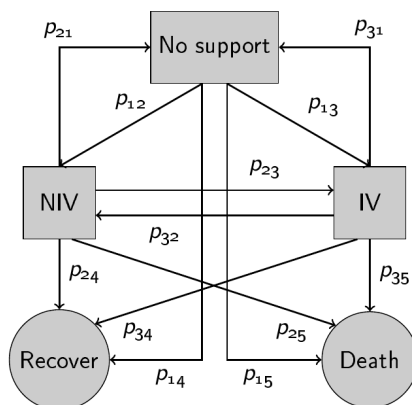
<sup>2</sup> Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia

\* silvaze@yandex.ru

**Key words:** Markov model, covid-19, interleukin-6

*Motivation and Aim:* The work is aimed on developing a mathematical method to evaluate the effectiveness of treatment regimens for hospitalized patients with COVID-19. One of the most dangerous complications for patients with COVID-19 is the respiratory distress syndrome, which requires respiratory support. The respiratory support includes the invasive and noninvasive mechanical lung ventilation. Because the respiratory distress syndrome is often associated with high interleukin-6 levels [1], COVID-19 is recommended to be treated with interleukin-6 inhibition, yet its effectiveness is under question [2]. Our idea is to use a Markov multistate model to describe the changes in patients' respiratory and resulting statuses via cohort analysis and then compare the dynamics for two groups: those who got interleukin-6 inhibitors, and those who didn't. Thus, we can study the effectiveness of interleukin-6 inhibition for patients with COVID-19.

*Methods and Algorithms:* We built the Markov chain model, that included five stages: no respiratory support, invasive (IV) and noninvasive (NIV) types of ventilation, recover with discharge from hospital, and death (Fig. 1). Patients were divided in two groups: those who were treated with inhibitors and those without. The transition probabilities of the model were estimated for two groups separately using the multinomial logistic regression. The 95 % confidence intervals were found using bootstrap method.



**Fig. 1.** Transition diagram of the Markov model of hospitalized patients with COVID-19. Arrows represent transitions between states, circles represent absorbing states. NIV = noninvasive ventilation, IV = invasive ventilation.  $p_{ij}$  are the probabilities of transitions from state  $i$  to state  $j$

*Results:* The results show that using inhibitors reduces the probability of transition to death state from no support or NIV states by 4.5 times on the first five days of patient treatment. After that time the death case probabilities have no significant differences. The patients with no support who got inhibitors have 3 times lower transition rates to IV state and 1.6 times higher transition rates to NIV state. The dynamics of patients already on mechanical ventilation does not differ whether inhibitors were used or not.

*Conclusion:* The Markov model based on respiratory support types showed that inhibitors can prevent patients, who were not under mechanical ventilation, from receiving invasive type of ventilation. There is a reduction of death risk for those patients, but it is short termed. This indicates that the Markov modelling can be used to clarify the impact of different treatment methods on COVID-19 dynamics. The multistate approach also affords a further elaboration to model improvement, as well as estimation of cost effectiveness and the amount of required resources.

*Acknowledgements:* This work was carried out with the financial support of the Ministry of Education and Science of Russia: Grant No. 075-11-2020-011 (13.1902.21.0040).

### *References*

1. Coomes E.A., Haghbayan H. Interleukin-6 in COVID-19: a systematic review and meta-analysis. *Rev Med Virol.* 2020;30(6):1-9.
2. Cavalli G. et al. Interleukin-1 and interleukin-6 inhibition compared with standard management in patients with COVID-19 and hyperinflammation: a cohort study. *Lancet Rheumatol.* 2021;3(4):e253-e261.

## Fighting ticks with functional-analytic guns

Kolokoltsov V.

*Higher School of Economics, Moscow, Russia*

*Sankt Petersburg State University, Faculty PMPU, St. Petersburg, Russia*

*University of Warwick, Coventry, UK*

*kolokoltsov59@mail.ru*

**Key words:** propagation of ticks, ticks' growth control, critical patch size, KISS model, control zones, vector-valued diffusion models, discontinuous diffusion coefficients, intermittency

*Motivation and Aim:* Ticks and tick-borne diseases present a well-known threat to the health of people in many parts of the globe. The scientific literature devoted both to field observations and modelling the propagation of ticks continues to grow. So far, the majority of the mathematical studies were devoted to models based on ordinary differential equations, where spatial variability was taken into account by a discrete parameter. Only few papers use spatially nontrivial diffusion models, and they are devoted mostly to spatially homogeneous equilibria. Here we aim to develop diffusion models for the propagation of ticks stressing spatial heterogeneity. This allows us to assess the sizes of control zones that can be created (using various available techniques) to produce a patchy territory, on which ticks will be eventually eradicated.

*Methods and Algorithms:* The starting point for our analysis is the so-called KISS model for the description of several interacting species (developed for the analysis of planktons). However, we are interested in creating control zones (kind of barriers) with non-beneficial conditions that can stop the propagation of ticks. To this end, we first extend the KISS model to the case when the non-beneficial region is bounded, interpreting its size as an additional control parameter, and when the diffusion coefficient is discontinuous at the interface of beneficial and non-beneficial regions, together with the mortality coefficient. Then we consider the modifications of the model with various boundary conditions, most importantly with the periodic boundary conditions, which provide the key intermediate step to tackle our main questions: how to organise the sizes and properties of the regular patches of (intermittent) control zones that would prevent the growth of the population of ticks?

We suppose that in the control zone one is able to achieve the required level of the death rates or to decrease the reproduction rates of ticks. In literature one can find various methods that can be used locally for this purpose. It can be achieved by the treatment of the soil (planned burns, cleaning and decreasing humidity levels, chemical treatment), by tick-targeted strategies such as TickBots, by specific actions on hosts, like dipping of cattle with acaricide. Alternatively it can be achieved by biological methods like infesting woods with ants (related experiments on ants were conducted in USA and Russia), or the natural enemies of ticks like the insects *Ixodiphagus hookeri*.

From the mathematical point of view, what we are doing is getting criteria for negativity of a class of diffusion operators with piecewise continuous symbols. The negativity of the diffusion operators that govern the propagation of ticks ensures that the population will eventually die out (in the region considered).

*Results:* We give rather precise results for the required lengths of critical zones for a homogeneous population, and then extend them partially to the structured populations, with several stages, like larvae, nymphs and adults in ticks. Using averaged parameters taken from various field observations, we apply our theoretical results to the concrete cases of the lone star ticks of North America and of the taiga ticks of Russia, and thus give concrete assessment of barrier zones with concrete numbers.

*Conclusion:* The main goal was to give exact estimates for the properties of control zones needed for the eradication of ticks. The numbers obtained in our analysis are based on very rough estimates, and our calculations show that they are quite sensitive to the input data. Further detailed concrete observations can make the estimates more precise and more specific for concrete regions.

On the other hand, we used one-dimensional diffusion models aiming at the approximate description of the propagation in a certain direction. Of course, it would be natural to extend the results to a more realistic (for ticks and many other species) two-dimensional diffusion models. One can also argue that approximations that are more appropriate represent other popular models, like the telegraph equations, or sub-diffusions or super-diffusions.

*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of the Russian Federation Grant ID: 075-15-2020-928.

*References* (We quote only some key sources used in our research.)

1. Filippova N.A. (Ed.) Taiga tick *Ixodes persulcatus* Schulze (Acarina, Ixodidae): morphology, systematics, ecology, medical significance. Moscow: Nauka, 1985. (in Russian)
2. Lashin S.H.D., White G A., Leas K., Kelman J., Squire C., Livingston D.L., Sullivan G.A., Baker E.W., Sonenshine D.E. TickBot: A novel robotic device for controlling tick populations in the natural environment. *Ticks Tick Borne Dis.* 2015;6(2)146-151.
3. Harris W.G., Burns E.C. Predation on the lone star tick by the imported fire ant. *Environ Entomol.* 1972;1:362-365.
4. Korotkov U.S. Life cycle of the taiga tick *Ixodes Persulcatus* in the conifer forests of the bottom land of the East Sayan mountain range. *Parasitology.* 2014;48:1:20-36. (in Russian)
5. Tosato M., Nah K., Wu J. Are host control strategies effective to eradicate tick-borne diseases (TBD)? *J Theor Biol.* 2021;508:110483.
6. Vshivkova O.A., Komarov A.S., Frolov P.V., Khlebopros R.G. The role of the heterogeneity of the habitat under the control of the ixode ticks number: cellular-automaton model. *Control Sci.* 2013;4:57-63. (in Russian)
7. Schaefer W.E., Thompson Ch.W., Kribs Ch.M., Gaff H. Dynamics of two pathogens in a single tick population. *Lett Biomath.* 2019;6:1:50-66.
8. Zhang X., Sun B., Lou Y. Dynamics of a periodic tick-borne disease model with co-feeding and multiple patches. *J Mathemat Biol.* 2021;82:27.

# Mathematical modeling and identifiability aspects in epidemiology

Krivorotko O.<sup>1,2,3\*</sup>, Kabanikhin S.<sup>1,2,3</sup>, Zyatkov N.<sup>1</sup>, Sosnovskaia M.<sup>2</sup>

<sup>1</sup> *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

\* *krivorotko.olya@mail.ru*

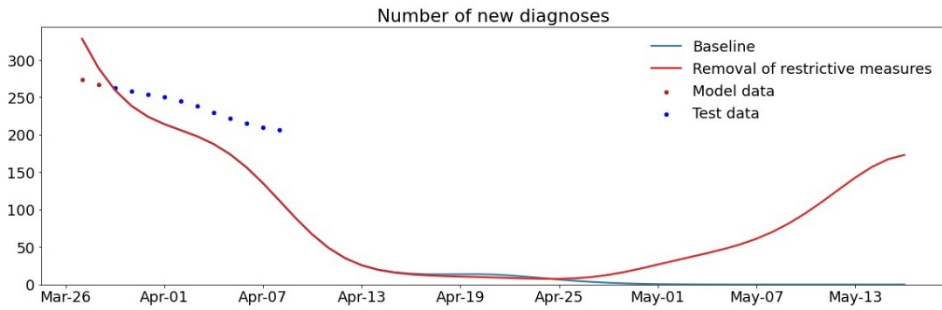
**Key words:** epidemiology, compartment models, agent-based modelling, stochastic models, data analysis, machine learning, optimization, regularization, identifiability, inverse problem, COVID-19, forecasting scenarios, high-performance computing

*Motivation and Aim:* Mathematical modeling of the spread of epidemics is an important tool for analyzing scenarios of infectious disease in the fixed region subject to external influences (restrictive measures, vaccination, virus mutation). Disease spread models are described by differential equations (compartment [1], reaction-diffusion and stochastic models), graph structure (agent-based models) with control and are characterized by their coefficients, individual for a particular disease and region. The aim of the work is to construct algorithms for identifying unknown parameters of compartment and agent-based models using additional information about the number of diagnosed, hospitalized, deceased cases at fixed points in time.

*Methods and Algorithms:* The problem of identification of parameters (the inverse problem) was reduced to the solution of the minimization problem of the misfit function. The inverse problem was solved by several methods and algorithms:

1. A sensitivity-based identifiability analysis of the selected model is performed [2, 3]. As a result, the intervals of variation of unknown parameters are reduced. The probability of hospitalization has been chosen from the web-site Invitro (concentration of IgG antibodies).
2. The regularization algorithm for solving of the inverse problem based on the identifiability analysis is developed using parameter variation limits, a priori information, etc.
3. The stochastic optimization: Tree-structured Parzen Estimator [4], nature-inspired algorithms (genetic, differential evolution, particle swarm), methods of data assimilation, big data analysis and machine learning are applied to solve the minimization problem.

*Results:* The agent-based model was constructed for describing the spread of COVID-19 and the scenarios of its propagation in the nearest 45 days in the Novosibirsk region with external influences [4, 5]. The parameters of agent-based model (infection rate, date of change in the infection parameter, test probability, and initial number of asymptomatic patients) were identified using additional data about number of daily diagnosed and deceased cases from March 12, 2020 to March 31, 2022 (red dots). The figure shows two scenarios of changes in the number of diagnosed cases in Novosibirsk region from April 1 to May 15, 2022: the baseline scenario (blue line) and a removal of restrictive measures (red line). Blue dots (April 1 to April 8, 2022) are the test real data that do not used in inverse problem solution.



**Conclusion:** It is shown that the number of diagnosed cases will increase from April 25, 2022 due to removal of restrictive measures, i.e. cancellation of wearing masks, social distance, isolation (red line). In this case, up to 150 diagnosed cases of COVID-19 are expected by mid-May 2022.

**Acknowledgements:** The study is supported by the Russian Foundation for Basic Research (project No. 21-51-10003) and by the Council for Grants of the President of the Russian Federation (project MK-4994.2021.1.1).

### References

1. Krivorotko O.I., Kabanikhin S.I., Zyatkov N.Yu., Prikhodko A.Yu., Prokoshin N.M., Shishlenin M.A. Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region. *Numer Anal Appl.* 2020;13(4):332-348.
2. Krivorotko O.I., Andornaya D.V., Kabanikhin S.I. Sensitivity analysis and practical identifiability of some mathematical models in biology. *J Appl Ind Math.* 2020;14(1):115-130.
3. Krivorotko O.I., Kabanikhin S.I., Sosnovskaya M.I., Andornaya D.V. Sensitivity and identifiability analysis of COVID-19 pandemic models. *Vavilov J Genet Breed.* 2021;25(1):82-91.
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 Model.* 2022;7:30-44.
5. Kerr C.C., Stuart R.M., Mistry D., Abeyesuriya R.G., Rosenfeld K., Hart G.R. et al. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol.* 2021;17(7):e1009149.

## Autoimmune effect of antibodies against the SARS-Cov-2 nucleoprotein

Matyushkina D.<sup>1\*</sup>, Shokina V.<sup>1</sup>, Tikhonova P.<sup>2</sup>, Manuvera V.<sup>2</sup>, Kharlampieva D.<sup>2</sup>, Lazarev V.<sup>2,3</sup>, Varizhuk A.<sup>2</sup>, Vedekhina T.<sup>2</sup>, Pavlenko A.<sup>1</sup>, Arapidi G.<sup>2,3</sup>, Govorun V.<sup>1,3</sup>

<sup>1</sup> *Scientific Research Institute for Systems Biology and Medicine, Moscow, Russia*

<sup>2</sup> *Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia*

<sup>3</sup> *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow Region, Moscow, Russia*

\* *d.matyushkina@gmail.com*

**Key words:** COVID-19, coronavirus infection, SARS-Cov-2, nucleoprotein, antibodies, autoimmunity

*Motivation and Aim:* New pandemic caused by the SARS-Cov-2 forced the scientific community to throw all capabilities to fight against this disease. Also, it is important not only to eliminate the pathogen from the body, but also to analyze the mechanisms of the occurrence of long-term post-COVID complications. *Methods and Algorithms:* We performed analysis of antibody response of patients with COVID-19 to linear and conformational epitopes of viral proteins by several methods: ELISA, chemiluminescent microparticle immunoassay (CMIA), chip array and western blot. We used downloaded standalone version of NetMHCpan 4.1 and NetMHCIIpan 4.0 [1] to predict protein fragments likely binding the most common alleles of HLA class I and pan-specific binding of peptides to HLA class II alleles of all known sequences respectively. All comparisons of continuous values distributions were performed using ANOVA tests. *Results:* The most intriguing and interesting results were obtained for the nucleocapsid protein. The titer of antibodies to conformational epitopes to nucleoprotein was 2.5 fold higher in the group of severe patients and correlated with the patient's oxygen demand. No less important is the presence of IgG to linear epitopes of the nucleoprotein, which can indicate a risk of development of an autoimmune process. We analyzed cross reactivity of this antibodies with own human proteins. Using competitive western blot analysis, we revealed that the autoimmune manifestation is caused by antibodies against the nucleocapsid protein. According with the possibility of HLA presentation main possible targets were identified by bioinformatic and mass-spectrometry approach. *Conclusion:* As a result of the study, we found out that in monitoring the COVID-19 disease progression it is essential to measure not only the level of antibodies to Spike protein, but especially to NP (both linear and conformational epitopes) because main autoreactive antibodies are associated with nucleoprotein. People with HLA alleles A01:01; A26:01; B39:01; B15:01 are most susceptible to the development of autoimmune processes after COVID-19. So the obtained results can be used for optimizing categorization of patients in terms of disease severity degree and for understanding the possible development of autoimmune complications and post-COVID consequences, for creating protocols for the management and prognosis of patients with COVID-19 as well as for further development of vaccines against coronavirus infection.

*Acknowledgements:* This work was supported by Rosпотребнадзор service state task No. 122030900051-9.

### References

1. Reynisson B. et al. *Nucleic Acids Res.* 2020;48(W1):W449-W454.

## **Coronavirus SARS-CoV2: preliminary results of the pandemic**

Netesov S.V.

*Novosibirsk State University, Novosibirsk, Russia*

*netesov.s@nsu.ru*

**Key words:** SARS-CoV2

In the report presented the modern taxonomy of the coronavirus family and the information about coronaviruses as pathogens of humans, domestic and farm animals and birds. The variety of variants of the current coronavirus SARS-CoV-2 and their differences from each other, as well as their relative sensitivity to vaccines are presented. The dynamics of the course of the pandemic in different countries of the world, the types and effectiveness of vaccines developed and produced, and the rate of vaccination against coronavirus are shown. Concentrated and presented information on methods of treatment and medicinal preparations. Finally, options for forecasting the development and possible attenuation of the coronavirus pandemic in the near future were discussed.



# Coupled agent-based model of COVID-19 and major depressive disorder spread in Moscow

Novikov K. \*, Romanyukha A.

*Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

\* [konst.novikov@gmail.com](mailto:konst.novikov@gmail.com)

**Key words:** COVID-19, agent-based model, major depressive disorder, QALY

*Motivation and Aim:* During the first year of COVID-19 pandemic the prevalence of major depressive disorder (MDD) in some populations reportedly increased by 29.4 %, the increase is believed to be associated with the lack of personal communications caused by lockdowns, distant work and education [1]. Using agent-based modeling approach for COVID-19 and MDD spread in Moscow population we investigate the effect of epidemic containment measures on physical and mental health.

*Methods and Algorithms:* We developed an agent-based coupled COVID-19 – MDD spread model, designed for Moscow population. The COVID-19 submodel is based on realistic age-sex composition of Moscow households, correlated chronic diseases, and statistic-based description of places of study and work. Modeling of MDD is based on the psychological stress concept; high level of stress leads to increased risk of getting depression. We considered mobility and a number of infected relatives as predictors of stress level. Modified (considering the stochastic nature of agent-based model and integer-valued parameters) multistart gradient descent optimization was used to fit the model to Moscow COVID-19 statistics data [2].

*Results:* The modeling results reveal that human toll in terms of quality adjusted life years related to mental disorders (MDD and additionally anxiety disorder) is as large as 48 % of COVID-related human toll. Model-based analysis shows that lockdowns despite the significant mobility decrease lead to better outcome both for physical and mental health.

*Conclusion:* The results obtained using the model show that mental disorders associated with epidemic containment measures leads to considerable human toll. Therefore, further cost-benefit analyses should consider both the estimates of economic loss related to lockdowns and mental disorders caused by mobility restrictions.

*Acknowledgements:* This work was carried out with the financial support of the Ministry of Education and Science of Russia: Grant No. 075-11-2020-011 (13.1902.21.0040).

## References

1. Santomauro D.F. et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the covid-19 pandemic. *Lancet*. 2021;398(10312):1700-1712.
2. Moscow Mayor official website. <https://www.mos.ru/> (Accessed 21.04.2022).

## Mathematical modeling of the dynamics of incomes of the population under epidemic process

Petrakova V.<sup>1\*</sup>, Krivorotko O.<sup>2,3</sup>

<sup>1</sup> *Institute of Computational Modeling, SB RAS, Krasnoyarsk, Russia*

<sup>2</sup> *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *vika-svetlakova@yandex.ru*

**Key words:** SIR model, macroeconomic indicators, ordinary differential equation, optimization problem

*Motivation and Aim:* The covid-19 epidemic has had an impact not only on the health of the population, but also on the economy of individual settlements and, in general, the whole world. Any significant increase in the incidence, even on a local scale of the city, leads to a significant adjustment in the functioning of the labor market. This paper considers a mathematical model of the impact of the spread of morbidity on the macroeconomic indicators of a particular region, such as the average per capita income of the population and the size of the labor force.

*Methods and Algorithms:* The paper presents a mathematical model built on ordinary differential equations of the SIR type [1] and the supplemented urban Lotka–Volterra model [2]. Based on the proposed differential model, an optimal control problem has been presented that allows taking into account not only the economic and epidemiological components, but also the social interaction of people within the population. The Pontryagin maximum principle was used to solve the optimal control problem.

*Results:* As a result, two well-analytically analyzed models of the impact of the spread of the virus on the most important macroeconomic indicators of the region were obtained. Various scenarios of the dynamics of the economic and epidemiological components of the Novosibirsk region are constructed depending on the different social behavior of its inhabitants.

*Conclusion:* The models obtained in the work can be improved by refining the epidemiological model and expanding the differentiation of the population relative to the epidemiological layers. The results of the analysis of the optimization model show a significant influence of social sentiments on the spread of the virus within the population and the economic situation. Consideration of epidemiological parameters in the form of control functions allows, in a way, to take into account the heterogeneity of the population and consider model parameters that are sensitive to changes within the interval, avoiding determining their exact value.

*Acknowledgements:* This work was supported by the Russian Science Foundation (project No. 18-71-10044).

### References

1. Kermack W., McKendrick A. A contribution to the mathematical theory of epidemics. *Proc Math Phys Eng Sci.* 1927;115:700-721.
2. Kamann D.-J.F., Nijkamp P. Technogenesis: Incubation and Diffusion. In: Research Memorandum. Amsterdam: VU University, 1988;24:1-40.

# Qualitative and quantitative features of delay differential equations and applications to biology

Rihan F.A.

*Department of Mathematical Sciences, College of Science, UAE University, Al-Ain, UAE  
frihan@uaeu.ac.ae*

**Key words:** Delay Differential equations, Continuous RK schemes, Stability, parameter estimations, Optimal control and dynamic diseases

*Motivation and Aims:* Recently much attention has been given to mathematical modeling of real-life phenomena using differential equations with memory, such as delay differential equations (DDEs). This is due to the fact that introduction memory terms in a differential model significantly increases the complexity of the model. Such a class of DDEs is widely used for analysis and predictions in various areas of life sciences and modern topics in population dynamics, computer science, epidemiology, immunology, physiology, and neural networks. In this paper, we provide a wide range of delay differential models that have a richer mathematical framework (compared with ODEs) for the analysis of biosystems. Qualitative and quantitative features of DDEs are discussed. Some numerical simulations are also provided to show the effectiveness of the theoretical results.

*Results and Conclusions:*

1. DDEs of real-life phenomena have potentially more interesting dynamics than equations without memory;
2. The time-delay (memory) is incorporated into systems to describe resource regeneration times, maturation periods, reaction times, feeding times, gestation period, etc.;
3. Studying qualitative behaviour of DDEs is essential for ensuring safe applications or real problems;
4. The presence of memory in the model enriches the dynamics of the model: It can stabilize or dis-stabilize the system
5. Mono-implicit continuous RK schemes are stable and suitable for stiff and non-stiff DDEs
6. Sensitivity functions clearly demonstrate the measure of the importance of the input parameters.
7. Sensitivity functions enable one to assess the relevant time intervals for the identification of specific parameters.

*References*

1. Rihan F.A Delay Differential Equations and Applications to Biology. Springer, 2021. doi: 10.1007/978-981-16-0626-7.
2. Rihan F.A. Remarks and Current Challenges. In: Delay Differential Equations and Applications to Biology. Forum for Interdisciplinary Mathematics. Singapore: Springer, 2021. doi: 10.1007/978-981-16-0626-7\_14.
3. Rihan F.A. Delay Differential Equations with Infectious Diseases. In: Delay Differential Equations and Applications to Biology. Forum for Interdisciplinary Mathematics. Singapore: Springer, 2021. doi: 10.1007/978-981-16-0626-7\_8.
4. Rihan F.A. Delay Differential Equations of Tumor-Immune System with Treatment and Control. In: Delay Differential Equations and Applications to Biology. Forum for Interdisciplinary Mathematics. Singapore: Springer, 2021. doi: 10.1007/978-981-16-0626-7\_9.

# Modeling the COVID-19 transmission dynamics and identifying meteorological and sociodemographic predictors

Rodic A.<sup>1\*</sup>, Salom I.<sup>2</sup>, Milicevic O.<sup>3</sup>, Markovic S.<sup>1</sup>, Zigic D.<sup>2</sup>, Ilic B.<sup>2</sup>, Djordjevic M.<sup>2</sup>, Djordjevic M.<sup>1</sup>

<sup>1</sup> *Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia*

<sup>2</sup> *Institute of Physics, University of Belgrade, Belgrade, Serbia*

<sup>3</sup> *Department for Medical Statistics and Informatics, Faculty of Medicine, University of Belgrade, Belgrade, Serbia*

\* *andjela.rodic@bio.bg.ac.rs*

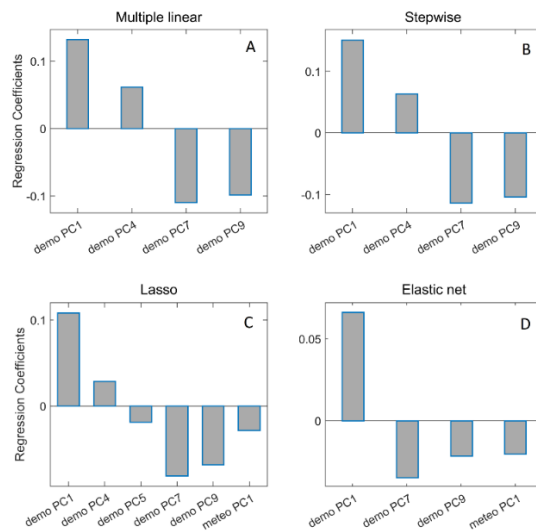
**Key words:** COVID-19, epidemiological model, basic reproduction number, SARS-CoV-2 transmissibility predictors, machine learning, feature selection and regularization, Principal Component Analysis, mitigation measures

*Motivation and Aim:* The spread of the SARS-CoV-2 virus during the first wave of the COVID-19 epidemic depended on its inherent transmissibility in a given population and the effectiveness of practicing the social distancing measures. To analyze the conditions favoring the fast infection spread in a population, one first needs to choose an appropriate measure of that spread. The Basic Reproduction Number ( $R_0$ ) measures the inherent virus transmissibility in a fully susceptible population in absence of any interventions. In addition, we argue that it is highly independent from the testing capacity. Then, one needs to identify the direct predictors of  $R_0$  among potentially relevant meteorological and sociodemographic factors. However, this is challenging due to high mutual correlations between some of the factors, which can therefore be obtained as indirect predictors. The social distancing measures can significantly impact the rate of the infection spread, but they have been introduced in different populations with different efficiencies. In China, the numbers of cases and deaths reached in Wuhan were orders of magnitude higher than the total numbers reported for the rest of the country. Therefore, one may ask if this prominent difference can be explained by an interplay of the inherent virus transmissibility and the effects of the applied control measures.

*Methods and Algorithms:* We constructed a dynamic compartmental model describing the infection propagation upon the virus introduction in a population. For 118 countries, we estimated the values of  $R_0$  from the fit of the model to the initial exponential phase of the case growth curves, characterized by the lack of intervention measures. First, we determined bivariate Pearson correlations of  $R_0$  with different meteorological and sociodemographic variables [1]. Then, we performed an advanced statistical analysis to infer the direct  $R_0$  predictors [2]. Namely, we applied the Principal Component Analysis to the meteorological and demographic variable sets to exchange them for the corresponding sets of non-correlated Principal Components (PCs). Next, we searched for robust  $R_0$  predictors among the PCs using independently four different methods: the custom multiple regression analysis, the Stepwise regression, the Lasso, and the Elastic net. The last two methods utilize both regularization and variable selection. To interpret the resulting predictor PCs, we analyzed their correlations with the starting variables. Regarding the epidemic growth phase influenced by social distancing measures, we enabled the model to describe it by incorporating the gradual flow of susceptible, healthy individuals practicing the measures to a “protected” compartment unavailable to the infection [3]. We used a constrained numerical

procedure to fit the full model to the cumulative case counts data for each of the China provinces and estimated the corresponding virus transmission ( $R_0$ ) and social distancing effectiveness parameters.

**Results:** The results of the advanced statistical analysis indicate that many of the strong correlations with  $R_0$ , obtained by the bivariate analysis [1] are probably the indirect consequences of the mutual variable correlations [2]. All four methods (Fig. 1, A–D) singled out the demographic PCs 1, 7 and 9 as the most direct predictors of  $R_0$ , which are dominantly correlated with the country development indicators, the built-up area per capita, and the time from the global epidemic onset respectively. Three methods (see Fig. 1, A–C) indicate the importance of an unhealthy lifestyle and environment through the PC4, while two methods (see Fig. 1, C–D) acknowledge the independent influence of the weather seasonality through the meteo PC1, despite its high correlation with the country development level. Among the China provinces, Wuhan appears as an outlier, with relatively high inherent virus transmission combined with less efficient epidemic control measures [3]. While it shows the highest Case Fatality Ratio, its Infection Fatality Ratio is average. The results also indicate that provinces tightened the measures with the increase of  $R_0$ .



**Fig. 1.** Inferring direct  $R_0$  predictors using four independent methods (A–D). Adapted from [2]

**Conclusion:** Among different meteorological and sociodemographic factors, which can be correlated both mutually and with  $R_0$ , the prosperity level, indoor crowdedness, the epidemic onset, and unhealthy lifestyle appear as the most robust, direct predictors of the SARS-CoV-2 transmissibility. The interplay of the inherent virus transmissibility and the effectiveness of the introduced intervention measures can reasonably explain very different epidemic outcomes in the provinces of China.

**Acknowledgements:** This study is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

## References

1. Salom I. et al. Effects of Demographic and Weather Parameters on COVID-19 Basic Reproduction Number. *Front Ecol Evol.* 2021;8:617841.
2. Djordjevic M. et al. Inferring the Main Drivers of SARS-CoV-2 Global Transmissibility by Feature Selection Methods. *GeoHealth* 2021;5(9):e2021GH000432.
3. Djordjevic M. et al. A systems biology approach to COVID-19 progression in population. *Adv Protein Chem Struct Biol.* 2021;127:291-314.

## Seeking the threshold for herd immunity in the synthetic population of a medium-sized city

Sannikova T.E.<sup>1\*</sup>, Danilova S.S.<sup>2</sup>

<sup>1</sup> *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

<sup>2</sup> *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

\* *te\_san@yahoo.com*

**Key words:** agent-based model, COVID-19, epidemiology, contact network, vaccination, reinfection

*Motivation and Aim:* The concept of herd immunity has been repetitively used by the public health authorities since the beginning of 2020 as a necessary condition for the end of the COVID-19 pandemic. It can be defined as a proportion of the population that is needed to be immune for the rate of the infection to decline [1]. From the mathematical point of view, the simplest way to calculate it is  $1-1/R_0$ , where  $R_0$  is a basic reproductive number of the infection. But the problem is that the estimates of  $R_0$  for COVID-19 varied widely, from 2 to 7 [2], that leads to different estimates for the threshold value of herd immunity. At the same time this formula does not contain the efficiency and the duration of the immunity and therefore it can be useful only within the short period of time when the recovered individuals are entirely protected. The aim of our work is to simulate the circulation of a respiratory virus in a synthetic population of 200 000 individuals and to reveal the leading factors in forming herd immunity.

*Methods and Algorithms:* We developed an agent-based model that consists of a synthetic population of agents and a set of rules that determine contacts between them. Each agent has a set of properties such as age, sex and social status the distributions of which were typical for medium-sizes Russian cities. Modeling disease transmission we considered only those physical contacts when people can be at a short distance from each other such as contacts between family members, classmates, colleagues and contacts in a public transport. For this purpose, we use data on the number of enterprises and public transport routs of several for medium-sizes cities. Using age-specific contact patterns [3] we simulate the spread of new virus in the population.

*Results:* By changing the number of contacts between agents we simulated the spread of viral infection with and without quarantine measures. We demonstrated that the quarantine measures strongly influence on the dynamics and help to control as well as an abrupt cancellation of all restrictive measures can provoke a postponed growth of the incidence. It was shown that the herd immunity can be achieved during one year only if the immunity after recovery is effective in more than 70 % cases of reinfection and without any restrictive measures. If the immune subpopulation protected from the viruses in less than 40 % cases of infections, the herd immunity is not achievable and the pathogen causes recurrent waves of infection.

*Conclusion:* The duration of natural immunity from COVID-19 and its effectivity against new strains are the subject of many investigations for the moment as well as the effectiveness of the vaccination. One way to assess the level of population immunity is to estimate the seroprevalence of anti-SARS-CoV-2 IgG antibodies [4, 5]. The following step is to incorporate such information in the model and to simulate the circulation of existing and new strains of SARS-Cov2.

*Acknowledgements:* This work was carried out with the financial support of the Ministry of Education and Science of Russia: Grant No. 075-11-2020-011 (13.1902.21.0040).

### *References*

1. McDermott A. Core Concept: Herd immunity is an important—and often misunderstood—public health phenomenon. *PNAS*. 2021;118(21):e2107692118.
2. Yu C.J. et al. Assessment of basic reproductive number for COVID-19 at global level: A meta-analysis. *Medicine*. 2021;100(18):e25837.
3. Ajelli M., Litvinova M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J Theor Biol*. 2017;419:1-7.
4. Stringhini S. et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet*. 2020;396(10247):313-319.
5. Krieger E. et al. Seroprevalence of SARS-Cov-2 Antibodies in Adults, Arkhangelsk, Russia. *Emerging Infect Dis*. 2022;28(2):463.

# Consumer loan demand modeling of households in Russia under sanctions

Shananin A.<sup>1,2,3,4\*</sup>, Trusov N.<sup>1,2,3\*\*</sup>

<sup>1</sup> *Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Moscow, Russia*

<sup>2</sup> *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow Region, Russia*

<sup>3</sup> *Federal Research Center "Computer Science and Control", RAS, Moscow, Russia*

<sup>4</sup> *Peoples' Friendship University of Russia (RUDN University), Moscow, Russia*

\* alexshan@yandex.ru, \*\* trunick.10.96@gmail.com

**Key words:** modified Ramsey model, consumer behavior, forecasts

Consumer credit is an important instrument of the economic relations in Russia. It supports the effective demand of the population. The majority of borrowers in Russia have low real incomes, which leads to an increase in consumer lending. The current situation may turn into a financial pyramid; the borrowers in the mid-term will not be able to service their loans. This will give a rise to a banking system crisis. Nowadays, the problem of growing debt on consumer loans has sharply escalated. The sanctions are aimed at lowering the standard of living of households due to the crisis in the banking system, there is a sharp increase in the key rate of the Central Bank of the Russian Federation; the inflation expectations are significantly increasing.

The modeling the economic behavior of households is based on the modified Ramsey model [1-3]. The model is identified according to Russian statistics. A special software has been developed: with the help of the Household Budgeted Survey collected by the Federal State Statistics Service have been, the user can easily construct the classification, identification of the representative household in a certain group of regions, and build forecasts. In this work, the authors investigated the dynamics of the economic situation of borrowers in different regions of the Russian Federation, considering the sharp change in monetary policy and the growth of inflationary expectations. Various options for state support and monetary policy are considered in terms of their impact on the state of the consumer lending market.

*Acknowledgements:* The work has been supported by Russian Foundation for Basic Research (grant No. 20-07-00285).

## References

1. Shananin A.A., Tarasenko M.V., Trusov N.V. Mathematical Modeling of household economy in Russia. *Comput Math Math Phys.* 2021;61(6):1030-1051.
2. Shananin A., Tarasenko M.V., Trusov N.V. Consumer Loan Demand Modeling. *Mathematical Optimization Theory and Operations Research: Recent Trends. CCIS.* 2021;1476:417-428.
3. Shananin A.A., Trusov N.V. The household behavior modeling based on Mean Field Games approach. *Lobachevskii J Math.* 2021;42(7):1738-1752.



## Forecast of the development of the COVID-19 epidemic situation in Moscow in 2022–2023

Taranik A.V.\*, Lebedev S.N., Litvinenko I.A., Baydin G.V., Pavlenko O.N.,  
Belova M.G., Besova E.V.

*Russian Federal Nuclear Center – Zababakhin All-Russia Scientific Research Institute of Technical  
Physics, Snezhinsk, Russia*

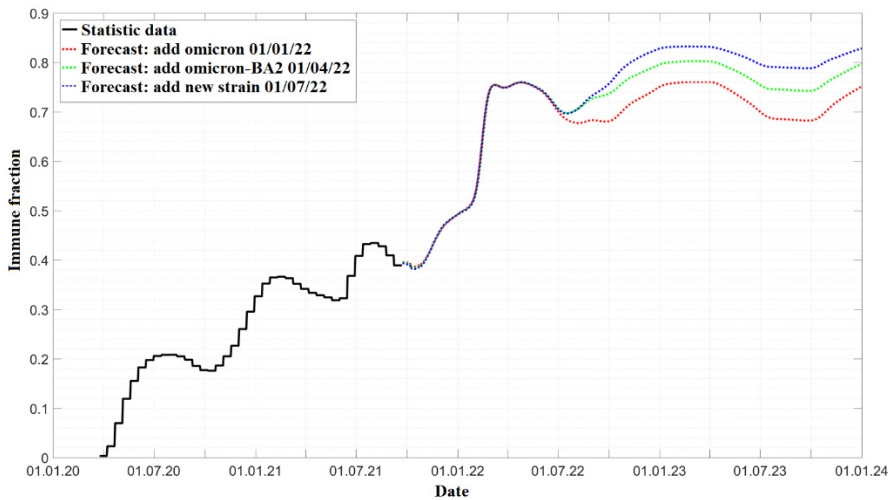
\* *a.v.taranik@mail.ru*

**Key words:** COVID-19, agent stochastic model, social contacts, epidemic situation

*Motivation and Aim:* The COVID-19 pandemic is a global challenge. The analysis of the development of the epidemic situation and its forecasting for the governments of many countries has been a problem for the last two years. The purpose of our work is to obtain an assessment of the development of the situation in Moscow and the UK in the coming years.

*Methods and Algorithms:* We have created an agent-based stochastic model for predicting the development of the epidemiological situation in the focus of the spread of viral infection through daily social contacts [1]. When constructing an epidemiological model, numerical parameters of the virtual epidemic process are coordinated with retrospective actual data from the beginning of the epidemic. The forecast is constructed as a continuation of the obtained epidemiological model within the framework of the selected assumptions: the appearance of new strains, seasonal factors and possible actions of the authorities. To analyze the ways of development of the epidemic situation, we also use the results of unique population studies in the UK conducted under the REACT [2] and ZOE [3] programs.

*Results:* An epidemiological model has been built for Moscow, according to which a forecast for the next two years has been made. Using the example of Moscow, it is shown that the appearance of omicron in January 2022 and its faster "descendants" will lead to the circulation of the virus, in which the immune layer will be maintained at the level of 70–80 %, which indicates continuous infection of people with a drop in their immunity six months after the previous disease. The analysis of population studies in the UK does not contradict the obtained modeling results.



*Conclusion:* Two years of the Covid-19 pandemic have not led to a decline in the circulation of the virus in human communities. The example of the UK shows that the circulation of the virus has moved from an outbreak mode to a continuous one. It is suggested that it is difficult to change strains to more transmissible ones due to the small proportion of susceptible ones in the population and the inability to realize their advantages relative to already circulating strains. A decrease in the importance of the seasonal factor in the spread of COVID-19 infection was noted.

*Acknowledgements:* The study is supported by Ministry of Education and Science of Russia: grant No. 075-11-2020-011 (13.1902.21.0040).

### References

1. Taranik A.V., Lebedev S.N., Litvinenko I.A., Baydin G.V., Pavlenko O.N., Belova M.G., Besova E.V. Making a Forecast of the development of the epidemiological situation in the pesthole of the spread of viral infection through the daily social contacts of residents. In: XV International conference Zababakhin Scientific Talks: 27.09.2021-01.10.2021, Snezhinsk, Russia. 2021:255.
2. REACT 1 study materials. Imperial College London. <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/react-1-study-materials/> (Accessed 14.04.2022).
3. ZOE Wider Health Studies, <https://covid.joinzoe.com/wider-health-studies> (Accessed 14.04.2022).

# Transmission of acute respiratory infections in a city: agent-based approach

Vlad A.I.\*, Sannikova T.E., Romanyukha A.A.

*Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

\**firkhraag@gmail.com*

**Key words:** agent-based model, epidemiology, respiratory infections, viruses, contact networks

*Motivation and Aim:* Acute respiratory infections (ARI) are the major public health concern, being the principal cause of morbidity and mortality in many countries. They are caused by various pathogens with viruses accounting for the majority of cases. The purpose of this model is to combine heterogeneous data into a dynamic system that can reproduce a seasonal increase in the incidence of ARI, taking into account their etiology, in a city with a temperate climate and that can be used to explore additional scenarios, such as the effectiveness of interventions or the possible impact of global warming.

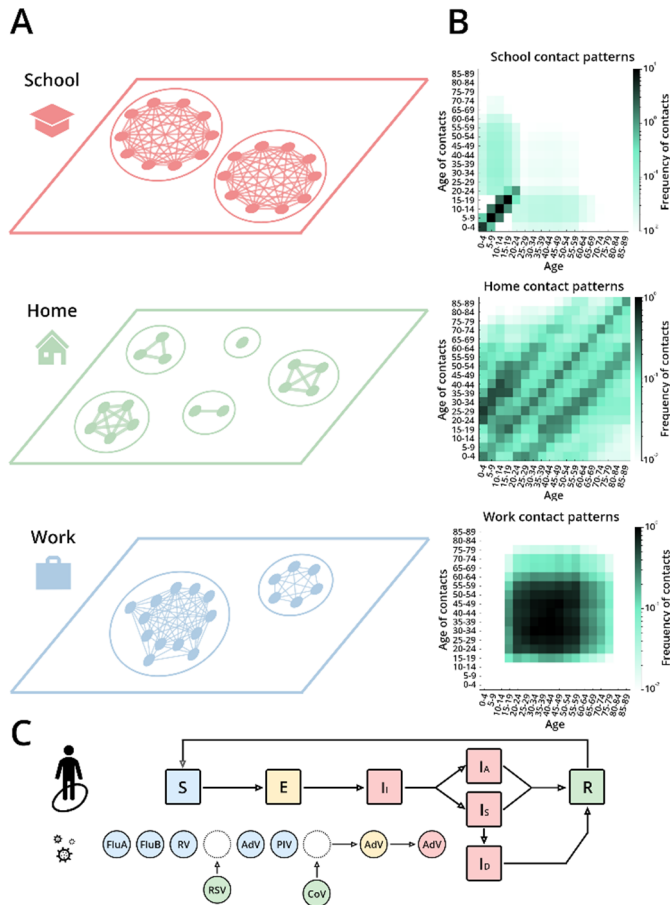
*Methods and Algorithms:* We developed an agent-based model to simulate the co-circulation of different respiratory viruses in a city (Fig. 1). We included seven predominant circulating viruses in Russian cities: influenza A and B, rhinovirus, respiratory syncytial virus, adenovirus, parainfluenza and coronavirus [1], with their properties, such as mean viral load [2], duration of incubation period [3], duration of symptoms [4] and probability of being asymptomatic [5], being estimated from the available literature. The model consists of multiple entities called agents, which are characterized by a set of properties. Such properties as age, sex, and economic status are set using city-specific demographic and socio-economic information from the Russian Census of 2010. Agents are assigned to households and can also be assigned to one main activity (school or work). The model simulates one year and has discrete time steps equal to a single day. Transmission of the viruses occurs when there is a contact between two agents in the same setting, which is modeled using contact networks represented as complete graphs for households and school classes, and Barabasi-Albert graphs for workplaces. Each agent can be as either susceptible, exposed, infectious, or recovered according to the stage of the infection. The probability of transmission is defined by the product of four main factors: infectivity of infectious agent, susceptibility to the virus of susceptible agent, duration of their contact and air temperature on the current step. Model calibration uses the simulation optimization that minimizes the normalized mean absolute error between the simulated and observed epidemic curves.

*Results:* We simulated the spread of respiratory infections observed in Moscow, Russia, in 1996–2002. The proposed model provided a good fit for the average epidemic curve, curves for 7–14 and 15+ age groups and curves for influenza A and B, respiratory syncytial virus, parainfluenza and adenovirus. We were able to capture the dynamics corresponding to the school holidays. Our estimates for 3–6 age group and rhinovirus were slightly underestimated. An incidence in 0–2 age group was greatly underestimated. Our model also greatly overestimated an incidence of coronavirus in February. We also examined two additional scenarios. In the first scenario, we simulated one-week school and class closures when student absences due to ARI were at or above 20 %, and found that it led to the reduction in the total number of diagnosed cases by

12 % and its maximum peak by 23 %. In the second scenario, representing a global warming, we increased the average air temperature by 2 °C and found that while the total number of diagnosed cases decreased by 7 %, the maximum peak size decreased by 12 %.

**Conclusion:** We created a model that can be used as a tool to simulate future outbreaks of emerging viruses in order to improve our understanding of the ongoing epidemic and to devise appropriate interventions.

**Acknowledgements:** This work was carried out with the financial support of the Ministry of Education and Science of Russia: Grant No. 075-11-2020-011 (13.1902.21.0040).



**Fig. 1.** Model structure. (A) Contact networks for different activities. (B) Age-specific contact patterns for different activities. (C) Within-host disease progression

## References

1. Karpova L.S. et al. The Impact of Influenza of Different Etiologies on other ARVI in Children and Adults in 2014 to 2016. *Epidemiology Vaccinal Prevention*. 2018;17(6):35-47.
2. Feikin D.R. et al. Is Higher Viral Load in the Upper Respiratory Tract Associated With Severe Pneumonia? Findings From the PERCH Study. *Clin Infect Dis*. 2017;64(3):S337-S346.
3. Lessler J. et al. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis*. 2009;9(5):291-300.
4. Carrat F. et al. Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies. *Am J Epidemiol*. 2008;167(7):775-785.
5. Galanti M. et al. Rates of asymptomatic respiratory virus infection across age groups. *Epidemiol Infect*. 2019;147:e176.

# 10

Symposium “Cognitive sciences,  
neurogenetics, neuroinformatics  
and systems computational biology”



## The effect of short photoperiod on the behavior, brain serotonin system and BDNF in mice that differ in predisposition to catalepsy

Adonina S.\*, Bazhenova E., Bazovkina D.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\**sadonina23@gmail.com*

**Key words:** depression, serotonin, catalepsy, behavior, SAD

*Introduction:* Seasonal affective disorder (SAD) is a complex of behavioral disorders that occur due to a decrease of the light phase of the day. According to the World Health Organization, such violations are observed in about 20–25 % of people in the autumn-winter period of the year. It is now believed that the formation of SAD is associated with an imbalance in the functioning of the brain serotonin (5-HT) system, neurotrophic factors and circadian rhythms. However, the pathophysiological and genetic mechanisms of the development of SAD remain poorly understood nowadays.

Catalepsy is a prolonged motor inhibition, presenting a syndrome of several psychopathologies as schizophrenia and affective disorders. A large amount of data has been accumulated on the involvement of the brain 5-HT system and BDNF (Brain-derived neurotrophic factor) in the regulation of hereditary catalepsy in mice of CBA/Lac strain that are characterized by high genetic predisposition to a pathological cataleptic reaction. It has been suggested that CBA/Lac mice may be sensitive to the exposure of decreased daylight length. The aim was to study and compare the effects of short photoperiod on the behavior and state of brain serotonin system in mice of non-cataleptic C57BL/6J and cataleptic CBA/Lac strains.

*Methods:* Adult males (2 months) of the experimental groups were kept under conditions of short photoperiod (light/dark: 4/20) for 6 weeks, whereas the animals of the control groups were kept under standard conditions (light/dark: 14/10). Behavior was assessed in the "open field", "forced swimming", "tail suspension" and "pinch-induced catalepsy" tests. The levels of 5-HT and its metabolite 5-HIAA in brain structures (prefrontal cortex, hippocampus, midbrain, hypothalamus) were studied by HPLC. The mRNA levels of genes encoding a) BDNF and its TrkB, p75 receptors and b) key elements of brain 5-HT system were determined using Real-Time PCR. The data were analyzed by two-way ANOVA followed by LSD post-hoc test.

*Results:* Exposure to a short photoperiod resulted in persistent rise of depression-like behavior in mice of both strains ( $p < 0.05$ ), as well as an increase in duration of cataleptic freezing in CBA/Lac animals ( $p < 0.05$ ) compared to control mice.

A decrease in the level of serotonin in the hypothalamus ( $p < 0.05$ ) and hippocampus ( $p < 0.05$ ) and the level of its metabolite 5-HIAA in the hypothalamus ( $p < 0.05$ ) was observed only in mice of the cataleptic CBA strain under conditions of keeping with shortened daylight hours.

Exposure to a short photoperiod led to a decrease in the expression of 5-HT<sub>1A</sub> receptor genes in the cortex ( $p < 0.05$ ) and 5-HT<sub>4</sub> receptor genes in the hippocampus ( $p < 0.05$ ) in C57Bl/6 mice, as well as a decrease in gene expression of 5-HT<sub>1A</sub> and 5-HT<sub>7</sub>

receptors in the hippocampus ( $p < 0.05$ ), 5-HT<sub>4</sub> receptors in the hypothalamus ( $p < 0.05$ ) and tryptophan hydroxylase 2 gene in the midbrain ( $p < 0.05$ ) in CBA mice. A significant decrease in the expression of the 5-HT<sub>2A</sub> receptor gene in the hippocampus was found in C57BL/6 animals relative to CBA animals, however, a significant effect of a short photoperiod on both lines was not observed. *Bdnf* gene expression in the cortex significantly decreased in the CBA/ Lac line in comparison with the control group ( $p < 0.05$ ). Though there were no significant changes in expression of genes encoding TrkB and p75 receptors.

*Conclusion:* An increase in indicators of depression-like behavior under shortened daylight hours was observed in mice of both strains and was accompanied by a decrease in the functional activity of the serotonin system, these changes were more significant in CBA animals. Thus, the CBA mouse strain can be proposed as an adequate model for studying the molecular and physiological mechanisms of SAD.

*Acknowledgements:* The work was supported by Russian Science Foundation, project No. 21-15-00051.

## **TrkB.T1 in the mechanisms of genetically-determined depressive-like behavior in mice**

Alsalloum M.\*, Ilchibaeva T., Tsybko A., Eremin D., Naumenko V.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *mr7salloum@gmail.com*

**Key words:** Depression, BDNF system, Serotonin system, TrkB receptor, TrkB.T1 receptor

Depression is one of the most severe mental disorders that significantly reduces the quality of life of patients. However, since available antidepressants are still not effective for all patients, an extremely urgent problem of modern neuroscience is the study of the mechanisms of depression.

Previously, we found that the level of TrkB.T1 receptor is decreased in ASC mice (mice with genetically-determined depressive-like behavior) in comparison with C57BL/6 “normal” mice. In our research, we induced TrkB.T1 overexpression in ASC mice to study its role in the mechanisms of depressive-like behavior. By using adeno associated viruses as carriers of TrkB.T1 coding gene, we induced TrkB.T1 overexpression and studied the effects of overexpression on the expression of key genes of serotonin and BDNF systems. Using behavioral tests, we assessed the effect of TrkB.T1 overexpression on behavioral level. We found that TrkB.T1 regulates aggressive-like behavior, anxiety, and depressive-like behavior in ASC mice. Considering our findings, we believe, that TrkB.T1 might serve as a drug target for new generation of antidepressants.

*Acknowledgements:* This research was supported by the Russian Foundation for Basic Research (grant No. 20-04-00253).



## Influence of administration of pro-inflammatory agents on development, individual and social behavior of BTBR mice

Ayriyants K.<sup>1\*</sup>, Mutovina A.<sup>3</sup>, Sapronova A.<sup>1</sup>, Ryabushkina Y.<sup>1</sup>, Kolesnikova M.<sup>3</sup>, Mezhlumyan E.<sup>3</sup>, Bondar N.<sup>1,3</sup>, Reshetnikov V.<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Sirius University of Science and Technology, Sochi, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

\* kaayriyants@bionet.nsc.ru

**Key words:** Autism spectrum disorders, ASD, Animal model, Autism, BTBR T(+)/tf/J, Early life adversity, Immune activation

*Motivation and Aim:* Autism spectrum disorder (ASD) is a neurodevelopmental disorder involving 1 out of every 68 children. ASD is defined by two core behavioral characteristics, namely, restricted repetitive behaviors and impaired social-communicative functioning. Accumulating evidence suggests that immune processes play a key role in the pathophysiology of ASD: extensive alterations in immune function have now been described in both children and adults with ASD, including ongoing inflammation in brain specimens, elevated pro-inflammatory cytokine profiles in the cerebrospinal fluid and blood, increased presence of brain-specific auto-antibodies and altered immune cell function. BTBR T+Itpr3tf/J (BTBR) mice provide a valuable animal model for ASD to elucidate the underlying mechanisms of these two behavioral characteristics of ASD. The BTBR mouse was reported to have significantly higher levels of serum IgG, brain IgG deposits and anti-brain IgG, elevated expression of cytokines and an increased proportion of MHC class II-expressing microglia compared to highly social C57BL/6 (B6) mice. The Th2-like immune profile of the BTBR mice and their constitutive neuroinflammation suggests that an autoimmune profile is implicated in their aberrant behaviors, as has been suggested for some humans with autism. Immune activation during early developmental stages has been proposed as a contributing factor in the pathogenesis of neuropsychiatric conditions such as obsessive-compulsive disorder, attention-deficit/hyperactivity disorder, and autism in both human and animal studies. However, the impacts of such an immune challenge on behavior and development are still not clear.

*Methods and Algorithms:* In this study, we used a model of early postnatal immune activation by the application of bacterial endotoxin lipopolysaccharide (LPS, 50 µg/kg), polyinosinic:polycytidylic acid (Poly I:C, 10 µg/kg) which simulates a viral infection, or combination (LPS + Poly I:C) to mice pups BTBR and B6 at postnatal day (PD) 3 and PD 5. The mice were tested in a battery of tasks: development of reflexes in neonatal period (negative geotaxis, grasping and cliff avoidance on 5 and 8 PD), anxiety (light-dark box test, 23 and 39 PD), social behavior (social interaction test, 24 and 39 PD) and stereotypic behavior (marble burying test, 55 PD). A blood test was performed on 24, 40 and 60 PD.

*Results:* Administration of LPS to males led to lower body weight on 5, 8 and 16 PD compared to saline (S) group. Similar effects were determined on 5 and 8 PD for LPS + Poly I:C group both males and females. Analysis of neonatal reflexes revealed a delay

on 8 PD in the formation of negative geotaxis (for LPS group) and grasping (for Poly I:C) for males compared to S group. An interesting effect was found in females: we found a delay in cliff avoidance on 5 PD for all groups compared to control. The analysis of inflammatory blood parameters did not reveal any changes between groups within each age, but changes in dynamics were found: on day 60, the administration of pro-inflammatory agents in males leads to an increase in lymphocytes relative to the control; in control females, no growth was observed after the administration of agents. The analysis of the number of red blood cells revealed an interesting result: on day 40, there was a significant increase in their number relative to other groups for both males and females. Evaluation of the effects of pro-inflammatory agents on the dynamics of platelet parameters showed a decrease in thrombocrit on day 40 and an increase by day 60 after the administration of Poly I:C for both males and females. On the 40 PD of the animals also enzyme immunoassay was performed. Analysis of 8 cytokines (IL-2, IF- $\gamma$ , TNF- $\alpha$ , IL-5, GM-CSF, IL-10, IL-1 $\beta$ , IL-4) revealed no changes other than a decrease in interleukin-5 in males treated with LPS, Poly I:C and combination compared to control. Next, we assessed the anxious behavior of the animals in the light-dark test and found that the interstrain differences were the most pronounced: BTBR mice spent significantly less time in light arms in 24, 40 and 60 than B6 mice. On 55 PD we analyzed stereotypic behavior in marble burying test and determined number of marble with more than 50 % of sinking. Male BTBR mice have shown to burrow more than B6 mice. Administration of pro-inflammatory agents resulted in decreased stereotypic behavior in male B6 mice but not in BTBR mice.

*Conclusion:* Thus, the introduction of bacterial and viral mimetics at an early age leads to various delayed effects of the immune system of mice in the juvenile, adolescent and adult periods, and also affects the development of animals.

*Acknowledgements:* This study was supported by State Budget Project FWNR-2022-0016.

# Changes in EEG spectral power during interaction with a player in a computer game

Bocharov A.<sup>1,3\*</sup>, Savostyanov A.<sup>1,2,3</sup>, Tamozhnikov S.<sup>1</sup>, Saprygin A.<sup>1,2</sup>, Rudych P.<sup>3</sup>, Zavarvin E.<sup>1,3</sup>, Merkulova E.<sup>1,3</sup>, Knyazev G.<sup>1</sup>

<sup>1</sup>Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>2</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup>Novosibirsk State University, Novosibirsk, Russia

\* bocharov@physiol.ru

**Key words:** EEG, theta rhythm, alpha rhythm, competition, cooperation

*Aim and Methods:* The aim was to investigate the oscillatory dynamics accompanying competition and cooperation with another player, as well as the individual mode during the construction of a figure in a computer game. Forty-two volunteers (24 females) between the ages of 18 and 45 participated in the study. Analysis of the differences in the spectral power of 126 channel EEG in different game modes was performed by using Loreta [1].

*Results and Conclusion:* During competition, spectral power in the theta range was higher than in cooperation mode in the anterior cingulate and medial prefrontal cortices. The increase of medial frontal theta rhythm has been known to be associated with focused attention and increased cognitive control [2–4]. It could be suggested that a greater increase in theta spectral power during competition is related to the need of mobilizing more cognitive resources and efforts in order to win the competition.

During cooperation, the spectral power of theta rhythm was lower than in individual mode in the medial prefrontal cortex and the medial prefrontal gyrus. According to Scheeringa et al. (2008) [5], default mode network (DMN) activity was negatively related to the medial frontal theta rhythm, suggesting that lower theta spectral power in medial cortical structures during cooperative mode may be related to DMN processes during interactions with another person. Alpha spectral power in the parietal and occipital cortex areas in the interaction (cooperation and competition) modes was lower compared to the individual mode. During cooperation the alpha spectral power was lower compared to the competition mode. The greatest decrease in alpha rhythm found in our study in the cooperation mode may be related to the processes of understanding the intentions of the other person.

*Acknowledgements:* The study was supported by the Russian Science Foundation (RSF) under Grant No. 22-15-00142 and budgetary funding of SRINM for basic scientific research (theme No. AAAA-A21-121011990039-2).

## References

1. Pascual-Margui R.D. Standardized low-resolution brain electromagnetic tomography (sLORETA). Technical details. *Methods Find Exp Clin Pharmacol.* 2002;2:5-12.
2. Cavanagh J.F., Frank M.J. Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci.* 2014;18(8):414-421.
3. Cohen M.X. A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci.* 2014;37(9):480-490.
4. Wascher E. et al. Frontal theta activity reflects distinct aspects of mental fatigue. *Biol Psychol.* 2014;96:57-65.
5. Scheeringa R. et al. Frontal theta EEG activity correlates negatively with the default mode network in resting state. *Int J Psychophysiol.* 2008;67(3):242-251.

## Impact of *P2RX7* and *TNF/LTA* gene polymorphisms in non-verbal intelligence in mentally healthy individuals

Enikeeva R.<sup>1,2\*</sup>, Kazantseva A.<sup>1,2,3</sup>, Davydova Yu.<sup>1,2</sup>, Mustafin R.<sup>4</sup>, Malykh S.<sup>5,6</sup>, Lobaskova M.<sup>5</sup>, Takhirova Z.<sup>2</sup>, Abushakhmina A.<sup>2</sup>, Khusnutdinova E.<sup>1,2,6</sup>

<sup>1</sup> *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre, RAS, Ufa, Russia*

<sup>2</sup> *Bashkir State University, Ufa, Russia*

<sup>3</sup> *Ufa State Petroleum Technological University, Ufa, Russia*

<sup>4</sup> *Bashkir State Medical University, Ufa, Russia*

<sup>5</sup> *Psychological Institute of Russian Academy of Education, Moscow, Russia*

<sup>6</sup> *Lomonosov Moscow State University, Moscow, Russia*

\* *enikeevarf@gmail.com*

**Key words:** non-verbal intelligence, cognitive functions, single nucleotide polymorphism (SNP), association analysis

*Motivation and Aim:* The study of cognitive functions as an important part of mental health has become challenging because of the enhanced requirements for effective intellectual activities in all fields of social functioning [1]. Nonverbal intelligence as one of cognitive abilities implies an individual's ability to use problem-solving strategies and manipulate visual information without using verbal skills [2]. The heritability (percent of individual differences accounted for genes) of intelligence in infancy is estimated at about 20 %, increases to about 40 % in middle childhood and may be as high as 80 % in adulthood [3]. Nowadays, the study of nonverbal intelligence from a biological point of view is considered the most justified. Inflammatory mediators belong to one of the promising and poorly studied biological systems in relation to non-verbal intelligence [4]. In this regard, the purpose of this work was to study the SNPs of genes involved in regulation of inflammatory mediators (*TNF/LTA rs1041981*, *rs1800629*, and *P2RX7 rs2230912*) and their role in the formation of individual differences in the level of nonverbal intelligence in mentally healthy individuals.

*Methods and Algorithms:* The study involved 1011 mentally healthy individuals (Russians, Tatars and Udmurts) from Republic of Bashkortostan (80 % women) (mean age  $19,79 \pm 1,69$  years). A written informed consent was obtained from all the participants after they were acquainted with all the procedures. SNPs genotyping was performed via real-time PCR with a fluorescent detection using the KASP method. Detection was carried out on the CFX96 DNA Analyzer (Bio-Rad, USA) with the endpoint fluorescence analysis. The level of nonverbal intelligence was measured using a black-and-white version of the Raven Progressive Matrices [5]. Statistical analysis was conducted with Plink v.1.09.

*Results:* As a result of haplotypic analysis we detected the association of the *TNF/LTA\*AA* haplotype (*rs1041981*, *rs1800629*) ( $\beta_{ST} = 1.19$ ;  $p = 0.033$ ;  $p_{perm} = 0.047$ ) with an enhanced level of nonverbal intelligence in individuals without cognitive decline. In the present study we also conducted the analysis of gene-by-environment (G×E) interactions estimating the effect of 14 socio-demographic parameters. As a result of G×E interactions we observed that sibship size significantly affected the association

of the *rs1041981* in the *TNF/LTA* gene ( $\beta = 2.08$ ;  $\beta_{ST} = 0.16$ ;  $p = 0.001$ ) and a dose-dependent effect of *P2RX7 rs2230912* minor *G*-allele on a decreased level of nonverbal intelligence among smoking individuals ( $\beta = -1.70$ ;  $\beta_{ST} = -0.10$ ;  $p = 0.022$ ).

*Conclusion:* The results obtained within the framework of this study confirm the role of the inflammatory mediator's system in the formation of differences in the level of nonverbal intelligence in mentally healthy individuals.

*Acknowledgements:* The study was partially supported by the Russian Science Foundation (project No. 17-78-30028), by the Ministry of Science and Higher Education of Russian Federation (project No. 075-15-2021-595) and by the mega-grant of the Ministry of Science and Higher Education of the Republic of Bashkortostan.

### *References*

1. Kanzafarova R.F. et al. Genetic and environmental aspects of mathematical disabilities. *Russ J Genet.* 2015;51(3):281-9.
2. Enikeeva R.F. et al. The role of inflammatory system genes in individual differences in nonverbal intelligence. *Vavilov J Genet Breed.* 2022;26(2):179-181.
3. Mustafin R.N. et al. Longitudinal genetic studies of cognitive characteristics. *Vavilov J Genet Breed.* 2020;24(1):87-95.
4. Vyshedskiy A. et al. Neurobiological Mechanisms for Nonverbal IQ Tests: Implications for Instruction of Nonverbal Children with Autism. *Res Ideas Outcomes.* 2017;3:e13239.
5. Raven J. The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol.* 2000;41(1):1-48.

## Effects of central administration of Cerebral dopamine neurotrophic factor (CDNF) on the behavior and serotonin system in the mouse brain

Eremin D.V.\*, Khotskin N.V., Naumenko V.S., Tsybko A.S.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *dima.969696@mail.ru*

**Key words:** CDNF, behavior, serotonin

*Motivation and Aim:* It is known that CDNF is able not only to protect the dopaminergic neurons of the brain, but also to participate in the control of various types of behavior. In turn, the serotonin (5-HT) system of the brain is involved in the regulation of many physiological processes and behavior and can potentially be under the influence of CDNF. Thus, the aim of this work was to study the effects of central administration of the recombinant CDNF protein on the behavior and of 5-HT in the mouse brain.

*Methods and Algorithms:* Male C56B16/J mice were injected with recombinant CDNF protein (3, 10, or 30 µg) or PBS into the left lateral ventricle of the brain. Movement activity, water and food intake, and sleep were recorded automatically in a home cage setting using the PhenoMaster for 3 days post-injection. Motor activity, anxious and exploratory behavior were measured in the open field test (OF) and elevated plus maze (EPM). Learning and memory were assessed using the Operant Wall module. 5-HT metabolism was assessed by HPLC in the midbrain raphe nuclei, frontal cortex, hypothalamus, thalamus, and hippocampus. The mRNA levels of *Htr1a*, *Htr2a*, *Htr7*, *Tph2*, *Maoa*, *Slc6a4*, *Creb1*, *Fos*, *Arc* and *Bdnf* were measured by real-time RT-PCR.

*Results:* It was found that the total distance traveled, as well as the distance and speed during the first 24 hours, were significantly increased in mice receiving 3 µg of CDNF. Also, this dose of CDNF significantly reduced the duration of sleep on the first day of testing. The total sleep time in the light phase of the day decreased in the mice of the 3 µg group. At the same time, the number of sleep episodes did not change in all experimental groups. Mice received CDNF demonstrated anti-anxiety behavior in the OF and EPM tests. CDNF at all doses significantly improved associative learning in mice compared to controls. CDNF caused significant changes in the mRNA level of *Htr1a*, *Htr2a*, *Htr7*, *Maoa*, *Creb1*, *Fos*, *Arc* genes in the different brain structures. An increase in 5-HT catabolism was shown in all brain structures studied, except for the hypothalamus, in mice treated with CDNF at all dosages.

*Conclusion:* Thus, the anti-anxiety, antidepressant and procognitive effects of a single central injection of CDNF protein were demonstrated for the first time, which were accompanied by significant changes in 5-HT metabolism, as well as changes in the transcription of many genes importantly involved in the reception and degradation of 5-HT. It has been shown that central administration of CDNF at low doses can cause disturbances in the sleep/wake cycle.

*Acknowledgements:* This work was supported by RSF grant No. 19-75-00016 and basic-research project FWNR-2022-0023.

## **Hippocampal overexpression of cerebral dopamine neurotrophic factor (CDNF) affected the behavior of mice with genetically-defined depressive-like behavior**

Kaminskaya Y.\*, Ilchibaeva T., Naumenko V., Tsybko A.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\*yanka2410@gmail.com

Dopamine neurotrophic factor (CDNF) is a promising target in the treatment of Parkinson's disease. However, the role of CDFN in the regulation of non-motor behaviour, including different kinds of psychopathologies, remains unclear. So, the aim of present work was to study the effect of CDFN overexpression in the hippocampus on the behaviour of mice with a genetic predisposition to depression-like behaviour. Overexpression of CDFN in hippocampal neurons of ASC (Antidepressant Sensitive Cataleptics) mice, characterized by genetically-defined depressive-like behaviour, was induced using an adeno-associated viral vector. Four weeks after stereotactic delivery of AAV-CDNF into dorsal hippocampus the locomotor activity, anxiety, exploratory and depressive-like behaviour were analysed in open-field (OF), elevated plus-maze (EPM) and tail-suspension (TST) tests. The Morris water maze (MWM) was used to test spatial learning and memory. In the MWM we have found significant improvements in the dynamics of learning in animals with CDFN overexpression. The animals of the experimental and control groups demonstrated approximately the same learning abilities. However, overexpression of CDFN boosted the dynamics of learning, which resulted in a decrease of distance covered, total path, and time spent even on the second day of training. At the same time CDFN overexpression failed to produce any significant changes in OF, EPM and TST suggesting that locomotor activity, anxiety, exploratory and depressive-like behavior were not affected. In addition we made an attempt to define underlying molecular basis for changes mentioned above. We have found the increase in mRNA level of unfolded protein response (UPR) marker IRE1 $\alpha$  and spliced form of XBP1 known to be spliced by IRE1 $\alpha$ . In its turn it is known that spliced XBP1 act as transcriptional regulator for different neuroplasticity-related genes.

Thus, we have shown for the first time that hippocampal CDFN improved learning of mice with genetically-defined depressive-like behavior. This effect may be linked with UPR signalling.

*Acknowledgements:* The work supported by Russian Foundation for Basic Research (grant No. 22-15-00011) and the basic research project FWNR-2022-0023.

## EEG reactions in Yakuts and labor migrants under execution of Stop-signal paradigm

Karpova A.<sup>1\*</sup>, Borisova N.V.<sup>1</sup>, Tamozhnikov S.<sup>2</sup>, Saprygin A.<sup>2,3</sup>, Savostyanov A.N.<sup>2,3</sup>

<sup>1</sup> M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

<sup>2</sup> Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* karpova74@list.ru

**Key words:** behavioural control, Stop-signal paradigm, EEG, anxiety disorder, labour migrants

*Motivation and Aim:* Executive control is a person's ability to voluntarily control their goal-directed actions. Executive control includes two mechanisms - 1) activation control, i.e. the ability to implement the action and 2) inhibitory control or the ability to suppress inadequate actions. Executive control can be violated in a number of diseases, including anxiety and stress disorders. In this study, we identified EEG correlates of inclination to anxiety disorder in labor migrants in comparison with permanent residents of Yakutia in the conditions of completing the Stop-signal paradigm.

*Participants and Methods:* The study involved 50 healthy young medical students of North-Eastern Federal University in Yakutsk (35 men, 15 women, mean age 22 years) and 50 healthy students of the same University (all men, mean age 24 years) who moved to Yakutia from the southern regions (Asia, Africa, South America). Migrants were examined twice – immediately after moving to Yakutia and six months after moving. The examination included EEG registration (64 channels, Amplificatory NeoRec, Russia) and filling out questionnaires for susceptibility to depression and anxiety disorder.

*Results:* During the first examination, migrants showed an increased level of predisposition to anxiety disorder in comparison with the Yakuts, which significantly decreased after six months of adaptation. Susceptibility to anxiety disorder correlated with the amplitude of desynchronization in the alpha frequency band when performing the SSP activation tasks and the amplitude of delta synchronization when performing the SSP inhibitory tasks.

*Conclusion:* EEG correlates under execution of SSP can be used as markers of inclination to anxiety disorder. Besides, SSP can be used as a method for diagnostic of success of adaptation to climatic changing among migrants.

*Acknowledgements:* The study is supported by the grant No. 18-415-140021 of Russian Foundation for Basic Research (RFBR).



## The use of genetic risk score approach to predict higher depression risk in young adults

Kazantseva A.<sup>1,2,3\*</sup>, Davydova Yu.<sup>1,2</sup>, Enikeeva R.<sup>1,2</sup>, Mustafin R.<sup>4</sup>, Malykh S.<sup>5,6</sup>, Lobaskova M.<sup>5</sup>, Kosareva A.<sup>2</sup>, Mikhailova A.Y.<sup>2</sup>, Khusnutdinova E.<sup>1,2,6</sup>

<sup>1</sup> Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre, RAS, Ufa, Russia

<sup>2</sup> Bashkir State University, Ufa, Russia

<sup>3</sup> Ufa State Petroleum Technological University, Ufa, Russia

<sup>4</sup> Bashkir State Medical University, Ufa, Russia

<sup>5</sup> Psychological Institute of Russian Academy of Education, Moscow, Russia

<sup>6</sup> Lomonosov Moscow State University, Moscow, Russia

\* Kazantsa@mail.ru

**Key words:** genetic risk score, depression, neurotransmitter system, hypothalamic-pituitary-adrenal system, inflammatory response genes, miRNA, GWAS

*Motivation and Aim:* Various genetic risk score (GRS) studies reported a genetic overlap between different affective psychopathologies. However, GRS studies of depression simultaneously based on both genetic and environmental factors remain scarce. Suggesting that depression should be viewed on a continuum [1], individuals with elevated depression scores in early adulthood may be at risk for developing clinical forms of depression later in ontogenesis [2]. Since previous studies implicated deregulation in monoaminergic [3, 4], HPA-axis [5, 6], inflammatory system [7], functioning in liability to depression and related phenotypes, the present study aimed to examine whether a GRS of depression-related traits based on genetic and environmental parameters predicts higher depression risk in the testing sample of age- and ethnicity-similar individuals.

*Methods and Algorithms:* The study included 1067 mentally healthy individuals (79 % women;  $18.85 \pm 4.85$  years) - students at the Universities of Russia of Caucasian origin (357 Russians, 341 Tatars, 234 Udmurts, and 135 of mixed ethnicity). The sample was divided into training (N = 767) and testing sets (N = 300). Depression score was assessed via self-report Beck Depression Inventory (BDI), while style of parental rearing – via Parental Bonding Instrument (PBI). Genotyping of examined SNPs in the *SLC6A4* (rs4795541, rs1042173), *MAOB* (rs6651806), *DRD4* (rs1800955), *AVPR1A* (rs1042615, rs3803107), *AVPR1B* (rs33911258, rs28632197), *OXTR* (rs53576, rs237911, rs7632287, rs2254298, rs2228485, rs13316193), *CRP* (rs3093077), *IL1B* (rs16944), *TNF* (rs1041981), *MIR124* (rs531564), *MIR135* (rs10459194), *MIR2113* (rs10457441), *MRAS* (rs9818870), *HTR1B* (rs13212041), *HTR2A* (rs7322347), *FYN* (rs2148710), *PCLO* (rs2715157) genes was performed using PCR-based KASP genotyping. Statistical analysis included multiple linear regressions with genotypes and environmental parameters as predictors for the training sample. GRS assessment at different P-values ( $P < 0.05$ ,  $P < 0.1$ ,  $P < 0.2$ ,  $P < 0.3$ ) was conducted in the testing sample with linear regression (PLINK v.1.09). The study was approved by the Biological Ethics Committee at the Institute of Biochemistry and Genetics.

*Results:* A series of linear regression analysis of depression score as dependent variable and SNPs and environmental factors as independent predictors were conducted. We succeeded to identify the best GRS model at  $P < 0.1$  significance level ( $P = 0.001$ ,  $\beta = 1.099$ ) in the testing sample, which explained 13.4 % of variance in depression score in mentally healthy young adults. This GRS model included the minor alleles of *AVPR1B* rs28632197 ( $\beta = 1.23$ ;  $P = 0.010$ ), *MRAS* rs9818870 ( $\beta = 0.90$ ;  $P = 0.089$ ), *PCLO* rs2715157 ( $\beta = 0.81$ ;  $P = 0.025$ ), *AVPR1A* rs3803107 ( $\beta = 0.96$ ;  $P = 0.055$ ), *MIR135A2* rs10459194 ( $\beta = 0.91$ ;  $P = 0.020$ ) gene variants together with such environmental factors as a decreased income level ( $\beta = 4.02$ ;  $P = 0.007$ ), the presence of chronic disorders ( $\beta = 1.68$ ;  $P = 0.049$ ), low level of paternal care ( $\beta = -2.03$ ;  $P = 0.013$ ), paternal overprotection ( $\beta = 2.53$ ;  $P = 0.002$ ).

*Conclusion:* Our findings established the best GRS model, which simultaneously assessed genetic and environmental factors and successfully predicted an increased depression level in the testing sample of the similar-age individuals from the same geographic region of Russia. Notably, the GRS model confirmed the involvement of genes encoding for arginine vasopressin receptors, hsa-miR-135a2 together with its target *MRAS* in individual variance in depression-related traits. In addition, the data obtained confirmed a significant effect of parental style of rearing and income level on individual differences in depression-related traits.

*Acknowledgements:* The study was partially supported by the Russian Science Foundation (project No. 17-78-30028), by the Ministry of Science and Higher Education of Russian Federation (project No. 075-15-2021-595) and by the mega-grant of the Ministry of Science and Higher Education of the Republic of Bashkortostan.

## References

1. Kwong A.S.F. et al. Genetic and Environmental Risk Factors Associated With Trajectories of Depression Symptoms From Adolescence to Young Adulthood. *JAMA Netw Open*. 2019;2(6):e196587.
2. Davydova Yu.D. et al. Genetic basis of depressive disorders. *Vavilov J Genet Breed*. 2019;23(4):465-472.
3. Noskova T.G., et al. Association of several polymorphic loci of serotonergic genes with unipolar depression. *Russ J Genet*. 2009;45(6):842-848.
4. Kazantseva A.V. et al. Role of dopamine transporter gene (*DAT1*) polymorphisms in personality traits variation. *Russ J Genet*. 2009;45(8):974-980.
5. Kazantseva A.V. et al. *AVPR1A* main effect and *OXTR*-by-environment interplay in individual differences in depression level. *Heliyon*. 2020;6(10): e05240.
6. Kazantseva A.V. et al. Arginine-vasopressin receptor gene (*AVPR1A*, *AVPR1B*) polymorphisms and their relation to personality traits. *Russ J Genet*. 2014;50(3):298-307.
7. Davydova Yu.D. et al. The role of presynaptic cytomatrix protein Piccolo and inflammatory genes in the development of depressive behavior. *Medical Genetics*. 2020;19(4):36-38.

# Analysis of behavioral and EEG responses given the performing tasks to control attention and behavior in children with emotional instability

Klemeshova D.I.<sup>1\*</sup>, Dashkevich G.E.<sup>3</sup>, Tamozhnikov S.S.<sup>1</sup>, Saprygin A.E.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Neuroscience and Medicine, Novosibirsk, Russia*

<sup>3</sup> *Neuropsychologist in private practice*

\* *minoko@mail.ru*

**Key words:** stop signal paradigm, EEG, executive function, emotional instability, Stroop effect, P300

Executive brain functions are a system of cognitive processes in the brain, including cognitive flexibility, inhibition (the ability to control impulsive or automatic reactions and generate rational ones), self-control, activity initiation, multitasking, monitoring, planning, working memory, decision making, abstract thinking, and the ability to switch attention. These functions are involved in practically every activity, learning and memory processes and, when impaired, can affect the correct regulation of other cognitive processes. Disruption on executive functions causes disorganized behavior, lack of adaptation to the environment, personality and mood changes. Also control of executive functions directs selective attention to the right stimuli. The P300 component in healthy people is detected during cognitive tasks and is associated with the involvement of attention, especially to new or meaningful stimuli. In an EEG study in the "oddball" paradigm, when polyphonic and monophonic stimuli were presented, an increase in the P300 amplitude on a non-target monophonic stimulus was observed in the central and parietal areas of the brain in children with various emotional disturbances. The ability to concentrate attention on the stimulus was normal, but the ability to ignore the non-target and irrelevant signal was impaired. The sound structure of the signal is likely to be important for the ability to isolate the desired signal.

*Materials and Methods:* In our work, we examined a sample of 12 children with diagnosed emotional instability compared to 12 healthy control subjects (mean age in both samples  $8 \pm 2$  years, samples balanced by gender).

All children took part in EEG study while they performed tests for directed attention (monophonemic and polyphonemic odd-ball), motor control (stop-signal paradigm), and cognitive control (Stroop test). EEG was recorded on an amplifier from Brain Products, Germany, using 64 electrodes arranged according to the international 10–10 % scheme. In addition, the children's parents completed a set of questionnaires aimed at assessing the child's ability to self-monitor behavior. The amplitude of the P300 peak in the frontal and parietal cortices recorded in different behavioral tests was chosen as neurophysiological markers.

*Results:* The most pronounced differences between the groups of children with emotional instability and healthy control children were found for the monophonemic odd-ball test. In this test, in response to the target stimulus, the amplitude of the P300 peak in children with the diagnosis did not differ from that in healthy children, but differences in this index were found for the response to the non-target stimulus. In

children with emotional instability, the amplitude of the P300 peak in the medial frontal and medial parietal cortex was significantly higher compared to the group of healthy children. This result can be interpreted as a marker of impaired selective attention in children with emotional instability. Healthy children responded more strongly to the target signal and ignored the non-target signal, whereas children with emotional control disorders responded equally to both types of signals.

*Conclusion:* Analysis of the P300 peak amplitude under the conditions of the selective attention test revealed differences in cognitive signal processing between healthy children and children with emotional instability. For children with emotional instability, impairments related to the evaluation of signals as non-targeted were diagnosed, but no impairments related to the evaluation of target signals were diagnosed.

*Acknowledgements:* The work of D.I. Klemeshova and A.E. Saprygin was also supported from the foundation budgetary project of the ICG SB RAS No. 0259-2021-0009.

# Actin microfilaments in the submembrane zone of olfactory receptor cells as a potential factor in the regulation of their sensitivity

Klimenkov I.V.\*, Sudakov N.P.

*Limnological Institute, SB RAS, Irkutsk, Russia*

\* [iklimen@mail.ru](mailto:iklimen@mail.ru)

**Key words:** olfaction, chemoreception, neurogenesis

*Motivation and Aim:* It is known that olfactory sensitivity to odorants correlates with area, width of sensory epithelium, quantity and maturity of the olfactory sensory neurons (OSNs) which has specific receptors with signal transduction system [1]. Previously, we have found that when the immature OSNs makes contact with its environment, the dendritic terminal develops a wide actin layer, inside which a pore is formed. It is assumed that the functional receptors of odorants generate across this pore the first intracellular signal from environmental water-soluble odorants [2]. According to our ideas, the wide layer of actin microfilaments serves as barrier for biochemical signals transduction inside the cell. We assume that actin microfilament polymerisation/depolymerisation in submembrane layer of dendrite terminal may be characteristic both for immature OSNs, and mature OSNs. It may provide sensitivity regulation to different odorant concentrations. Namely, in the absence of odorants, grade of actin polymerisation should be minimal, but in the excess of odorant concentration, amount of polymerised actin increased.

*Methods and Algorithms:* It was studied the peculiarities of actin microfilament development in the mature OSNs dendrite terminales of Baikalian fishes *Cottocomephorus inermis* (Jakowlew, 1890), *Cottocomephorus grewingkii* (Dybowski, 1874), *Comephorus baicalensis* Pallas, 1776 (Cottidae) and Wistar rats. Olfactory organs were fixed in paraformaldehyde, permeabilized in Triton-X100, stained with Phalloidin-FITC, and analysed on LSM 710 laser confocal microscope (Zeiss Germany).

*Results:* It was revealed that in all analyzed animals the observed actin-belt network consists of two types of regularly arranged meshes of different diameters: small meshes related to receptor cells and large meshes related to the apexes of supporting cells. Moreover the submembrane actin layer in the apical part mature OSNs has different grade of polymerisation. It is possible this reflects different level of their functional state which depends on specificity and mode of odorant action.

*Conclusion:* Our studies indirectly shows that regulation of sensitivity to odorants may be realised both biochemical signal transduction level, and by structural rearrangements of actin cytoskeleton of submembrane area of dendrite terminal. Thus the process of actin polymerization/polymerization in the apical part of receptor cells may be involved in primary process of chemoreception of OSNs by supporting the adequate level of their stimulation during the variable modes of exposure to odorants.

## References

1. Doty R.L. Handbook of Olfaction and Gustation. Third Edition. Book. Doty, R.L. (eds.). New York: Wiley-Liss, 2015.
2. Klimenkov I.V. et al. Rearrangement of Actin Microfilaments in the Development of Olfactory Receptor Cells in Fish. *Sci Rep.* 2018;8:3692.

## Approach-avoidance conflicts

Liou M.

*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*

\* *michelleliou@gmail.com*

**Key words:** anterior hippocampus, BOLD responses

*Motivation and Aim:* Area anterior hippocampus (aHC; homolog of rodent ventral hippocampus) is a common target in experiments on approach-avoidance conflicts. A recent report hypothesized that decreased activity in Area aHC reflects the suppression of aversive and emotional states, which would otherwise compete with goal-directed or motivational behavior [1]. Stimulation of Area aHC was shown to produce sustained increases in heart rate and blood pressure in normal rats, whereas bilateral inhibition of this area enhanced parasympathetic responses with no changes in sympathetic responses [2, 3]. Given the existing work, this study demonstrated that an abrupt instruction to open the eyes while the subject was resting quietly with eyes closed could probe brain reactivity to an aversive (or novel) stimulus.

*Methods and Algorithms:* A total of 49 young adults (26 females, mean age  $22.94 \pm 3.18$ ) underwent 8-min resting-state fMRI (rs-fMRI) scans under 4-min eyes-closed (EC) followed by 4-min eyes-open (EO) conditions. During the experiment, each subject received exclusively auditory instructions to close or open the eyes. Image data were preprocessed by including motion correction, slice timing, linear detrending, and spatial transformation to the MNI-152 template. Standardized intraclass correlation (ICC) was applied to masking out voxels showing high degrees of synchronicity among subjects. The size of ICC was mainly contributed by temporal responses to the EC and EO instructions. The supra-threshold maps were co-registered to the JuBrain Cytoarchitectonic Atlas available at <http://www.fz-juelich.de/inm/inm-7/tools> for topographical definition of area-specific effects [4].

*Results:* Under the EC-then-EO paradigm, the auditory cortex presented a sharp increase in activity followed by a short-term decrease (the sharp increase could be partially caused by head motion). Area aHC presented an immediate increase followed by a short-term decrease in activity after the onset of the eyes-open instruction. The effect size of decreased activity was far larger in Area aHC than in the auditory cortex. We also observed an immediate increase followed by a short-term decrease in activity after EO onset in the caudal premotor cortex (6d1, 6d3, and 6mc-SMA), anterior parietal cortex (1, 2, 3a, 3b), and motor cortex (4a, and 4b), echoing the responses observed in Area aHC.

*Conclusion:* Decreased activity in Area aHC could release subjects from behavioral inhibition in favor of goal-directed behavior. Fear and stress have been shown to induce rapid and dramatic alterations in the respiratory effort, cardiac output, and blood pressure [5]. When reacting to novel (aversive) stimuli, the duration of an orienting response is normally shorter than that of a freezing response (specific to threatening stimuli), both of which occur prior to the fight-or-flight response and engage motor immobility as a parasympathetic brake on the otherwise active motor system [6, 7]. We hypothesize that the EO auditory instruction induced an immediate increase followed by a short-term and sharp decrease in activity in brain areas that react to auditory stimulation and participate

in the suppressed fight-or-flight response aiming to sustain on-going behavior (i.e., remaining motionless with eyes fixed on a central crosshair).

*Acknowledgements:* This study was supported by Grant 110-2410-H-001-046 from the Ministry of Science and Technology, Taiwan.

### *References*

1. Yoshida K., Drew M.R., Mimura M., Tanaka K.F. Serotonin-mediated inhibition of ventral hippocampus is required for sustained goal-directed behavior. *Nat Neurosci.* 2019;22(5):770-777.
2. Kuntze L.B., Ferreira-Junior N.C., Lagatta D.C., Resstel L.B.M. Ventral hippocampus modulates bradycardic response to peripheral chemoreflex activation in awake rats. *Exp Physiol.* 2016;101(4):482-93.
3. Ajayi I.E., McGovern A.E., Driessen A.K., Kerr N.F., Mills P.C., Mazzone S.B. Hippocampal modulation of cardiorespiratory function. *Respir Physiol Neurobiol.* 2018;252:18-27.
4. Eickhoff S.B., Stephan K.E., Mohlberg H., Grefkes C., Fink G.R., Amunts K. et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage.* 2005;25(4):1325-35.
5. Guyenet P.G. The sympathetic control of blood pressure. *Nat Rev Neurosci.* 2006;7(5):335-46.
6. Hagens A.M.A., Oitzl M., Roelofs K. Updating freeze: aligning animal and human research. *Neurosci Biobehav Rev.* 2014;47:165-76.
7. Roelofs K. Freeze for action: neurobiological mechanisms in animal and human freezing. *Philos Trans Royal Soc B: Bio Sci.* 2017;372(1718):20160206.

## Correlation of resting-state EEG indicators with personal and genetic characteristics of subjects living in different socio-climatic conditions

Milakhina N.<sup>1\*</sup>, Saprygin A.<sup>2,3</sup>, Tamozhnikov S.<sup>3</sup>, Bazovkina D.<sup>1</sup>, Savostyanov A.<sup>1,2,3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Neuroscience and Medicine, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

\* [tashamilka@mail.ru](mailto:tashamilka@mail.ru)

**Key words:** EEG, mental disorder, anxiety, depression, polymorphism 5-HTTLPR

*Motivation and Aim:* EEG is a universal method for the study of functional activity of the human brain. EEG rhythms reflect the individual, genetic and behavioral characteristics of individuals, as well as the influence of climatic conditions [1]. As you know, mental disorders have a biopsychosocial model [2]. To date, depression and anxiety disorders are the leading disorders of affective disorders according to WHO. The issue of the relationship of depression-prone EEG markers to genetic and environmental factors is still relevant.

*Methods and Algorithms:* The questionnaires “Big Five Factor Markers”, Goldberg Big Five IPIP Markers [3] and Eysenck Personality Questionnaire were used to measure personality within the five-factor model. We also used well-known questionnaires to measure personal anxiety (State Trait Anxiety Inventory). Emotional intelligence was measured using Barchard’s questionnaire [4]. Two questionnaires were used to assess the manifestations of individualistic and collectivist tendencies: the Singelis questionnaire on the collective and independent self-concept (Self Construal scale) [5] and the second, which selectively measures affiliative tendencies to next of kin or close relatives (RISC, The Relational-Independent Self-construal). Beck’s questionnaire was used to assess the severity of depressive symptoms [6]. EEG electrodes were placed on the head according to the international scheme 10–10 %. In adult participants in Novosibirsk, resting state EEG was recorded through 128 channels, in children through 64 channels. For participants from Tuva and Yakutia, EEG was recorded through 64 channels. The Cz channel was always used as a reference electrode, and the AFpz channel was used as a ground electrode. The resistance under the electrode was measured before the start of EEG recording and was less than 5 k $\Omega$ . Genomic DNA was isolated using a kit to extract DNA from samples of buccal epithelium («Biosilica», Russia). The frequencies of the short (S-) and long (L-) alleles of the serotonin transporter gene were determined using allele-specific PCR using 5'-ggcgttgctctgagc-3 or 5'-gaggtgcacat3 primers (Lesch et 1996) and the genomic DNA of the subjects as a matrix. The allele L- was also subdivided into La- and Lg-. In further analysis, the allele Lg- was treated as S-. PCR products were separated by electrophoresis in 3 % agarose gel.

*Results:* There was a different relationship of anxiety with the spectral power in the alpha-2 rhythm range in the frontal cortex at open eyes in a dormant state for men and women. The difference between men and women was observed in the group of high anxiety subjects and was not observed for the group with low and medium anxiety: the



high anxiety women have an increased spectral power of alpha-2 rhythm, for highly anxious men - reduced spectral power alpha-2-rhythm. We also found differences in the relative spectral power of alpha 2 in the parietal cortex in the conditions of functional penis with open eyes between urban residents and those living in the village, which may indirectly indicate that in Siberian regions, urban residents are more at risk of developing depressive symptoms than villagers. We also noted the reliable effect of the local influence on the relative spectral power of delta rhythm under opened eyes in the frontal cortex: it was more pronounced in the city residents than in the village, which may indirectly indicate that that citizens have a greater ability to suppress behavior. This effect was also positively correlated with gender (women had a higher effect than men) and consciousness, which may indicate that women are more likely to suppress impulsive behavior than men. The Goldberg neurotism effect was different in alpha-2 in open eyes and delta in closed eyes for the entire sample of subjects. In the case of alpha-2: the higher the neuroticism, the lower the alpha-2 ( $F(1.431) = 4,15, p = 0.042, \eta^2 = 0,010$ ). In the case of delta: the higher the neurotism, the lower the delta. When considering the polymorphism 5-5-HTTLPR (SS> LL+ LS - more so in the Mongoloid group) effect on brain activity in resting-state conditions with closed eyes, the brain activity in the delta range was clearly expressed in the superior frontal gyrus and the frontal cortex, the right gyrus and parietal cortex were involved in the theta range. In Caucasians (LL+LS > SS), the polymorphism 5-HTTLPR effect was strongly expressed in alpha 1 rhythm in the superior frontal gyrus and frontal lobe, and in alpha 2 rhythm in the postcentral gyrus and parietal cortex.

*Conclusion:* It can be proposed that the S- allele of the serotonin transporter is associated with better adaptability to life conditions in high-risk environments, which is indirectly confirmed by the frequency of this allele in various ethno-regional groups. The question of spectral-power interaction and the psychological characteristics of different ethnic groups, as well as the polymorphism of the serotonin transporter and psychological characteristics, is still open and requires further detailed study.

*Acknowledgements:* The work of N.S. Milakhina, A.E. Saprygin, and A.N. Savostyanov was supported from the funds of the budget project No. 0259-2021-0009 at the Institute of Cytology and Genetics of the SB RAS.

### References

1. Knyazev G., Merkulova E., Savostyanov A., Bocharov A. Effect of Cultural Priming on Social Behavior and EEG Correlates of Self-Processing. *Front Behav Neurosci.* 2018;12:236.
2. Tripathi A., Das A., Kar S. Biopsychosocial model in contemporary psychiatry: current validity and future prospects. *Indian J Psychol Med.* 2019;41(6):582-585.
3. Knyazev G., Mitrofanova L., Bocharov A. Validization of russian version of Goldberg's "Big-Five factor markers" inventory. *Psycholog J.* 2010;31(5):100-110.
4. Knyazev G., Mitrofanova L, Razumnikova O., Barchard K. Adaptation of russian language version of K. Barchard's "Emotional intelligence questionnaire". *Psycholog J.* 2012;33(4):112-120.
5. Dorosheva A., Knyazev G., Kornienko O. Validation of two russian language version of self-conception questionnaires. *Psycholog J.* 2016;37(3):99-112.
6. Beck A. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation, 1996.

## Investigation of influence of the gut microbial composition associated with colitis on mice behavior

Morozova M.V.<sup>1,2\*</sup>, Borisova M.A.<sup>3</sup>, Snytnikova O.A.<sup>4</sup>, Tsentlovich Y. P.<sup>4,5</sup>, Kozhevnikova E.N.<sup>1, 2, 6, 7</sup>

<sup>1</sup>Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>2</sup>Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

<sup>3</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup>International Tomography Center, SB RAS, Novosibirsk, Russia

<sup>5</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>6</sup>Novosibirsk State Agrarian University, Novosibirsk, Russia

<sup>7</sup>Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Novosibirsk, Russia

\* morozova.maryana@mail.ru

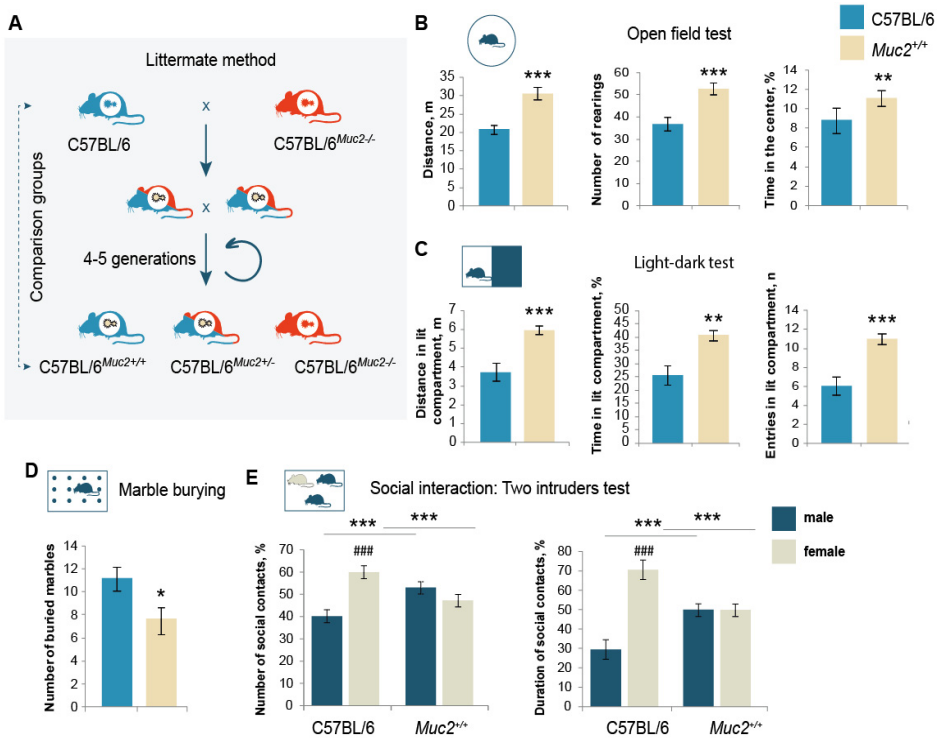
**Key words:** Inflammatory bowel diseases, gut-brain axis, intestinal microbiota, behavior, glycine

*Motivation and Aim:* Inflammatory bowel diseases (IBD) are chronic and relapsing inflammatory disorders of the gastrointestinal tract with complex etiology and no strategies for complete cure. IBD are often complicated by mental disorders like anxiety and depression, indicating substantial shifts in the gut-brain axis. However, the mechanisms connecting IBD to mental diseases are still under debate.

*Methods and Algorithms:* Here we use *Muc2* mutant mouse model of chronic colitis to uncouple the effects of the intestinal microbiota on host behavior from chronic inflammation in the gut.

*Results:* *Muc2* mutant male mice exhibit high exploratory activity, reduced anxiety-related behaviors, impaired sensorimotor gating, and altered social preference towards males and females. Microbial transfer to wild-type mice via littermate co-housing shows that colitis-associated microbiota rather than inflammation *per se* defines behavioral features in *Muc2* colitis model (Fig. 1). Metagenomic profiling and combination of antibiotic treatments revealed that bacterial species *Akkermansia muciniphila* is associated with the behavioral phenotype in mutants, and that its intestinal abundance correlates with social preference towards males. Metabolomic analysis together with pharmacological inhibition of Gly and NMDA receptors helped us to determine that brain glycine is responsible for the behavioral phenotype in *Muc2* mice. Blood and brain metabolic profiles suggest that microbiota-dependent changes in choline metabolism might be involved in regulation of central glycine neurotransmission.

*Conclusion:* Taken together, our data demonstrates that colitis-associated microbiota controls anxiety, sensorimotor gating and social behavior via metabolic regulation of the brain glycinergic system, providing new venues to combat neurological complications of IBD.



**Fig. 1.** Behavioral traits are associated with microbiota of *Muc2*<sup>-/-</sup> mice. (A) The scheme of the littermate co-housing method. (B) Open field test ( $n = 18-20$ ). Distance:  $t = -3.895$ ,  $p < 0.001$ ; rearing:  $t = -4.97$ ,  $p < 0.001$ , Student's  $t$ -test; time in the center:  $Z = -2.997$ ,  $p < 0.001$ , Mann-Whitney  $u$  test. (C) Light-dark test ( $n = 17-20$ ). Distance:  $t = 4.019$ ,  $p < 0.001$ ; time:  $t = -3.584$ ,  $p < 0.001$ ; entries:  $t = -4.26$ ,  $p < 0.001$ , Student's  $t$ -test. (D) Marble burying test ( $n = 18-20$ ). Number of buried marbles:  $Z = 2.22$ ,  $p = 0.026$ ; Mann-Whitney  $u$  test. (E)  $Z = 2.22$ ,  $p = 0.026$ ; Mann-Whitney  $u$  test. (F) Two intruders test ( $n = 10-11$ ). Two-way ANOVA revealed a statistically significant interaction between the resident group and the intruder gender (number of contacts:  $F(1, 38) = 15.490$ ,  $p < 0.001$ ; duration:  $F(1, 38) = 28.306$ ,  $p < 0.001$ ). Number of contacts (C57BL/6, male vs. female):  $p < 0.001$ , Fisher's LSD test. Number of contacts with a male/female intruder vs. C57BL/6:  $p = 0.008$ , Fisher's LSD test; duration of contact with a male/female intruder vs. C57BL/6:  $p < 0.001$ , Fisher's LSD test. \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , vs. C57BL/6. #### =  $p < 0.001$ , male vs. female

**Acknowledgements:** This study was supported by the Russian Science Foundation (RSF) Grant No. 20-74-10022.

## The emergence and evolution of living systems compensating for entropy processes

Ratushnyak A.S.\* , **Zapara T.A.**, Proskura A.L., Sklyarov A.N.

*Federal Research Center for Information and Computing Technologies, Novosibirsk, Russia*

\* *ratushniak.alex@gmail.com*

**Key words:** modeling emergence of living systems, entropy, negentropy, molecular machines

Understanding the principles and mechanisms of the functioning of biological systems and the brain, in particular, is necessary both for developing methods and approaches to the prevention, diagnosis and correction of neuropathologies, and for advancing in the field of creating cognitive systems. However, progress in the field of understanding the principles and mechanisms of the functioning of biological systems has not yet had fundamental success. To achieve significant results, it is necessary to search for conceptually new approaches. The construction of new concepts can be based on an understanding of the basic principles underlying the emergence and evolution of biological functional systems. It is important to define the very concept of living systems. There is uncertainty in understanding the term itself. Now there are several hundred definitions of such systems, essentially reducing to the concept of “self-reproduction with variations”. NASA, for example, defines life as “a self-sustaining chemical system capable of undergoing Darwinian evolution”. At the same time, not self-reproducing, but of course, living organisms are known – neurons, erythrocytes, etc. Naturally, scientific analysis does not allow us to reasonably assert that even the simplest systems capable of self-reproduction could spontaneously form in the foreseeable time interval. Another assumption about the basic function of any biosystem is the concept that their emergence and existence are based on the processes of counteracting entropy. Self-reproduction, at the same time, is a derivative that arose in the process of evolution, but not at all their main property. On the basis of such a concept, one can try to create a model of the primary prebiont and consider possible logical schemes for its functioning and evolution. One can proceed from the idea of living systems as molecular, digital, logical (with non-binary logic) machines, the main function of which is to create and maintain a certain order (homeostasis, complex information and energy) based on forecasting using previously obtained (recorded) information. In this case, the mechanism of emergence, spontaneous formation of extremely simple molecular machines without an auto-replication complex, but having the features of living systems that maintain their stability in an entropic environment, becomes more understandable. The elements of such machines can be rather simple organic molecules that perform the functions of receptors of external influences, logical elements that process, fix information and act as effectors. In this case, logical operations are reduced to fixing the relationship between receivable events in the environment and the reactions of effector molecules, which make it possible to avoid events that increase the entropy of such a molecular system. Environmental factors interact with molecular structures that act as P1 and P2 receptors. The influences transformed by these structures are transmitted to the molecular logical structures that perform the selection of time-related signals and the reaction to them by turning on the effector, which allows optimizing the loss/gain of

energy and information by the system. At the initial moment, the effector turns on only under the action of a factor that changes the information-energy state. In this case, the previous factor (factors) are remembered as signal ones. At the first stage, the basis of the interconnection of signals is not of fundamental importance – association, correlation or cause-and-effect relationships. The connection of two (or more) signals is remembered. When the next pair of interconnected signals occurs in the environment, the effector turns on already under the action of the first factor – the signal. At the same time, the system avoids losses or optimizes acquisitions. Spontaneous emergence of such molecular constructs consisting of several simple biomolecules seems quite probable. And the stability they receive in an entropic environment can lead to their accumulation. The fusion of such complexes differing in the set of receptors and effectors contributed to the expansion of the possibilities of avoiding damage and the acquisition of additional information and energy. At the same time, natural selection can be based not on the basis of reproducing organisms, but when primary agents appear after fusion and fragmentation with the selection of the most relevant molecular structures. For example, sets of receptors for associative factors, times and directions for the implementation of logical processes, the volume and depth of “memory” about a changing environment. On the basis of such natural scientific ideas about biological functional systems and their primary prebiotic prototypes, a model of a logical agent has been developed. A model of the medium has been developed in which there are signals of various modalities and agents can be spontaneously formed from the elements (molecules) present. In the process of modeling in the environment, the number of elements (having logical functions) that determine the probability of the emergence and persistence of the agent is optimized. A set of conditions, coefficients of their values and rules for the interaction of agents and the environment, stabilizing the system as a whole, is determined.

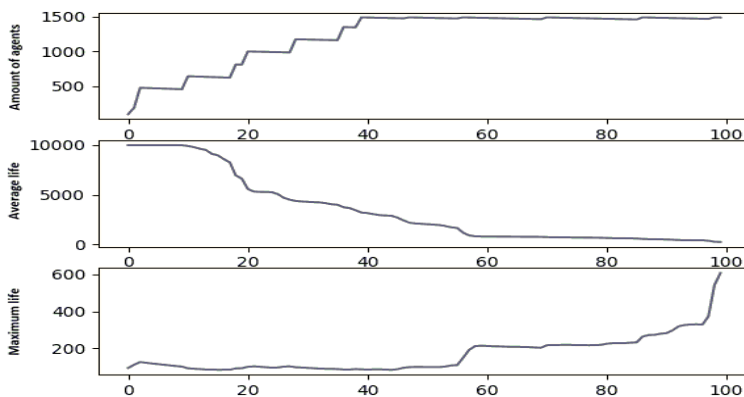


Figure 1 shows an example of the dynamics of adaptation of negentropic agents for given parameters of the environment and the agent. When a pool of agents is formed, their interaction with the environment leads to the selection of the most stable ones, their number stabilizes and then remains constant until the end of the analysis cycle. It is possible to increase the complexity of the model of biologically inspired cognitive systems by combining them with a network or matrix approach up to the limits of the capabilities of the tools.

# Behavioral and neurophysiological study of subject's personality traits under recognition of sentences about self and others

Saprygin A.<sup>1,2</sup>, Lebedkin D.<sup>3</sup>, Savostyanov A.<sup>1,2,3</sup>, Vergounov E.<sup>2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

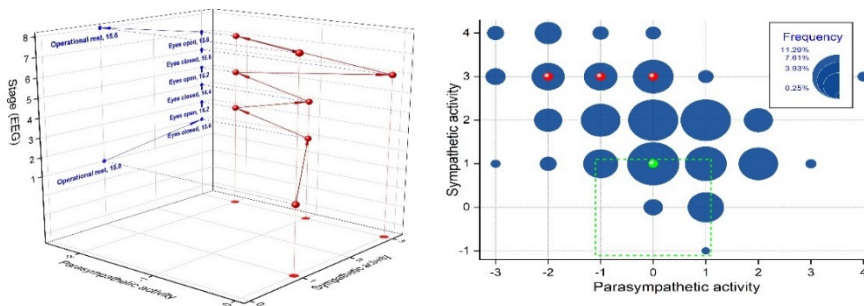
<sup>2</sup> Scientific-Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

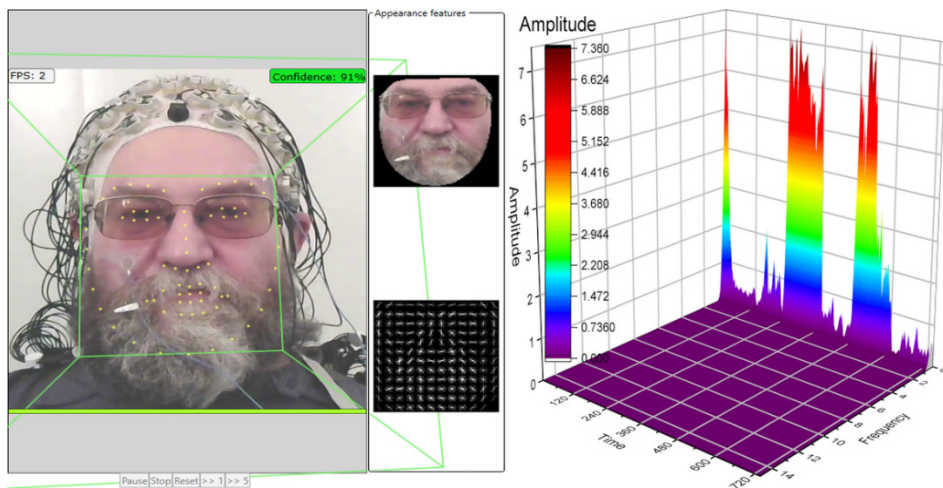
**Key words:** Speech Comprehension Task, EEG, Sympathovagal activity, Electrocardiogram Derived Respiration, Mental Chronometry, Face Image Analysis, Data Science, Intelligent Database, Personality Theory

*Motivation and Aim:* One of the key problems of modern cognitive research is the development of the objective methods for test task assessment and correlation of their results with participant's personality traits. This problem is particularly acute during the study of behavioral reactions while doing neurolinguistic tasks under different conditions [1]. Thus follow: the problem of composing implicit test tasks [1, 2] (in which a participant cannot deliberately give misleading responses) and the problem of using instrumental and mathematical methods of studying participant's reactions [3]. This work is aimed at the second problem.

*Methods and Algorithms:* Study of behavioral responses to emotional phrases, describing aggression and anxiety of either participant themselves or unrelated strangers, was conducted using EEG, HRV (with EDR estimations; Fig. 1), Face Image Analysis (Fig. 2), Mental Chronometry. Analysis employed Data Science approaches (including Permutation Approach).



**Fig. 1.** Sympathovagal activity during resting state EEG of a student participant. Left: “individual trajectory” Sympathovagal activity by stages of a resting state EEG (blue color, numbers – EDR/min; Electrocardiogram Derived Respiration.); stages: 1 (before “start” command); 2, 4, 6 (eyes closed); 3, 5, 7 (eyes open), 8 (after “stop” command). Right: Sympathovagal activity of a participant and in the statistical population (1000 sampling units), based on Permutation Approach based on the results of 178 students of this age. Sympathovagal activity: –3 (significant decrease), –2 (moderate), –1 (slight), 0 (balance); +1 (slight increase), +2 (moderate), +3 (significant). Blue circles: share in the statistical population; green ball – Sympathovagal activity during stages 1 and 8, red balls – stages 2–7; area outlined by green dotted line – Sympathovagal balance range



**Fig. 2.** Face Image Analysis during resting state EEG by one of the authors (left), STFT-analysis of his face muscles in time (right). Deliberately chosen for additional interference (sunshades, beard)  
 1 point on the timescale equals 120 sec, on the frequency scale – 2 Hz  
 0–120, 240–360 and 480–600 sec – eyes closed (high activity of face muscles)  
 120–240, 360–480 and 600–720 sec – eyes open (low activity of face muscles)

*Results:* Developed an Intelligent Database for neurolinguistic tasks that allows to conduct analyses based on features of the data of instrumental methods itself. New results, obtained with the help of the Intelligent Database, have shown high efficiency of a complex of instrumental methods, adapted by authors for neurolinguistic tasks.

*Conclusion:* The authors' approach allows to unite opportunities of personalized approach and individual typological data with advantages of estimates based on statistical populations studied factors in neurolinguistic experiments.

*Acknowledgements:* The study is supported by Russian Science Foundation grant 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. doi: 10.3389/fnhum.2019.00437.
2. Greenwald A.G., McGhee D.E., Schwartz J.L. Measuring individual differences in implicit cognition: the implicit association test. *J Personality Social Psychol.* 1998;74(6):1464-1480. doi: 10.1037/0022-3514.74.6.1464.
3. Chin M.S., Kales S.N. Is There an Optimal Autonomic State for Enhanced Flow and Executive Task Performance? *Frontiers Psychol.* 2019;10. doi: 10.3389/fpsyg.2019.01716.

## Genetic markers and neurophysiological correlates of the psychological personality traits among people from different regions of Siberia

Savostyanov A.<sup>1,2,3\*</sup>, Tamozhnikov S.<sup>1</sup>, Karpova A.<sup>4</sup>, Kawai-Ool U.<sup>5</sup>, Ivanov A.<sup>2</sup>, Kazantsev F.<sup>2</sup>, Klimenko A.<sup>2</sup>, Lashin S.<sup>2</sup>

<sup>1</sup> *Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup> *Institute of Medicine at the Ammosov North-Eastern Federal University, Yakutsk, Russia*

<sup>5</sup> *Tuva Science Center, Kyzyl, Russia*

\* *a-sav@mail.ru*

**Key words:** psychological personality traits, allelic polymorphism, EEG, inter-regional differences

*Motivation and Aim:* Psychological genetics investigates the associations between molecular-genetic markers and the human psychological personality traits, including inclinations to mental pathology. Neurophysiological indicators such as EEG measures show associations with both genetic and behavioral indices. The climate environment is also a factor affecting both brain activity and human behavior. Our project investigated associations between genetic markers and neurophysiological correlates of psychological personality traits in people living in different regions of Siberia – Novosibirsk Region, Altai Territory, Republic of Tuva and Republic of Sakha (Yakutia).

*Participants and Methods:* Our study includes two stages. In the first stage, bioinformatic analysis of the scientific papers was performed (Ivanov et al., 2019). The 164 allelic polymorphisms of brain neurotransmitter systems were identified, for each of which its associations with the risk of anxiety-depressive disorders depend on the region of residence and ethnic characteristics of people. At the second stage, we collected and analyzed a collection of data obtained from a sample of 1744 people living in different regions of Siberia. The participants were of eight ethnic groups and differed by region of residence, social status, sex, and age. For all participants, the samples of biological materials were collected, genome sequencing was carried out, and the individual single nucleotide polymorphisms for all previously selected 164 pairs of alleles were defined. All participants completed the set of questionnaires to estimate psychological personality features and a tendency to anxiety-depressive disorder. The most of these people participated in EEG experiments with resting-state conditions and under recognition of emotional-related stimuli.

*Results:* Different alleles of these genes are associated with psychological individual traits in the scores of collectivism-individualism, extroversion, consciousness, trait anxiety and neuroticism. Regional differences in the frequency of distribution of alleles associated with socially significant personality traits in various regions of Siberia have been identified. Associations between allelic polymorphism of selected genes and delta and alpha rhythm scores under resting conditions and emotional speech recognition have been investigated.



*Conclusion:* The results make it possible to propose a model of the complex influence of genetic and regional factors on the formation of personal psychological characteristics of people.

*Acknowledgements:* Supported by the Russian Foundation of Basic Research, the grant No. 18-29-13027 and the budget project No. 0259-2021-0009 at the Institute of Cytology and Genetics of the SB RAS.

### *References*

1. 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*. 2019;10(9):699. doi: 10.3390/genes10090699.

## Calcium activity of neuron-glia networks upon HIF prolyl hydroxylase inhibition in hypoxia *in vitro*

Savyuk M.<sup>1\*</sup>, Poloznikov A.<sup>2</sup>, Vedunova M.<sup>1</sup>, Mitroshina E.<sup>1</sup>

<sup>1</sup> National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

<sup>2</sup> Higher School of Economics, Moscow, Russia

\* [mary.savyuk@bk.ru](mailto:mary.savyuk@bk.ru)

**Key words:** calcium imaging, neural networks, primary cell cultures

**Motivation and Aim:** Hypoxia is one of the main pathogenetic components of ischemic stroke. It also accompanies various neurodegenerative diseases, for example, Alzheimer's disease, Parkinson's disease, Huntington's disease, and brain traumas [1]. So, it is important to find a way to reduce neuronal damage and adapt nerve cells to hypoxia. One of the main molecular mechanisms of cell's hypoxic adaptation is hypoxia-inducible factor-1 (HIF-1), which controls many genes associated with angiogenesis and cells survival. HIF activity is regulated by oxygen-dependent HIF prolyl hydroxylase (PHD). Pharmacological inhibition of PHD by highly selective inhibitors is promising idea for hypoxic damage correction [2].

**Methods and Algorithms:** Primary mouse hippocampal cultures were used for *in vitro* studies [3]. Modeling of acute hypoxia was carried out by complete replacement of the culture medium with a low oxygen for 10 min on 14 DIV. In order to characterize the functional calcium dynamics of neuron-glia networks, the functional calcium imaging was carried out. Fluorescent calcium sensitive color dye Oregon Green 488 BAPTA-1 AM (OGB-1) (Invitrogen, USA) was used as a calcium sensor. Visualization was performed using an LSM 800 confocal laser scanning microscope (Carl Zeiss, Germany) [4]. Previously developed software was used for calcium activity analysis [5]. Statistical analysis was performed using GraphPad Prism v.9.3.1.471.

**Results:** We studied the effects of the novel D'014-0038 PHD inhibitor on network calcium activity of hippocampal cultures. Inhibitor was applied in cultural medium immediately after hypoxia simulation in concentration 1  $\mu\text{M}$ . The following calcium characteristics were analyzed: the duration and frequency of calcium events, the number of working cells, the level of correlation of the calcium signal in pairs of cells, the speed of signal propagation, and the number of existing functional connections relative to the maximum possible mathematically calculated connections number. It has been shown that hypoxia causes a significant decrease in all calcium activity characteristics compare to the intact group. Our data show that PHD blockades by D'014-0038 (1  $\mu\text{M}$ ) preserve the activity of nervous cells, but did not restore connectivity of the neuron-glia network in post-hypoxic periods. The frequency of calcium oscillations in the "PHD inhibitor" group was significantly higher than the "Hypoxia" group ( $0.24 \pm 0.07$  and  $0.12 \pm 0.01$  osc/min respectively). Moreover, a similar effect was shown for the parameter "Number of active cells" ("PHD inhibitor" –  $71.35 \pm 5.93$  % active cells and "Hypoxia"  $42.76 \pm 7.17$  % active cells). However, there is no significant difference in the level of calcium signal correlation and the speed of signal propagation.

*Conclusion:* It was shown that the use of the PHD inhibitor D'014-0038 maintain calcium activity of neuronal cells in hypoxia modeling, but does not preserve network connectivity

*Acknowledgements:* This research was funded by grant from the Russian Science Foundation (RSF), project No. 18-75-10071-p.

### *References*

1. Mitroshina E. et al. Hypoxia-Inducible Factor (HIF) in Ischemic Stroke and Neurodegenerative Disease. *Front Cell Dev Biol* 2021;9:703084.
2. Savyuk M. et al. Neuroprotective effect of HIF prolyl hydroxylase inhibition in an *in vitro* hypoxia model. *Antioxidants* 2020;9:662.
3. Vedunova M. et al. Seizure-like activity in hyaluronidase-treated dissociated hippocampal cultures. *Front Cell Neurosci.* 2013;7:149.
4. Kustikova V., Krivonosov M., Pimashkin A., Denisov P., Zaikin A., Ivanchenko M., Meyerov I., Semyanov A. CalciumCV: computer vision software for calcium signaling in astrocytes. In: Proceedings of the 7th International Conference, AIST 2018, Moscow, Russia, 5–7 July 2018.

## Identification of a novel small RNA encoded in the mouse urokinase receptor uPAR gene (*Plaur*) and its molecular targets

Semina E.<sup>1,2\*</sup>, Rysenkova K.<sup>1,2</sup>, Troyanovskiy K.<sup>2</sup>, Klimovich P.<sup>1,2</sup>, Bulyakova T.<sup>2</sup>, Ivashkina O.<sup>4,5,6</sup>, Anokhin K.<sup>4,5</sup>, Karagyaur M.<sup>2</sup>, Zvereva M.<sup>3</sup>, Rubina K.<sup>2</sup>, Tkachuk V.<sup>1,2</sup>

<sup>1</sup> *Institute of Experimental Cardiology, National Cardiology Research Center of the Ministry of Health of the Russian Federation, Moscow, Russia*

<sup>2</sup> *Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*

<sup>4</sup> *Institute for Advanced Brain Studies, Lomonosov Moscow State University, Moscow, Russia*

<sup>5</sup> *Laboratory of Neurobiology of Memory, FSBSI P.K. Anokhin Research Institute of Normal Physiology, Moscow, Russia*

<sup>6</sup> *Laboratory of Neuroscience, NBICS Center, Kurchatov Institute National Research Center, Moscow, Russia*

\* *e-semina@yandex.ru*

**Key words:** urokinase receptor, uPAR, *Plaur*, *Plaur-miR1-5p*, *Emx2*, *Mef2d*, *Snmp200*, neuroblastoma

The urokinase receptor (uPAR) is the glycosylphosphatidylinositol (GPI)-anchored receptor of urokinase (uPA), which is involved in brain development, nerve regeneration, wound healing and tissue remodelling. uPAR overexpression stimulates radial neuronal migration to the outer layers of the differentiating cortex, whereas uPAR knockout reduces the migration of parvalbumin-expressing GABA interneurons into the cerebral cortex. Recent papers have shown that mutations and polymorphisms in the *Plaur* gene or uPAR ligand *SRPX2* affect the formation of brain structures and induce severe developmental pathologies in humans (speech deficiency, mental weakness and autism spectrum disorders). Using a model of acute generalised seizures in mice, we revealed that *Plaur* operates as an immediate early gene, and is rapidly induced by neuronal activity in different brain regions independently of the *de novo* protein synthesis. This rapid and universal response confirms the important role of uPAR in regulating neuronal responses to excitation and/or damage.

Assumingly, diverse functions of *Plaur* might be attributed to hypothetical, unidentified microRNAs encoded within the introns of the *Plaur* gene. The *Plaur* gene consists of seven exons and seven introns. Prior to our study, miRNAs had not been identified in the *Plaur* gene sequence. The total *Plaur* gene size is 16,000 bp, while the mature mRNA (exons only) is composed of only 1,000 nt, suggesting that non-coding RNAs, including miRNAs, are encoded in the gene. We conducted a bioinformatic search and analysed the miRNAs that make up the *Plaur* gene. We identified a novel miRNA, which we named *Plaur-miR1*. Using wild-type uPAR-expressing Neuro2a cells, CRISPR-edited uPAR-deficient Neuro2a cells and *in vivo* model of endogenous induction of *Plaur* expression in the brain, we proved the existence of a new miRNA, *Plaur-miR1-5p*. *In silico* analysis of targets allowed us to identify its possible functions, namely determining cell fates in the developing central nervous system (plays a crucial role in neuronal apoptosis). We confirmed *Plaur*-dependent expression of *Plaur-miR1* in the mouse brain and mouse neuroblastoma Neuro2a cells. Utilising the *in silico* MR-microT algorithm in

DianaTools, we selected two target genes – *Emx2* (is a transcription factor that plays an essential role in specifying cell fates in the embryonic central nervous system) and *Mef2d* (encodes a developmental protein that regulates large-scale gene expression programmes necessary in embryogenesis and tissue architecture maintenance, including the brain, and contributes to the regulation of neurogenesis, neuronal apoptosis and differentiation) – with high binding score. Moreover, sequencing of the mouse brain samples for Plaur-miR1 target genes revealed two other target genes – *Nrip3* (nuclear receptor interacting protein 3) and *Snrnp200* (U5 small nuclear ribonucleoprotein). The expression of *Emx2*, *Mef2d* and *Snrnp200* in mouse brain and *Mef2d* and *Snrnp200* in Neuro2a cells correlated with the expression of *Plaur* and Plaur-miR1.

In conclusion, we identified the novel Plaur-miR1 as a functional miRNA of *Plaur* intron 3. Furthermore, we revealed that Plaur-miR1 expression regulates *Emx2* and *Mef2d* expression. Taking into account our previously published data and the present results we suggest a novel role of *Plaur* as a morphogenetic factor in the brain development and a marker of brain disorders.

Accession numbers. Sequences from the figure 4 E (clones 1, 9, 11 and 16) for the reported microRNA Plaur-miR1-5p have been deposited with the GenBank® under sequences names and accession numbers as follows: BankIt2532817 Seq\_1 OM105889, BankIt2532817 Seq\_2 OM105890, BankIt2532817 Seq\_3 OM105891, BankIt2532817 Seq\_4 OM105892.

*Funding:* This research was supported by the Russian Science Foundation, Project No. 19-75-30007 ‘Fundamental issues of regenerative medicine: regulation of human tissue renewal and repair’, and by the Ministry of Science and Higher Education of the Russian Federation, Project No. 075-15-2020-801. Isolation of mRNA and miRNA from the mouse brain was carried out with the support of project No. 075-15-2020-801. All other methods and research were supported by Project No. 19-75-30007.

## Protective properties of DCD and LRP2 proteins in schizophrenia and bipolar disorder

Seregin A.A. \*, Dmitrieva E.M., Smirnova L.P., Ivanova S.A., Semke A.V.

Tomsk National Research Medical Center, RAS, Mental Health Research Institute, Tomsk, Russia

\* [apocalips1991@mail.ru](mailto:apocalips1991@mail.ru)

**Key words:** bipolar disorder, schizophrenia, proteome, mass spectrometry, serum, biomarker

*Motivation and Aim:* Schizophrenia and bipolar disorder (BD) are a heterogeneous group of disorders with unclear etiology and pathogenesis, which makes it difficult to diagnose these severe diseases. Proteomic methods can make it possible to identify protein markers associated with the pathogenetic mechanisms of these disorders, which can serve as a basis for the development of new methods for diagnosing and treating schizophrenia and BD [1].

*Methods and Algorithms:* We analyzed the blood serum of 63 patients with paranoid schizophrenia (F20.0) aged 34[28;40] years and 23 patients with BD (F31) aged 32[21;52] years. Patients were hospitalized in an acute condition and blood samples were taken before the start of therapy. The control group consisted of 25 mentally and somatically healthy volunteers corresponding to the gender and age of the examined groups. For mass-spectrometric analysis of serum, 10 people from each group of patients with schizophrenia, BD and healthy individuals were selected. Sample preparation, mass spectrometry and data analysis were carried out as described earlier in our article [2]. Statistical analysis was performed using a two-tailed unpaired t-test (FDR = 0.05 and  $S_0 = 2$ ). To measure the level of dermcidin in the serum of the study groups, the ELISA Kit for Dermcidin (DCD) SEC896Hu<sup>®</sup> (Cloud-Clone Corp., USA) was used. The level of Low-density lipoprotein receptor-related protein 2 (LRP2) was determined using the ELISA Kit for Low Density Lipoprotein Receptor Related Protein 2 (LRP2) Hu (Cloud-Clone Corp., USA). The statistical significance of the result of the ELISA analysis was tested by the nonparametric Mann-Whitney U-test.

*Results:* As a result, more than 1600 proteins were detected during mass spectrometry. Data analysis using MaxQuant and further evaluation of normalized mean LFQ intensities relative to the studied groups revealed unique proteins that differed significantly for each group: 27 proteins specific for schizophrenia and 18 for BD [2]. Proteins found in patients with schizophrenia are mainly involved in the immune response, cell communication, regulation of protein and nucleic acid metabolism. Unique proteins identified in the serum of BD patients are involved in the regulation of DNA synthesis and cell cycle, differentiation of nerve cells, and transport processes across the cell membrane. Many of these proteins are involved in the pathogenesis of psychiatric disorders and 2 of the most interesting proteins were chosen for further study: DCD and LRP2. But a comparison of their levels between groups of patients with schizophrenia, BD and healthy individuals did not show significant differences. Further, we divided the patients with schizophrenia into 2 subgroups: 33 patients with progressive defect (F20.01) and 30 patients with stable defect (F20.02). In patients with schizophrenia with progressive disease, a decrease in the level of DCD (0.9 [0.56; 1.23] ng/ml Mann-Whitney  $p = 0.0038$ ) and a significant increase in the level of LRP2 (146.70

[102.175; 231.702] pg/ml Mann-Whitney  $p = 0.026$ ), as well as a decrease in DCD in patients with a stable defect (0.435 [0.26; 0.73] ng/ml Mann-Whitney  $p = 0.048$ ) compared with healthy controls (LRP2 99.48 [63.65; 142.29] pg/mL, DCD 1.50 [0.62; 3.22] ng/mL). In addition, a decrease in the level of DCD was found in patients with a stable defect (Mann–Whitney  $p = 0.045$ ) in comparison with the group of patients with BD (1.21 [0.53; 2.06] ng/ml).

**Conclusion:** The protein dermcidin, in addition to its antimicrobial activity, is a factor in the survival of neurons in the event of their damage [3]; its RNA is expressed in the pons varolii and the paracentral gyrus of the brain [4]. Also, the dermcidin-derived peptide Y-P30 was isolated from the cell lines of the nervous system [5]. It is known that Y-P30 has a neuroprotective effect [6] and promotes the growth of neurites from thalamic and cerebellar progenitor neurons [7]. It is likely that the decrease in DCD levels in patients with schizophrenia compared with controls and in patients with schizophrenia with a stable defect compared with bipolar disorder may be associated with more pronounced neuronal damage during the pathogenesis of schizophrenia. However, the role of DCD in this process is yet to be elucidated. LRP2 is involved in the regulation of the development of neural stem cells in the fetus and progenitor cells during maturation [8]. In adulthood, LRP2 regulates the development and migration of neural stem cells located in two key niches, the subventricular zone of the lateral ventricles and the dentate gyrus of the hippocampus [9]. The function of LRPs in neuronal development is poorly understood, but they have the potential to play a regenerative and projective role in neuronal injury in the pathogenesis of Schizophrenia. An increase in LRP2 levels in patients with progressive defect schizophrenia supports this hypothesis. Other works also confirm the involvement of LRP2 in the pathogenesis of schizophrenia [10]. The proteins we have identified can quite possibly be used as biomarkers upon detailed study, and the study of their neuroprotective mechanisms can serve to elucidate the pathogenesis of mental disorders and, probably, to develop new methods of therapy.

**Acknowledgements:** This research was funded by Grant of RSF No. 18-15-00053-P.

## References

1. Loginova L.V. et al. Mass-spectrometric analysis of serum proteins in patients with schizophrenia. *Bull Sib Branch Russ Academy Med Sci.* 2011;31(6):63-68.
2. Smirnova L. et al. The difference in serum proteomes in schizophrenia and bipolar disorder. *BMC Genomics.* 2019;20(7):535. doi: 10.1186/s12864-019-5848-1.
3. 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.
4. Porter D. et al. A neural survival factor is a candidate oncogene in breast cancer. *Proc Natl Acad Sci USA.* 2003;100:10931-10936.
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 neurogenic activity. *J Biol Chem.* 2008;283:25036-25045.
8. Pollen 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;13(9):1261-1264.
10. Gibbons A.S. et al. Density Lipoprotein Receptor-Related Protein and Apolipoprotein E Expression is Altered in Schizophrenia. *Front Psychiatry.* 2010;1:19. doi: 10.3389/fpsy.2010.00019.

## **P.Y. Halperin's activity theory and language consciousness theory: results of EEG-based research in the perspective of ICT-assisted foreign language learning**

Shtern E.N.<sup>1</sup>, Savostyanov A.N.<sup>2</sup>, Lebedkin D.A.<sup>3</sup>

<sup>1</sup> *Lingvator: German language laboratory, Novosibirsk, Russia*

<sup>2</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>3</sup> *Humanitarian Institute, Novosibirsk State University, Novosibirsk, Russia*

**Key words:** activity theory, language consciousness, ICT in language learning, brain oscillatory activity in language tasks

This article presents the results of an EEG-based study of peculiarities of brain oscillatory activity during language tasks (based on the German language). The aim of the study was to reveal neurophysiological indicators, which allow to claim a significant difference between the ways of information input when considering language activity of learners at the neurolinguistic level. The methodological approach in the experimental group was based on the theory of stage-by-stage formation of mental activity and language consciousness in the understanding of P.Y. Galperin, according to which the native language turns into a language-model, and already with its help the mechanisms of language work are realized, and this skill in turn is transferred to the process of mastering the German language, gradually making the rejection of the native language possible. The results obtained allow us to claim the existence of certain differences between the groups of subjects, consisting in a smaller amplitude of the P300 peak and its smaller duration in the experimental group in comparison with the control one. Less activity in the alpha range for the experimental group was noticed. Taken together, these parameters may suggest less difficulty (compared to the control group) in completing tasks. The results of the study will serve as a neuro- and psycholinguistic basis for the development of information and communication technologies in the teaching of German on the basis of the existing linguodidactic course leading to the language proficiency test in accordance with international requirements CEFRL. The guiding foundations of the action theory of type 3 from P.Y. Halperin, the effectiveness of which is confirmed by neuro- and psycholinguistic parameters, can constitute an alternative basis for the development of digital learning programs which are based so far in foreign and Russian applications on the Ebbinghaus curve.



## Concordance between the *in vivo* concentrations of neurospecific proteins (BDNF, NSE) in the brain and blood in patients with epilepsy

Shvaikovskaya A.A.<sup>1\*</sup>, Zhanaeva S.Y.<sup>1</sup>, Olkov N.M.<sup>1,2</sup>, Moysak G.I.<sup>1,3</sup>, Danilenko K.V.<sup>1</sup>, Aftanas L.I.<sup>1,2</sup>, Tikhonova M.A.<sup>1</sup>

<sup>1</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

<sup>2</sup> *V. Zelman Institute of Medicine and Psychology, Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Federal Centre for Neurosurgery, Novosibirsk, Russia*

\* *a.a.shvaikovskaya@mail.ru*

**Key words:** human, brain, blood, brain-derived neurotrophic factor (BDNF), neuron-specific enolase (NSE), epilepsy

The state of the brain is usually assessed by analyzing brain-specific substances in peripheral biological fluids that are easily available for research (blood, urine, saliva, and less often the cerebrospinal fluid). On the other hand, blood-brain barrier and a number of other factors may make such association weak or completely absent. The extent to which the concentration of neurospecific molecules at the periphery corresponds to their content in the brain remains unknown for the most analytes. The study was aimed at elucidating a concordance of the content of neurospecific proteins, namely brain-derived neurotrophic factor BDNF (a universal marker of neurogenesis and neuroplasticity) and neuron-specific enolase NSE (a marker of neuronal damage in brain diseases including epilepsy) in the hippocampus and blood. The subjects were patients with epilepsy who underwent a planned resection of the hippocampus. Specimens for the analysis were hippocampal tissue and venous blood leukocytes/serum taken one hour prior the surgery. The analysis included a measurement of BDNF by multiplex solid phase analysis, NSE by enzyme-linked immunosorbent assay (ELISA), and total protein by biochemical methods. NSE concentration corrected by total protein correlated positively between the brain and leukocytes, with the strength of this correlation to be mild ( $r = 0.45$ ,  $p = 0.047$ , Pearson test,  $N = 20$ ). No other correlations were significant. This preliminary finding indicates that the concentration of some neurospecific proteins in the periphery reflects their concentration in the brain in humans.

*Acknowledgements:* the study was supported by Russian Science Foundation (grant No. 22-25-00588).

## IgG-Dependent dismutation of Superoxide and DPPH radical in patients with schizophrenia

Smirnova L.P.<sup>1</sup>, Mednova I.A.<sup>1</sup>, Vasilyeva A.R.<sup>2</sup>, Kazantseva D.V.<sup>2</sup>, Krotenko N.M.<sup>1,2</sup>, Ivanova S.A.<sup>1</sup>

<sup>1</sup> Mental Health Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia

<sup>2</sup> Siberian State Medical University, SSMU, Tomsk, Russia

\* [lpsmirnova@yandex.ru](mailto:lpsmirnova@yandex.ru)

**Key words:** Schizophrenia, oxidative stress, superoxide dismutase, abzyme, DPPH radical

*Motivation and Aim:* There is controversy over the schizophrenia genesis, but identified neu-rotoxicity and alterations in structural and brain connectivity suggest the presence of degenerative processes in the brain. Oxidative stress is an important factor in schizophrenia, since neuronal damage can be associated with either an increase in the oxidative process, or a de-crease in antioxidant protection, or both [1]. An imbalance in the antioxidant defense system is observed both in patients with a first episode of schizophrenia who have not previously used antipsychotics, and in patients with chronic schizophrenia receiving drug treatment [2, 3]. A number of studies have shown the ability of antibodies isolated from the blood of humans and other mammals to catalyze redox processes [4, 5]. IgG in schizophrenia patients have catalase activity that could possibly play a role in protecting organisms from oxidative stress [6]. The aim of this work was to study the ability of immunoglobulins G patients with schizophrenia to dismutate superoxide and DPPH radicals.

*Methods:* Serum of 34 patients with paranoid schizophrenia in the acute phase and 14 mentally and somatically healthy persons matched by sex and age was used. IgG was purified from serum by affinity chromatography on Protein-G-Sepharose columns. The homogeneity of the preparations was confirmed by the method of gradient electrophoresis in 12.5 % PAGE. SOD activity of antibodies was determined spectrophotometrically. SOD activity of IgG was determined by assessing the degree of inhibition of the reduc-tion of nitroblue tetrazolium (NST) in diformazan superoxide radicals, which are gen-erated during the oxidation of xanthine to uric acid in the presence of xanthine oxidase on spectrophotometer Cary 60 (Agilent) at a wavelength of 560 nm. To determine the kinetic parameters of SOD activity of IgG we used different NST concentration (1–37,5 µM). Evaluation of the degree of inhibition of the SOD activity of IgGs was performed using specific nonspecific inhibitor SOD enzyme – triethylenetetramine (TETA). SOD activity of IgGs without inhibitor was taken as 100 %. The antiradical activity of immunoglobulins G was determined using the DPPG radical, by the drop in the optical density of its solution at a wavelength of 520 nm. Statistical analysis was performed using the SPSS software, release 20.0, for Windows. Median, first and third quartiles were calculated. Between-group differences were evaluated using Mann–Whitney U test.

*Results:* We have shown for the first time that antibodies from patients with schizophrenia have SOD activity and this activity is an intrinsic property of antibodies. The maximum increase in SOD activity was registered in the group of patients with acute schizophrenia ( $p = 0.001$ ). Based on the data of inhibitory analysis using a specific TETA

inhibitor, we can assume that the mechanism of the SOD activity of IgG is similar to the mechanism of classical enzyme catalysis. According to the kinetic analysis, the affinity of the abzyme to the substrate is higher than that of the classical SOD enzyme. Immunoglobulins G have not been shown to bind DPPH radical.

*Conclusions:* Thus, Immunoglobulins from schizophrenic patients have a high ability to bind the superoxide radical, but do not have other antiradical activity. It is likely that superoxide dismuting antibodies help to reduce the level of oxidative stress in the body of patients.

*Acknowledgements:* This work was supported by the Russian Science Foundation, grant No. 18-15-00053P (2021–2022).

### References

1. Ermakov E.A., Dmitrieva E.M., Parshukova D.A., Kazantseva D.V., Vasilieva A.R., Smirnova L.P. Oxidative Stress-Related Mechanisms in Schizophrenia Pathogenesis and New Treatment Perspectives. *Oxid Med Cell Longev.* 2021;2021:8881770. doi: 10.1155/2021/8881770.
2. Ivanova S.A., Smirnova L.P., Shchigoreva Y.G., Semke A.V., Bokhan N.A. Serum glutathione in patients with schizophrenia in dynamics of antipsychotic therapy. *Bull Exp Biol Med.* 2015;160(2):283-285.
3. Krotenko N.M., Smirnova L.P., Loginov V.N., Ivanova A.S., Semke A.V. Influence of neuroleptic therapy on state of lipid peroxidation and glutathione system in schizophrenic patients. *Siberian Herald Psychiatry Addiction Psychiatry.* 2010;2(59):133-135.
4. Kulberg A., Petyaev I.M., Zamotaeva N.G. Catalytic properties and catalytic destruction of cellular receptors (R-proteins). *Immunology.* 1988;3:37-40.
5. 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.
6. 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.

## **Cerebral dopamine neurotrophic factor (CDNF) affects hippocampal neuroplasticity and fear memory in rats**

Tsybko A.<sup>1\*</sup>, Ilchibaeva T.<sup>1</sup>, Yarkov R.<sup>2</sup>, Eremin D.<sup>1</sup>, Vedunova M.<sup>2</sup>, Naumenko V.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Biology and Biomedicine, Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia*

\* *antoncybko@mail.ru*

Cerebral dopamine neurotrophic factor (CDNF) is considered to be a protective factor for brain dopaminergic neurons. A little is known about the involvement of CDNF in the regulation of behavior aside from locomotor activity. There are some evidence indicating the role of CDNF in spatial memory, however, it is unknown whether CDNF is involved in the regulation of other forms of memory, e.g. fear memory. To address this question we used AAV-mediated gene delivery to induce CDNF overexpression in hippocampal neurons of adult Wistar rat males. We found that CDNF overexpression impaired conditioned response recall in fear conditioning paradigm by enhancement of the fear memory extinction. These alterations in fear memory were accompanied by elevated mRNA level of Creb implicated in Long-Term Potentiation formation. At the same time, CDNF overexpression led to increase in mRNA levels of unfolded protein response markers Atf6, Perk and spliced Xbp1. In the rat primary hippocampal culture the CDNF overexpression stimulated the formation of burst activity and significantly increased not only the frequency of spikes, but also the number of spikes in a network burst. Also, it has been shown that CDNF increased not only the number of bonds per cell, but also the correlation of signals in neuronal culture. At the same time, both correlations between neighboring and distant cells were increased. The percentage of active cells in culture was increased by more than two times.

Thus, for the first time we have demonstrated that endogenously overexpressed CDNF can negatively regulate the fear memory formation that may be linked with the marked effects on hippocampal neuroplasticity produced by CDNF.

*Acknowledgements:* The work supported by the Russian Science Foundation (grant No. 22-15-00011) and basic research project No. FWNR-2022-0023.

## Influence of sulfur metabolism genes on processes related to learning and memory in *Drosophila melanogaster*

Zakluta A.S.<sup>1</sup>, Nikitina E.S.<sup>2,3</sup>, Chuvakova L.S.<sup>1</sup>, Rezvykh A.P.<sup>1</sup>, Zatsepina O.G.<sup>1</sup>, Evgen'ev M.B.<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

<sup>2</sup> Department of Neurogenetics, Pavlov Institute of Physiology, St. Petersburg, Russia

<sup>3</sup> Department of Human and Animal Anatomy and Physiology, Herzen State Pedagogical University, St. Petersburg, Russia

**Key words:** transcriptome, RNA-seq, *Drosophila melanogaster*, H<sub>2</sub>S, learning, memory

The gas-transmitter hydrogen sulfide (H<sub>2</sub>S) is one of the most important biological mediators produced by the transsulfuration pathway (TSP) and is involved in many physiological and pathological processes related to oxidative stress, reproduction, learning, and memory. The transsulfuration pathway (TSP) converts homocysteine to cysteine after the breakdown of methionine. In *Drosophila melanogaster* and other eukaryotes, the enzymes cystathionine- $\beta$ -synthase (CBS) and cystathionine- $\gamma$ -lyase (CSE), play a central role in H<sub>2</sub>S formation and metabolism. We obtained fly lines with sulfur metabolism gene deletions using the CRISPR/Cas9 technique. We also obtained lines with double deletion of CBS and CSE genes. In this study, we investigated the role of H<sub>2</sub>S in learning and memory processes using the courtship paradigm with mating rejection. It was shown that in fly lines with double gene deletion and CBS gene deletion, short- and long-term memory formation was completely impaired. In the fly line with CSE gene deletion, only impaired memory consolidation processes were observed. We obtained and analyzed transcriptome libraries from the heads of naive males with gene deletions. Sequencing was performed on the Illumina NextSeq 500 platform. Sequencing yielded approximately 15-20 million reads for each library. Statistical analysis showed that there were approximately 4 times as many down-regulated genes as up-regulated genes. Of particular interest is that the maximum number of genes that significantly changed the level of gene expression was found in lines with single deletion of the CBS gene (295 genes). At the same time, lines with double deletion showed a significantly lower number with altered gene expression levels (190 genes).

Transcriptome analysis revealed multiple changes by Gene Ontology terms. We found significant down-regulation of genes that are involved in learning and memory processes in all lines with CBS gene deletion and especially in lines with double deletion of sulfur metabolism genes.

## Memory formation in *Drosophila* with neurospecific suppression *limk1* gene expression in nervous system

Zalomaeva E.S.<sup>1,2\*</sup>, Egozova E.S.<sup>1,3</sup>, Medvedeva A.V.<sup>2</sup>, Zhuravlev A.V.<sup>2</sup>,  
Nikitina E.A.<sup>1,2</sup>

<sup>1</sup> Herzen State Pedagogical University of Russia, St. Petersburg, Russia

<sup>2</sup> Pavlov Institute of Physiology, RAS, St. Petersburg, Russia

<sup>3</sup> Saint-Petersburg Institute of Bioregulation and Gerontology, St. Petersburg, Russia

\* Zalomaeva.E@yandex.ru

**Key words:** *limk1*, learning, memory, forgetting

Nowadays one of the significant problem of neurobiology is the investigation of the etiology and progression of different neurodegenerative diseases. One of the causes of neurodegenerative diseases is disturbance of actin remodeling cascade whose key enzyme is LIMK1. *Drosophila* constitutes a convenient model for studying the link between genome organization and chromosome architecture observed in cognitive disorders. The revealed association between *limk1* gene's mutational damage, changes in its expression and activity as well as cognitive impairment allows to use current model for the study of neurodegenerative and genomic diseases. We analyzed the formation and dynamics of short-term and medium-term memory in *Drosophila* stocks with neurospecific suppression *limk1* gene expression in nervous system, without such suppression and wild type *CS* stock. Conditioned courtship suppression paradigm was used to assess learning ability and memory formation. A 5-day-old virgin male was put in a special box with a fertilized female *CS* and was left for 30 min. Courtship behavior was analyzed in naive males and in males in 0, 15, 30, 60 min and 24 h after training to assess the formation and dynamics of memory. The behavior of the male was recorded in a special program for 300 sec. We calculated the learning index (LI) to assess the effectiveness of learning. Randomization test was used to statistical analysis. Research demonstrated that LI of all stocks on all points after training significantly different from zero, but not different from each other within the same stock. LI *CS* stock and stock with suppression *limk1* gene after 24 h was significantly less than LI after 0 min ( $p \leq 0,05$ ). LI stock without suppression *limk1* gene was not decline on all the points. Thus, forgetting *Drosophila* with no suppression *limk1* gene is less pronounced.

*Acknowledgements:* The study was funded by RFBR (No. 20-015-00300 A).

# EEG preprocessing algorithm to measure DMN connectivity

Zavarzin E.<sup>1\*</sup>, Milakhina N.<sup>3</sup>, Karpova A.<sup>4</sup>, Savostyanov A.<sup>1,2,3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup> North-Eastern Federal University, Yakutsk, Russia

\* zavarzinevg@gmail.com

**Key words:** EEG, default-mode network, depression

*Motivation and Aim:* Depression is a very dangerous psychophysiological disease. By now, many methods of depression diagnostics, based on brain activity (for example, evaluated by EEG), developed. There is the concept of default-mode network, clusters of neurons that are highly correlated by their activity when a subject is distracted or daydreaming. High DMN connectivity in active states can be interpreted as a marker of depression [1], so a DMN connectivity measuring algorithm should be developed.

*Methods and Algorithms:* To test the algorithm on a large set of data, EEG data from ICBrainDB [2] was used, 173 EEG recordings at rest and during an active task. To measure DMN activity, correlation analysis was performed. Total DMN activity depends on inner correlations of DMN regions.

*Results:* Algorithm was applied to EEG of healthy subjects. Because DMN is highly correlated in resting state and lower correlated in active state, binary classification models were trained to test DMN activity quantities for EEG samples between states. RandomForest and KNeighbours models have above 80 % precision.

*Conclusion:* Preprocessing algorithm should be used in further studies to test more experimental groups and conditions.

*Acknowledgements:* Development of the data analysis algorithm was supported by the grant of Russian Science Foundation No. 22-15-00142.

## References

1. Knyazev G.G., Savostyanov A.N., Bocharov A.V., Brak I.V., Osipov E.A., Filimonova E.A., ... Aftanas L.I. Task-positive and task-negative networks in major depressive disorder: a combined fMRI and EEG study. *J Affective Disorders*. 2018;235:211-219.
2. Ivanov R., Kazantsev F., Zavarzin E., Klimenko A., Milakhina N., Matushkin, Y.G., ... Lashin S. ICBrainDB: An Integrated Database for Finding Associations between Genetic Factors and EEG Markers of Depressive Disorders. *J Personalized Med*. 2022;12(1):53.

11

Symposium “Systems biology  
and bioinformatics of DNA  
repair processes and programmed  
cell death”





## Role of PAR and RNA in biomolecular condensate formation

Alemasova E.\*, Sukhanova M., Lavrik O.

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

\* [lisenok.istreb@gmail.com](mailto:lisenok.istreb@gmail.com)

**Key words:** poly(ADP-ribose) (PAR), RNA, biomolecular condensate

Condensates are intracellular compartments that accumulate certain biomolecules without the help of membranes. These molecular assemblies may emerge as a result of sub-stoichiometric mechanisms (binding, bridging), liquid-liquid phase separation (LLPS) or their superposition [1]. Understanding how a specific structure is formed is necessary for correct predictions of how it will behave. Nucleic acids (DNA, RNA, and poly(ADP-ribose) [PAR]) are key scaffold molecules that function during the architectural organization and dynamic restructuring of condensates. PAR is produced in the cells of higher eukaryotes and has features of structure and synthesis that distinguish it from canonical nucleic acids [2]. Similarities and differences between PAR and RNA can determine their interplay during the formation and functioning of condensates with complex architecture. In our review work we analyze the basic principles of how the properties of constituent nucleic acids (length, sequence, structure, and rigidity) influence condensate organization. We discuss the role of RNA in the formation of condensates with different morphologies and varying RNA-protein composition. We also examine in detail the majority of known PAR-containing condensates (the nucleolus, DNA repair foci, transcription hubs, the spindle, RNP granules and extracellular mineral particles) to understand the function of PAR in their assembly. We conclude that both being polymeric scaffolds, PAR and RNA can enter into multiple specific and nonspecific interactions. For nonspecific electrostatic contacts, these anionic polynucleotides are interchangeable. However, PAR may function as a more rapidly synthesized and less site-dependent RNA analog initiating RNA-dependent systems at early stages of their activity. It is possible that similarities between PAR and RNA can be used in systems with phase re-entrant behavior [3]. In particular, based on our data, we hypothesize that PAR synthesis itself may have the features of re-LLPS process. Despite its low-complexity, PAR can be recognized through a variety of protein domains [4]. Therefore, PAR and RNA can bind with different partner proteins, ensuring the immiscibility of liquid phases and the formation of multilayer condensates. Due to reversibility of PAR synthesis and degradation, this polymer is more advantageous for the assembly of transient compartments, while RNA can contribute to organization of less dynamic condensates.

*Acknowledgements:* The study is supported by the Russian Science foundation (grant No. 20-14-00086).

### References

1. Erdel F., Rippe K. Formation of chromatin subcompartments by phase separation. *Biophysical J.* 2018;114(10):2262-2270.
2. Leung A.K.L. Poly(ADP-ribose): a dynamic trigger for biomolecular condensate formation. *Trends Cell Biol.* 2020;30(5):370-383.
3. Portz B., Shorter J. Biochemical timekeeping via reentrant phase transitions. *J Mol Biol.* 2021;433(12):166794.
4. Teloni F., Altmeyer M. Readers of poly(ADP-ribose): designed to be fit for purpose. *Nucleic Acids Res.* 2016;44(3):993-1006.

## Using fluorescently labeled proteins to study the interactions of APE1 with DNA and Pol $\beta$ during the BER pathway

Bakman A.<sup>1\*</sup>, Boichenko S.<sup>1,2</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* *art-bakman@yandex.ru*

The most common forms of DNA damage are the single base lesions which are corrected through the base excision repair (BER) pathway. BER is a multistep process involving several repair enzymes including apurinic/apyrimidinic endonuclease 1 (APE1) and DNA polymerase  $\beta$  (Pol $\beta$ ). During the BER process APE1 recognizes and incises DNA strands containing AP site. Then Pol $\beta$  adds single nucleotide to the 3'-hydroxyl and removes the dRP group with its intrinsic dRP lyase activity. According to the “passing the baton” model of the BER process, DNA intermediates in BER are processed and then passed from one enzyme to another in a coordinated fashion. The aim of the present study was to create an experimental system based on the fluorescent labeled APE1 that would allow us to analyze the protein-DNA and protein-protein interaction in BER process.

In this work we have prepared Alexa-488 fluorophore Cys-labeled human APE1. The wild type enzyme and several mutant forms containing substitutions of surface Cys residues by Ser were analyzed. It was found that Alexa-488 fluorophore was sensitive to the process of the interaction between labeled APE1 and THF-containing (THF is an analog of AP-site) DNA. The changes in the fluorescence intensity have been attributed to the stages of the enzyme-DNA interaction process. Using the stopped-flow fluorescence technique we have studied the mechanisms of interaction between labeled APE1 and DNA substrates of different length in the presence or absence of Pol $\beta$ . Obtained data revealed that Pol $\beta$  stimulates both APE1 binding with THF-containing DNA duplex with formation of the catalytic competent complex and the APE1 dissociation from its product. These findings confirm the “passing the baton” model for APE1 with Pol $\beta$  coordination in BER and explain Pol $\beta$  stimulation effect on APE1 activity.

*Acknowledgements:* This work was supported by Russian Science Foundation grant No. 21-74-10103.

## Structural features of substrate recognition by APE1-like endonucleases

Bulygin A.A.<sup>1,2\*</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* *abulygin@niboch.nsc.ru*

The integrity of DNA in the cell is preserved by a complex repair system that implements the recognition and removal of DNA damage and restoration of normal DNA structure. AP endonuclease hydrolyzes the phosphodiester bond on the 5' side of the abasic nucleotide in the base excision repair pathway. In the nucleotide incision repair pathway, an enzyme directly recognizes and catalyzes the cleavage of DNA on the 5' side of various specific nonbulky lesions such as 1,N<sup>6</sup>-ethenoadenosine ( $\epsilon$ A), 5,6-dihydrouridine (DHU), deoxyuridine (dU), the  $\alpha$ -anomer of adenosine ( $\alpha$ A) and some other. Despite notable efforts to identify the causes of wide substrate specificity of AP endonucleases, the question on how a given nucleotide is accommodated by the active site of the enzyme remains unanswered. The main purpose of this work was to elucidate the structural features of substrate recognition by APE1-like endonucleases. Molecular dynamic (MD) approach was used to build and to analyze complexes of APE1-like endonucleases including human APE1, zAPE1 from *D. rerio*, xAPE1 from *X. laevis* and Rrp1 from *D. melanogaster* with damaged DNA. Obtained MD data for all tested enzymes revealed that damaged nucleotides have steric interactions with a protein loop Asn229-Thr233. It was found that significant reorganization and relocation of this loop are required to accommodate damaged nucleotide in the active site of the enzymes. To verify MD simulation data three mutant forms of APE1 from *D. rerio*, containing substitution N229G, A230G or E236A were obtained by a site-directed mutagenesis. Pre-steady-state kinetic and PAGE analysis of enzyme catalytic activity revealed that E236A substitution significantly improved the enzyme toward NIR-substrates. A comparison of MD simulation data with catalytic efficiency of these enzymes allowed us to gain deep insight into the mechanism of damage recognition. It was concluded that the Asn229-Thr233 loop could control enzyme specificity.

*Acknowledgements:* This work was supported by Russian Science Foundation Grant No. 21-14-00018.

## Tyrosyl-DNA phosphodiesterase 1 Inhibitors as topotecan sensitizers

Chernyshova I.A.<sup>1\*</sup>, Dyrkheeva N.S.<sup>1</sup>, Zakharenko A.L.<sup>1</sup>, Drenichev M.S.<sup>3</sup>,  
Oslovsky V.E.<sup>3</sup>, Ivanov G.A.<sup>3</sup>, Kurochkin N.N.<sup>3</sup>, Mikhailov S.N.<sup>3</sup>, Lavrik O.I.<sup>1,2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

<sup>3</sup> Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

\* [chernyshova0305@gmail.com](mailto:chernyshova0305@gmail.com)

**Key words:** topotecan, tyrosyl-DNA phosphodiesterase 1, inhibitors, sensitizers

**Motivation and Aim:** Topotecan is a cytostatic drug widely used in the treatment of cervical, ovarian [1] and small cell lung cancer [2]. This drug, getting into the cell, leads to DNA damage due to the ability to stabilize covalent complexes of topoisomerase 1 (Top1)-DNA [3], arising during the catalytic reaction of the enzyme. The accumulation of damage eventually leads to the death of tumor cells. However, there are mechanisms for removing such adducts in the cell, for example, the enzyme tyrosyl-DNA phosphodiesterase 1 (Tdp1) can hydrolyze the covalent bond of Top1-DNA [4], which reduces the effectiveness of Top1 inhibitors used in clinical practice. Therefore, the additional use of Tdp1 inhibitors in chemotherapeutic cocktails is a promising strategy for increasing the sensitivity of tumor cells to topotecan. Previously, it was shown that disaccharide nucleosides with lipophilic groups and similar monosaccharide derivatives inhibit Tdp1 in the range of submicromolar concentrations [5, 6]. Thus, the aim of the work was to study a series of new lipophilic nucleoside derivatives as Tdp1 inhibitors for the further development of drugs capable of sensitizing tumor cells to topotecan.

**Methods and Algorithms:** To evaluate the inhibiting effect of lipophilic nucleoside derivatives on the purified recombinant human Tdp1 enzyme, we conducted fluorescent screening and kinetic experiments, using a fluorescent method based on Tdp1's ability to remove the BHQ1 fluorescence quencher from the 3'-end of the oligonucleotide. To study the biological properties of the most active inhibitors, we identified their toxicity and the synergistic effect with the antitumor drug topotecan *in vitro* and *in vivo*.

**Results:** Effective Tdp1 inhibitors with IC<sub>50</sub> from 0.18 to 6 μM were found among a series of lipophilic nucleoside derivatives. It was shown that the studied compounds are non-toxic in the concentration range up to 100 μM, and five of them increased the cytotoxic effect of topotecan on HeLa cells. Treatment of HeLa cells with a combination of topotecan and a leader compound resulted in an increase the level of DNA damage compared to the use only topotecan. Mice with ascitic carcinoma Krebs-2 treated with topotecan and the leader compound showed a significant decrease in the mass of ascites and the number of tumor cells in it compared with the control group of mice who did not get a therapy.

**Conclusion:** Lipophilic nucleoside derivatives are a promising class of Tdp1 inhibitors. A non-toxic compound capable to sensitize tumor cells towards the topotecan *in vitro* and *in vivo* was found in the investigated series. This leader compound is a promising candidate for the development of antitumor drugs that increase the sensitivity of tumors to the Top1 inhibitor used in clinical practice.

*Acknowledgements:* The work was supported by the RSF grant No. 21-14-00105.

### *References*

1. Lorusso D. et al. Review role of topotecan in gynaecological cancers: current indications and perspectives. *Crit Rev Oncol Hematol.* 2010;74(3):163-174.
2. O'Brien M., Eckardt J., Ramlau R. Recent advances with topotecan in the treatment of lung cancer. *Oncologist.* 2007;12(10):1194-1204.
3. Pommier Y., Marchand C. Interfacial inhibitors: targeting macromolecular complexes. *Nat Rev Drug Discov.* 2012;11(1):25-36.
4. Pommier Y. et al. Tyrosyl-DNA-phosphodiesterases (tdp1 and tdp2). *DNA Rep.* 2014;19:114-129.
5. Komarova A.O. et al. Novel group of tyrosyl-DNA-phosphodiesterase 1 inhibitors based on disaccharide nucleosides as drug prototypes for anti-cancer therapy. *J Enzyme Inhib Med Chem.* 2018; 33(1):1415-1429.
6. Zakharenko A. et al. Inhibition of Tyrosyl-DNA Phosphodiesterase 1 by Lipophilic Pyrimidine Nucleosides. *Molecules.* 2020;25(16):3694.

## **Top1-PARP1 association and beyond: from DNA topology to break repair**

Das B.B.

*Laboratory of Molecular Biology School of Biological Sciences Indian Association for the Cultivation of Science, Jadavpur Kolkata, India*

Selective trapping of human topoisomerase 1 (Top1) on the DNA (Top1 cleavage complexes; Top1cc) by specific Top1-poisons triggers DNA breaks and cell death. Poly(ADPribose) polymerase 1 (PARP1) is an early nick sensor for trapped Top1cc. New mechanistic insights have been developed in recent years to rationalize the importance of PARP1 beyond the repair of Top1-induced DNA breaks. This review summarizes the progress in the molecular mechanisms of trapped Top1cc-induced DNA damage, PARP1 activation at DNA damage sites, PAR-dependent regulation of Top1 nuclear dynamics, and PARP1-associated molecular network for Top1cc repair. Finally, we have discussed the rationale behind the synergy between the combination of Top1 poison and PARP inhibitors in cancer chemotherapies, which is independent of the 'PARP trapping' phenomenon.

## Features of the substrate specificity of a novel AP endonuclease from *Pyrococcus furiosus*

Davletgildeeva A.T.<sup>1\*</sup>, Ishchenko A.A.<sup>2</sup>, Saparbaev M.<sup>2</sup>, Fedorova O.S.<sup>1</sup>,  
Kuznetsov N.A.<sup>1</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Group "Mechanisms of DNA Repair and Carcinogenesis", Université Paris-Saclay, Villejuif, France*

\* *davleta94@gmail.com*

AP endonucleases play one of the key roles in base excision repair (BER) pathway, and present main enzymes of the nucleotide incision repair (NIR). These enzymes catalyze DNA hydrolysis 5' side to the AP site or number of damaged nucleotides. The majority of pro- and eukaryotic AP endonucleases belong to the ExoIII or EndoIV structural families. Nevertheless, a novel endonuclease class of enzymes from archaea and some bacteria, designated endonuclease Q (EndoQ), was identified recently. EndoQ is not structurally related to the known AP endonucleases from ExoIII or EndoIV structural families. However, this enzyme possesses substrate specificity and catalytic activity similar to well-studied AP endonucleases. Present work was aimed to study the kinetic mechanism of damaged nucleotide recognition by EndoQ enzyme from *P. furiosus*. We analyzed kinetics of conformational changes of enzyme and DNA molecules during their interaction. The model DNA substrates containing various damages, including an F-site (the analog of AP site), alpha-adenosine, uridine, hypoxanthine and others, were tested to elucidate the features of lesion recognition. The influence of the nucleotide placed opposite to the lesion on the efficacy of DNA cleavage was analyzed. Strong effect of divalent metal ions on the catalytic stage was shown. The data obtained in current study allowed us to determine kinetic features of specific recognition of various damaged DNA substrates and to evaluate the catalytic activity of EndoQ enzyme in the BER and NIR pathways in archaea.

*Acknowledgements:* This work was supported by the Russian Foundation for Basic Research grant No. 20-34-90008.

## Functional role of active site amino acid residues of human ALKBH2 during recognition of methylated DNA bases

Davletgildeeva A.T.<sup>1</sup>, Zhao M.<sup>2\*</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* *hdzhaomingxing@163.com*

Epigenetic regulation of gene expression is an important control method over the implementation of genetic information by living organisms. Site-specific methylation of DNA is one of the crucial ways of performing this regulation. The enzymatic oxidative demethylation is aimed to control different types of DNA methylation. The main pathway of the removal of methyl groups in DNA is catalyzed by nonheme Fe(II)/ $\alpha$ -ketoglutarate-dependent DNA dioxygenases. Commonly accepted, that the AlkB family dioxygenases are the repair enzymes that catalyze the direct removal of alkyl groups from DNA bases such as 3-methylcytosine, 1-methyladenine, 1,N6-ethenoadenine. Recently was shown, that ALKBH2, one of the human orthologues of AlkB, is able to oxidize alkyl groups not only in the damaged bases, but also methyl group in 5-methylcytosine, which is the main epigenetic mark. It was suggested that the ability of the ALKBH2 to process 5-methylcytosine strongly depends on the ability to accommodate the damaged base in the active site of enzyme. We used the active site mutational remodeling strategy to identify the functional role of Tyr122, Ile168 and Asp173 amino acid residues in the active site. A set of ALKBH2 mutant forms was constructed by site-directed mutagenesis. The amino acid residues Tyr122, Ile168 and Asp173, which form contacts with the methylated bases accordingly to the structural data, were substituted to alanine. The PAGE and pre-steady-state kinetics experiments were conducted to analyze the activity of these mutant forms towards 5-methylcytosine, 3-methylcytosine and 1-methyladenine. It allowed us to evaluate the role of the active site residues in ALKBH2 enzyme's specificity. The data obtained provide insights into the mechanism of substrate recognition by ALKBH2 in epigenetic DNA demethylation process.

*Acknowledgements:* This work was supported by Russian Science Foundation Grant No. 21-14-00018.



## Base excision repair in non-canonical DNA structures

Diatlova E.<sup>1,2\*</sup>, Eroshenko D.<sup>1,2</sup>, Zharkov D.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *e.diatlova@gs.nsu.ru*

**Key words:** DNA glycosylase, non-canonical DNA, base excision repair

*Motivation and Aim:* Maintaining genome integrity is a necessary condition for survival of all organisms. One of the most important DNA repair systems is base excision repair, which starts with recognition and cleavage of small non-bulky DNA modifications (AP sites, modified bases). Although genomic DNA in living cells mainly exists in the B form, many functionally important elements of the genome can adopt alternative structures *in vivo*. Such structures (thereafter referred to as “non-canonical DNA”) include cruciform DNA, hairpins, triplexes, quadruplexes, intercalated motifs (i-motifs), single-stranded DNA, bulges, Z-DNA, DNA/RNA heteroduplexes, and displacement loops (D-loops) containing either RNA (R-loops) or DNA invading strands. Some of them are known as regulatory elements of gene activity and genome stability, others could be intermediates of harmful processes causing genome instability. Non-canonical DNA itself is highly sensitive to DNA damage, causing gene activity misregulation and severe diseases [1, 2]. Repair process in such structures remains poorly investigated.

*Methods and Algorithms:* In this work we applied qualitative and quantitative methods to estimate glycosylase activity at oligonucleotide substrates containing damaged bases or nucleotides – uracil, 5-hydroxyuracil (hoU), 8-oxoguanine (oG) or tetrahydrofuran – and forming non-canonical structures. Glycosylase activity assay was used for cruciform DNA and hairpins, steady-state kinetics was used for heteroduplexes DNA/RNA and 1–2-nucleotide bulges, single turnover and/or burst kinetics were used for bulges, heteroduplexes, i-motifs, and G-quadruplexes. Each type of substrate was tested with several prokaryotic and eukaryotic glycosylases.

*Results:* Uracil-DNA-glycosylases from *E. coli* and human showed activity on all non-canonical structures tested: i-motifs, G-quadruplexes, heteroduplexes and bulges, with no significant effect of the position in the structure or of an opposite base.

Functional but not structural homologs recognizing oxoguanine (Fpg from *E. coli* and human OGG1) were more disparate: they both showed activity on cruciform DNA, hairpins formed by (CAG)<sub>n</sub> triplet repeats, and regular hairpins (activity for both enzymes was detected only when oG was in the stem but not in the loop). Activity of OGG1 at i-motifs was very low whereas Fpg demonstrated cleavage depending the position of oG in the i-motif structure. Both glycosylases showed activity at oG in G-quadruplexes, but OGG1 was very sensitive to the position of oG while Fpg was not. DNA/RNA heteroduplexes and bulges were also cleaved by Fpg with higher efficacy than by OGG1. For these substrates kinetic parameters were obtained where possible. Fpg retained its specificity at heteroduplexes, discriminating substrate with adenine opposite to oG. OGG1 removed oG only from the oG/rC pair. The rate constants of glycosidic bond cleavage ( $k_2$ ) for OGG1 on oG/rC and bulges were two orders of magnitude lower than  $k_2$  for oG in canonical DNA. Fpg removed oG from a one-

nucleotide bulge with the same efficiency as from the oG/C pair, but for two nucleotide bulges the specificity constant ( $k_{\text{cat}}/K_M$ ) was an order of magnitude lower.

Nei from *E. coli* and its mammalian homologs NEIL1 and NEIL2 remove the same damaged bases (mostly oxidized pyrimidines) but with different DNA structure specificity. They all cleaved substrates tested in this work but to varying degrees. NEIL2 worked expectedly better than others on hoU in the loop part of hairpins and in bulges, and much worse on DNA/RNA heteroduplexes.

*Conclusion:* The ability of most glycosylases to recognize and remove its specific damages from non-canonical structures depends of its location in the structure and its accessibility for the catalytic amino acids. In the cases when the glycosylases showed activity, the catalytic efficacy was usually much lower compared to their specific substrate structure.

*Acknowledgements:* The study was supported by Russian Foundation for Basic Research (No. 20-04-00554-a).

### *References*

1. McMurray C.T. Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet.* 2010;11:786-799.
2. Zhao X., Krishnamurthy N., Burrows C.J., David S.S. Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts. *Biochemistry.* 2010;49:1658-1666.

## The new general biological property of stem-like tumor cells

Dolgova E.V.\*, Bogachev S.S.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [dolgova.ev@mail.ru](mailto:dolgova.ev@mail.ru)

**Key words:** stem-like tumor cells, Krebs-2, Epstein-Barr Virus-induced B-lymphoma, double-stranded DNA, proteoglycans/glycoproteins, scavenger receptors, glycosylphosphoinositol-associated proteins, RNA-seq

*Motivation and Aim:* Previously, we have demonstrated poorly differentiated cells of various origin and genesis, including tumor stem-like ones, to be capable of native, i.e. without using any artificial compounds or procedures, internalizing extracellular double-stranded DNA (dsDNA) fragments. To detect and investigate the cells in question, we have designed the 500 bp dsDNA probe, which is TAMRA-labeled PCR-amplified human *Alu*-repeat (*Alu*-TAMRA DNA probe) [1]. *In vivo* engraftment and transcriptome analysis have revealed both tumor-initiating properties of DNA-internalizing cells of murine Krebs-2 carcinoma and overexpression in these cells of genes responsible for asymmetric division, pluripotency, epithelial-to-mesenchymal transition, drug resistance etc. [2]. In experiments with other tumor models (primary human glioblastoma cell lines, established U87 human glioblastoma cell line, Epstein-Barr Virus-induced B-lymphoma (EBV+ B-lymphoma), murine hepatocarcinoma and Lewis sarcoma), TAMRA+ cells have also been proved to manifest the properties characteristic of tumor stem-like cells (TSCs). It is these cells that are essential for the development of grafts *in vivo* and are the center of spheres formation due to secretion of a set of cyto- and chemokines [3-6]. However, the exact mechanism of extracellular dsDNA internalization into cells has not yet been clarified.

*Methods and Algorithms:* In the current report, for murine Krebs-2 carcinoma and human EBV+ B-lymphoma, there have been determined the mechanisms mediating the interaction of dsDNA with the surface of TSCs and its consequent internalization into these cells. The following approaches were used: (i) electrophoretic separation of cells in a free volume with the following estimation of the number of cells capable of binding TAMRA-dsDNA; (ii) pre-exposure of cells to various proteases and/or compounds inhibiting pinocytosis and endocytosis, followed by assessing the resulting efficiency of *Alu*-TAMRA DNA probe internalization; (iii) analysis of the transcriptome of the indicated cell lines followed by the qPCR to validate the differences in expression of some target genes identified as possible participants in the process of dsDNA internalization.

*Results:* It was shown, that the process of binding and internalizing is rather complicated and composed of the following successive stages: (i) initiating electrostatic interaction and contact of a negatively charged dsDNA molecule with a positively charged molecule(s) on the surface of a TSC; (ii) binding of the dsDNA probe to a tumor stem cell surface protein(s) via the formation of a strong chemical/molecular bond; (iii) the very internalization of dsDNA into the cell. Binding of DNA to cell surface proteins is determined by the presence of heparin/polyanion-binding sites within the protein structure, which can be competitively blocked by heparin and/or dextran sulfate, wherein heparin blocks only the binding, while dextran sulfate abrogates both binding and

internalization. The abrogation of internalization by dextran sulfate implies the role of scavenger receptors in this process. Cells were shown to uptake DNA in amounts constituting ~0.008 % of the haploid genome. Inhibitors of caveolae-dependent internalization abrogate the DNA uptake in Krebs-2 cells, and inhibitors of clathrin/caveolar mechanism block the internalization in EBV+ B-lymphoma cells.

*Conclusion:* It is shown for the first time that, in contrast to the majority of committed tumors cells, stem-like tumor cells of Krebs-2 and EBV+ B-lymphoma carry a general positive charge on their surface. It turns out that both proteoglycans/glycoproteins and complexes of scavenger receptors/glycosylphosphoinositol-associated proteins with caveolae or clathrin vesicles are the basic factors determining the interaction and internalization of dsDNA into TSCs. Thus, the basic provisions of the concept characterizing the principle of the organization of the stem-like tumor (and, possibly, some other poorly differentiated) cell surface factors to be formulated as a concept of “group-specific cell surface factors, which form the profile of tumor stem-like cells”. The results of bioinformatic and experimental research confirming the validity of the proposed concept are presented.

*Acknowledgements:* This research was funded by the Russian Ministry of Science and High Education via the Institute of Cytology and Genetics (State Budget Project No FWNR-2022-0016), LLC “KARANAHAN” and LLC “ACTIVATOR MAF”.

### *References*

1. Dolgova E.V. et al. Identification of cancer stem cells and a strategy for their elimination. *Cancer Biol Ther.* 2014;15(10):1378-1394.
2. Potter E.A. et al. Gene expression profiling of tumor-initiating stem cells from mouse Krebs-2 carcinoma using a novel marker of poorly differentiated cells. *Oncotarget.* 2017;8(6):9425-9441.
3. Dolgova E.V. et al. Non-adherent spheres with multiple myeloma surface markers contain cells that contribute to the sphere formation and are capable of internalizing extracellular double-stranded DNA. *Clin Lymphoma Myeloma Leuk.* 2016;16(10):563-576.
4. Dolgova E.V. et al. Evaluation of a strategy for tumor-initiating stem cell eradication in primary human glioblastoma cultures as a model. *Vavilov J Genet Breed.* 2018;22(7):825-836. (in Russian)
5. Dolgova E.V. et al. Efficacy of the new therapeutic approach in curing malignant neoplasms on the model of human glioblastoma. *Cancer Biol Med.* 2021;18(3):910-930.
6. Kisaretova P.E. et al. Approbation of the cancer treatment approach based on the eradication of TAMRA+ cancer stem cells in a model of murine cyclophosphamide resistant *lymphosarcoma*. *Anticancer Res.* 2020;40(2):795-805.

## TOP1, TDP1 and PARP1 inhibition: coupling and association

Dyrkheeva N.\*, Zakharenko A., Lavrik O.

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

\* *dyrkheeva.n.s@gmail.com*

**Key words:** DNA topoisomerase 1, tyrosyl-DNA-phosphodiesterase 1, Poly(ADP-ribose)polymerase 1, enzyme inhibition

DNA topoisomerase 1 (TOP1) is an enzyme that has evolved to resolve topological problems in DNA catalyzing DNA supercoil relaxation via transient single-stranded breaks in DNA. There are clinically used antitumor drug compounds that stabilize the TOP1-DNA covalent complex formed during the normal catalytic event. These TOP1 poisons, for example topotecan and irinotecan, are commonly used to treat ovarian, colon, and small cell lung cancers [1]. Tyrosyl-DNA-phosphodiesterase 1 (TDP1) is a DNA repair enzyme that purifies the 3'-end single-stranded breaks of DNA from covalent adducts of various origins, including those formed under the action of topotecan and irinotecan. Thus, TDP1 could be one of the factors of resistance of tumor cells to these drugs [2]. Poly(ADP-ribose)polymerase 1 (PARP1) is an enzyme that, through post-translational modification of poly(ADP-ribosylation), controls different processes in the cell, including DNA repair, maintenance of genome integrity, cell death, and others. Among these processes could be the repair of TOP1-DNA stalled complex caused by topotecan and irinotecan [3]. Olaparib is a PARP inhibitor indicated to treat breast cancer, ovarian cancer, fallopian tube cancer, peritoneal cancer, pancreatic cancer, and prostate cancer [4].

Our team discovered a number of TDP1 and PARP1 inhibitors among derivatives of natural biologically active compounds [5-7]. The anticancer effect of TOP1 inhibitors can be enhanced by the simultaneous inhibition of PARP1, and TDP1 [5, 8]. TDP1 inhibitors may sensitize tumor cells to the action of TOP1 poisons, thereby potentiating their effects. Some of the TDP1 inhibitors we found have a pronounced sensitizing effect on the cytotoxic effect of topotecan *in vitro* and antitumor *in vivo* [5].

*Acknowledgements:* The study is supported by the RSF (21-14-00105).

### References

1. Pommier Y. et al. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol.* 2010;17(5):421-433.
2. Comeaux E., Waardenburg R. 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.
3. Chowdhuri S.P., Das B.B. Top1-PARP1 association and beyond: from DNA topology to break repair. *NAR Cancer.* 2021;3(1):zcab003. doi: 10.1093/narcan/zcab003.
4. Slade D. PARP and PARG inhibitors in cancer treatment. *Genes Dev.* 2020;34(5-6):360-394.
5. Zakharenko A. et al. Dual DNA topoisomerase 1 and tyrosyl-DNA phosphodiesterase 1 inhibition for improved anticancer activity. *Med Res Rev.* 2019;(4):1427-1441.
6. Nilov D. et al. Inhibition of Poly(ADP-Ribose) Polymerase by Nucleic Acid Metabolite 7-Methylguanine. *Acta Naturae.* 2016;8(2):108-115.
7. Sherstyuk Y. et al. Design, synthesis and molecular modeling study of conjugates of ADP and morpholino nucleosides as a novel class of inhibitors of PARP-1, PARP-2 and PARP-3. *Int J Mol Sci.* 2019;21(1):214. doi: 10.3390/ijms21010214.
8. Das B.B. et al. PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage *Nucleic Acids Res.* 2014;(7):4435-4449.

# Bioorthogonality in the light of cell defense mechanisms: Interaction of non-natural nucleic acids with the base excision DNA repair system

Endutkin A.V.<sup>1\*</sup>, Yudkina A.V., Zharkov D.O.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *Aend@niboch.nsc.ru*

**Key words:** non-natural nucleotides, bioorthogonality, synthetic biology, DNA repair, DNA replication, translesion DNA synthesis

*Motivation and Aim:* The concept of bioorthogonality, which appeared in the field of synthetic biology, refers to any chemical reaction that can occur within living systems without interfering with natural biochemical processes. The modification of nucleic acids, especially DNA, in living cells is tightly controlled by DNA repair and DNA damage response systems. Any non-canonical nucleotide in DNA, with the exception of a very small number of natural modifications, is perceived as damage and removed. Thus, DNA repair, by definition, counteracts the bioorthogonal properties of modified DNA placed inside a living cell. However, although some types of non-natural nucleotides received certain attention, the activity of DNA repair and DNA damage response systems towards modifications used in synthetic biology has not been subject to systematic studies.

*Methods and Algorithms:* We have investigated the interaction of the base excision repair (BER) system enzymes with several classes of non-natural nucleotide units used in synthetic biology such as nucleotides used for click reactions (C8-alkyne-dU, C8-alkyne-dC, 5-ethynyl-dU, 1-ethynyl-AP site), nucleotides with alternative chirality ( $\beta$ -L-nucleotides of A, C, G, and T), and nucleotides forming modified internucleoside links (triazole linkage). The choice of these types of modifications is mostly guided by their widespread use as tools in synthetic biology or the steadily growing interest in such applications.

*Results:* We have studied the ability of human DNA glycosylases (and their homologues or functional analogs from *E. coli*) to remove bases from DNA containing these non-natural nucleotides, the ability of human and *E. coli* AP endonucleases to hydrolyze DNA containing them, and the ability of DNA polymerases to include correct and mismatches dNMP opposite to them. Generally, C5-modified pyrimidines were substrates for DNA glycosylases recognizing the products of 5-methylcytosine deamination, and 1-ethynyl-AP site was the substrate for AP endonucleases. The inverted chirality nucleotides were not substrates by themselves but blocked the action of DNA glycosylases that recognize both the damaged base and the opposite base.

*Conclusion:* The use of nucleic acids containing unnatural nucleotides to control and manipulate living cells is actively developing and will continue to develop in the coming decades. Therefore, the question of the degree of bioorthogonality of such molecules will become more and more urgent. We show that at least some unnatural nucleosides used in cell biology are subject to DNA repair, and the results obtained with such reagents should be interpreted with caution.

*Acknowledgements:* This work was supported by Russian Science Foundation (21-74-10104 to A.V.E.).

## Bacterial Argonaute proteins as sensors of DNA damage and repair

Esyunina D.<sup>1,2</sup>, Agapov A.<sup>1</sup>, Beskrovnaya M.<sup>1,3</sup>, Kropocheva E.<sup>1</sup>, Kuzmenko A.<sup>4</sup>, Lisitskaya L.<sup>1,2</sup>, Musabirov A.<sup>1</sup>, Oguienko A.<sup>1</sup>, Okhtienko A.<sup>1</sup>, Olina A.<sup>1</sup>, Aravin A.A.<sup>4</sup>, Kulbachinskiy A.<sup>1,2\*</sup>

<sup>1</sup> *Institute of Molecular Genetics, National Research Center “Kurchatov Institute”, Moscow, Russia*

<sup>2</sup> *Institute of Gene of Biology, RAS, Moscow, Russia*

<sup>3</sup> *Department of Molecular Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>4</sup> *Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, USA*

\* *avkulb@yandex.ru*

**Key words:** prokaryotic Argonaute proteins, DNA interference, DNA lesions, double-strand breaks

Prokaryotic Argonautes (pAgos) are programmable nucleases that participate in cell defense against invading genetic elements. They are evolutionary related to eukaryotic Argonautes (eAgos), which act as central components of RNA interference and are involved in silencing of foreign elements and gene regulation. Intriguingly, however, while all eAgos bind small RNA guides to recognize and silence RNA targets, all known pAgos preferentially target DNA. *In vitro*, pAgos can be programmed with small DNA or RNA guides to recognize and cleave complementary DNA targets. We show that nucleotide modifications and lesions in the DNA target can strongly affect the efficiency of its cleavage by pAgos, depending on the type of modification and its position relative to the site of cleavage. This potentially allows the use of pAgos as sensors of site-specific DNA damage. *In vivo*, pAgos preferentially target double-strand DNA ends and breaks and participate in their processing. We show that specific targeting of DNA breaks can be used for detection of DNA damage and repair at the whole-genome level. pAgos also target multicopy DNA elements including plasmids and phages, and can promote homologous recombination between plasmids and the chromosome, thus enabling genome editing in bacteria. Finally, we have discovered a new group of pAgos that use small DNA guides to recognize and cleave RNA targets. We show that these pAgos can sense modifications and lesions in RNA transcripts suggesting that they can find applications in RNA biotechnology.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grants 19-14-00359 and 22-14-00182) and the Russian Ministry of Science and Higher Education (grant 075-15-2021-1062).

## Interaction of triazinylamidophosphate-containing oligonucleotides with cells and DNA-binding proteins

Irina E.S.\*, Markov O.V., Kochetkova A.S., Zharkov T.D., Kuprushkin M.S., Lavrik O.I., Khodyreva S.N.

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

\* [ilina@niboch.nsc.ru](mailto:ilina@niboch.nsc.ru)

**Key words:** modified oligonucleotides, DNA-binding proteins

*Motivation and Aim:* A distinctive feature of the currently most widely applied anticancer therapy is the use of cytotoxic methods, namely radiation therapy and chemotherapy, which are accompanied by side effects that harm the patient's body. Increasing the effectiveness of the impact on tumor cells, which is not accompanied by nonspecific toxic effects on the body, as well as expanding the range of combination therapy, is an important and urgent task.

To develop strategies for synthesis of triazinylamidophosphate derivatives of oligonucleotides with different modifiers that may improve stability, cell uptake and interaction with key proteins of DNA repair.

*Methods and Algorithms:* A scheme for the introduction of a modified unit on a solid-phase support during the oxidation step with highly reactive 2-azido-4,6-dichloro[1, 3, 5]triazine according to the Staudinger reaction was proposed and implemented, followed by the insertion of various aliphatic residues upon treatment with the corresponding amines. Modifying reagents (azidotriazines) differed in the number of previously introduced functional groups into the triazine backbone. The most effective was the modifier with two acceptor substituents in the form of acid chloride groups [1].

To study the regularities of triazinylamidophosphate oligodeoxyribonucleotides (TAP-ODN) interaction with cells and target proteins within them, a panel of human cell lines including oncotransformed cells of different origins: HeLa, MCF-7, BJAB, T98G and immortalized non-cancer HEK293T cells were selected.

Affinity modification with chemically reactive DNAs containing TAP-ODN and either photoactivatable nucleotide at the 3'-end or 5'-deoxyribose phosphate residue (5'-dRP) or AP site was applied for search of target proteins in whole-cell extracts (WCE).

The degree of damage to genomic DNA as a result of exposure to genotoxic compounds was assessed by the comet method. TAP-ODN conjugates with fluorophores (Flu-TAP-ODN) were used to track the intracellular accumulation of these DNAs.

*Results:* It has been shown that Ku antigen (double strand break sensor) and PARP1 (single and double strand break sensor) are the main targets of DNA probes, with WCE being significantly different in the content of Ku and PARP1. This difference was confirmed by Western blot analysis. HEK293T and T98G, cell lines with high and low Ku antigen and PARP1 content, and WCE prepared from them were used in further experiments.

For Ku and PARP1 (both for purified proteins and in the extracts) the fundamental possibility of the formation of complexes with TAP-ODNs has been shown; moreover, the efficiency of protein binding to ODNs with the largest in volume substituent was significantly higher than for the native counterpart. It was found that, in comparison with



unmodified oligonucleotides, their triazinylamidophosphate derivatives exhibit increased resistance to the action of cellular nucleases, moreover this stability can be significantly increased by introduction of additional phosphorylguanidine (PG) groups. It was shown that TAP-ODNs and duplexes containing them act as activators of the synthesis of poly (ADP-ribose) catalyzed by PARP1, only moderately yielding to unmodified ODN counterparts.

It was found that Ku antigen and PARP1 play an important role in the protection of DNAs containing TAP-ODNs from hydrolysis by intracellular nucleases, which was confirmed by experiments with the addition of these proteins to the extract with a low concentration of its own Ku and PARP1.

By flow cytometry experiments, it was shown that the intracellular accumulation of Flu-TAP-ODNs is comparable to the transfection efficiency of the control oligonucleotides using the standard Lipofectamine 2000 transfectant for HEK293T and two times lower for T98G. For a set of oligonucleotides with different numbers of PG groups, it was shown that under standard transfection conditions (4 hours, without serum), the greatest penetration efficiency is observed for ODN with two PG groups, and with an increase in their number under the same transfection conditions, the penetration efficiency decreases.

When studying the localization of Flu-TAP-ODNs in the cells and the effect of PG groups on this parameter, it was shown that both oligonucleotides with and without these groups actually penetrate the cell membrane and are localized in the cytoplasm both within endosomal structures and diffusely.

During a series of kinetic experiments, it was shown that with an increased transfection time (8 and 24 hours), the efficiency of the ODN accumulation in the cells increases. The factors influencing the efficiency of Flu-TAP-ODNs penetration were determined, and a conclusion about the possibility of using these constructs for the delivery of ODN into cells was made.

The impact of TAP-ODNs on the cytotoxic effect of several DNA-damaging agents was evaluated on the example of HEK293T and T98G cells (the cell lines with high and low content of both Ku antigen and PARP1 respectively). It was shown that TAP-ODNs practically does not affect the cytotoxicity of temozolomide on HEK293T cells, while for T98G, a moderate increase in the cytotoxicity in presence of ODNs is observed. The slightly enhanced cytotoxicity of bleomycin and etoposide against T98G cells, but not HEK293T was also observed.

*Conclusion:* Thus, representatives of the class of TAP-ODNs derivatives were obtained for the first time. These TAP-ODNs are able to penetrate into cells without transfectant and interfere with cytotoxicity of temozolomide, bleomycin and etoposide with the influence being dependent of the amount of endogenous Ku antigen and PARP1 in the cells.

*Acknowledgements:* The study is supported by the RSCF (grant No. 19-14-00204).

## *References*

1. Markov A.V. et al. Antiviral activity of a new class of chemically modified antisense oligonucleotides against influenza a virus. *Russ J Bioorganic Chem.* 2019;45(6):773-781.

## Computational design of molecular probes targeting CD95 signaling pathway

Ivanisenko N.V.<sup>1, 2\*</sup>, Ivanisenko V.A.<sup>1, 2</sup>, Lavrik I.N.<sup>1, 3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Translational Inflammation Research, Medical Faculty, Otto von Guericke University Magdeburg, Magdeburg, Germany

\* n.ivanisenko@gmail.com

**Key words:** Apoptosis, cell death, virtual screening, molecular modelling, c-FLIP, molecular docking

**Motivation and Aim:** There are two types of apoptosis induction: intrinsic-mediated via mitochondria and extrinsic-mediated via death receptor (DR) activation. The induction of apoptosis via CD95 is largely controlled by the Death-Inducing Signaling Complex (DISC), which is formed upon CD95 stimulation. CD95 DISC comprises oligomerized, CD95, the adaptor protein FADD, procaspases-8/10 and cellular FLICE inhibitory proteins (c-FLIP). Deregulation of the CD95 pathway is associated with multiple disease development, including cancer and neurodegenerative diseases. Computational design of molecular probes targeting CD95 platform components is of high interest to study detailed molecular mechanisms of CD95 molecular platform organization and development of new treatment approaches.

**Methods and Algorithms:** Virtual screening of small molecule compounds was done using GLIDE software. Molecular modeling was carried out using Rosetta and AlphaFold software. Mathematical modeling of CD95 pathway activation was carried out by solving systems of ODE with Scipy module in Python.

**Results:** Molecular models of the CD95 platforms for cell death initiation were reconstructed *in silico*. Small molecule compounds targeting major components of CD95 platform were predicted using virtual screening approach of large databases of chemical compounds and validated experimentally. Kinetic mathematical model was further developed to analyze the observed effects of small molecule compounds on DISC activation.

**Conclusion:** The developed models suggest the significant role of c-FLIP in regulating the cell death activation, molecular platform stoichiometry and activity of developed small molecules compounds.

**Acknowledgements:** Work 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.

# Transcriptional mutagenesis-based reporter system for the analysis of DNA repair

Kim D.V.<sup>1,2\*</sup>, Khobta A.<sup>3</sup>, Zharkov D.O.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Institute of Nutritional Sciences, Friedrich Schiller University Jena, Jena, Germany*

\* *kim.daria.nsk@gmail.com*

**Key words:** base excision repair, knockout, reporter system

*Motivation and Aim:* Base excision repair (BER) is an important cellular process responsible for removing non-bulky DNA base modifications. The BER proficiency of human cells is commonly assessed either by analysis of cell extracts or using damaged plasmid DNA but these approaches have rather low sensitivity and specificity. As an alternative, the plasmids with defined positioned DNA lesions could be used [1].

*Methods and Algorithms:* The knockout of BER genes *APEX1*, *OGG1*, and *MUTYH* were generated using genome editing technology CRISPR/Cas9 from the parental 293FT cells. To detect the DNA repair in cells we use the assay based on plasmids containing single nucleotide substitution in the *EGFP* gene, which results in a loss of fluorescence [2]. The lesion is introduced in the position of the substitution. Inability to repair the DNA lesion in the transcribed strand leads to a heterogeneous mRNA pool with different ribonucleotides positioned opposite to the lesion (“transcriptional mutagenesis”). The evaluation of a cell’s fluorescence by cell cytometry provides information on the level of aberrant transcripts and thus on the repair of the original lesion in DNA.

*Results:* Constructs contacting synthetic analogous of abasic sites were used to detect the repair in cells deficient in the main apurinic/apyrimidinic endonuclease *APEX1*. Despite the *APEX1* knockout, the repair efficiency of this lesion was considerable, which could indicate the presence of alternative repair pathways. We also examined the repair of 8-oxoguanine in *OGG1* and *MUTYH* knockout cells. 8-oxoguanine is efficiently repaired by *OGG1* when paired with cytosine, and adenine is repaired by *MUTYH* when paired with 8-oxoguanine. Base pair 8-oxoguanine:adenine is non-canonical and arises during the replication.

*Conclusion:* The EGFP reporter system can efficiently be used to assess the repair of DNA base damages in different knockout cell lines. Besides, this system should allow discovery of alternative repair mechanisms.

*Acknowledgements:* The research was supported by Russian Foundation for Basic Research Graduate Student Fellowship (20-34-90092).

## References

1. Lühnsdorf B. et al. Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal Biochem.* 2012;425(1):47-53.
2. Rodriguez-Alvarez M., Kim D., Khobta A. EGFP reporters for direct and sensitive detection of mutagenic bypass of DNA lesions. *Biomolecules.* 2020;10(6):902.

## Activity of natural polymorphic variants G118V and R149I of DNA polymerase $\beta$

Kladova O.A.<sup>1\*</sup>, Soloviev N.O.<sup>2</sup>, Mikushina E.S.<sup>1</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* *kladova@niboch.nsk.ru*

**Key words:** DNA repair, protein-nucleic acid interactions, DNA polymerase beta, single nucleotide polymorphisms

DNA polymerase  $\beta$  (Pol $\beta$ ) is one of the main DNA polymerases involved in the base excision repair pathway, capable of filling single nucleotide gaps in DNA. It is known that functionally deficient Pol $\beta$  mutants may have a reduced efficiency of DNA repair, which can lead to an increased occurrence of mutations in the genome. At the moment, significant amount of studies showed the association of single nucleotide polymorphisms with various cancers.

In this study we analyzed the natural mutations located in DNA-binding domain of Pol $\beta$ . The analysis was carried out using software tools (SIFT, PolyPhen, CADD, REVEL, MetaLR, Provean) predicting whether an amino acid substitution will have an impact on the biological function of a protein. During the analysis, two natural polymorphic variants Gly118Val and Arg149Ile were selected with a high probability of affecting the protein functioning. Then, mutant forms of Pol $\beta$  were obtained by site-directed mutagenesis and purified to experimentally analyze activity of these polymorphic variants.

The EMSA analysis revealed that both mutant forms have reduced ability to bind single nucleotide gapped DNA compare to WT Pol $\beta$ . Using stopped flow analysis, it was shown that Gly118Val mutant form binds dNTP several time less effectively then WT Pol $\beta$  and Arg149Ile mutant form. PAGE analysis of enzyme activity revealed that both mutant forms have a reduced ability to fill the single nucleotide gapped DNA compared to the WT Pol $\beta$ .

Obtained data allow to conclude that single amino acid mutations Gly118Val and Arg149Ile mutations located in DNA-binding domain of Pol $\beta$  significantly influence on the enzyme catalytic activity and interaction with DNA.

*Acknowledgements:* This work was supported by Russian Science Foundation grant No. 21-74-10103.

## **Combined treatment with small molecules targeting c-FLIP and Mcl-1 promotes apoptosis and necroptosis in pancreatic cancer cells**

König C.<sup>1\*</sup>, Hillert-Richter L.<sup>1</sup>, Ivanisenko N.<sup>2</sup>, Lavrik I.<sup>1</sup>

<sup>1</sup> *Translational Inflammation Research, Medical Faculty, Otto von Guericke University Magdeburg, Magdeburg, Germany*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *corinna.koenig@st.ovgu.de*

There are three isoforms of c-FLIP reported to be recruited to the death inducing signaling complex (DISC): long (c-FLIPL), short (c-FLIPS) and raji (c-FLIPR). While c-FLIPS and c-FLIPR have an anti-apoptotic role, c-FLIPL can act in both anti- and pro-apoptotic manner depending on its concentration. We designed a small molecule called FLIPinB which interacts with c-FLIPL to stabilize the active center of caspase-8 within the caspase-8/c-FLIPL heterodimer and increases caspase-8 activity. We concentrated on pancreatic cancer cells, SUIT-020, which are resistant towards single treatments. To further enhance the pro-apoptotic effects of FLIPinB we carried out combinatorial treatments with the Mcl-1 inhibitor S63845 and gemcitabine. Our previous studies showed that moderate concentrations of gemcitabine have only small effects on SUIT-020 cells. Here we show that combinatorial treatment with CD95L/TRAIL, gemcitabine, FLIPinB and S63845 leads to an increased loss of cell viability. Moreover, this increased cell viability loss was accompanied by enhanced caspase activity. Further, we have shown that this combinatorial treatment might induce both apoptotic and necroptotic pathways via an increase in the amounts of complex IIa and complex IIb after combined treatment. Taken together, our study suggests new approaches for the development of novel cancer therapies targeting apoptosis.

## Usnic acid derivative antitumor, antimetastatic, and hematoprotective effects in the Lewis carcinoma model

Kornienko T.E.<sup>1\*</sup>, Zakharenko A.L.<sup>1</sup>, Filimonov A.S.<sup>2</sup>, Luzina O.A.<sup>2</sup>,  
Salakhutdinov N.F.<sup>2</sup>, Lavrik O.I.<sup>1</sup>

<sup>1</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

<sup>2</sup> N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia

\* t.kornienko1995@gmail.com

**Key words:** tyrosyl-DNA phosphodiesterase 1, inhibitor of tyrosyl-DNA phosphodiesterase 1, DNA repair, usnic acid, Lewis lung carcinoma

*Motivation and Aim:* The search and development of inhibitors of DNA repair pathways is one of the main priorities of medical chemistry, since DNA repair inhibitors can ensure effective treatment of oncological diseases. Tyrosyl-DNA phosphodiesterase 1 (Tdp1) is a promising molecular target because it participates in the repair of several types of DNA damage caused by anticancer drugs widely used in clinic [1]. Thus, Tdp1 inhibitors can increase the effectiveness of treatment of various types of cancers. Olaparib is an effective clinically approved poly(ADP-ribose) polymerase 1 (PARP1) inhibitor. Its mechanism of action is based on preventing the synthesis of poly-ADP-ribose, thereby blocking the response to DNA damage. When used as monotherapy and in combination with conventional chemotherapy drugs, it inhibits the growth of certain tumor cell lines *in vitro* and tumor growth *in vivo*. Poly(ADP-ribose)polymerase 1 (PARP1) is one of the main regulators of the repair of single-stranded DNA breaks, thus playing an important role in maintaining genome stability and cell survival. It is known that PARylation of Tdp1 enhances its recruitment to sites of DNA damage without affecting the catalytic activity of Tdp1 [2]. PARP1 stimulates Tdp1 enzymatic activity at AP-sites. Thus, the combined use of PARP1 and Tdp1 inhibitors may be a promising therapeutic strategy for the treatment of various cancers. The development of new inhibitors of DNA repair enzymes based on natural compounds and their derivatives is particularly relevant, since such compounds often have complementarity to targets of biological origin.

A promising biologically active compound is usnic acid. It is a secondary metabolite of various lichen species with a wide range of biological activity (anti-inflammatory, antiviral, antitumor, etc.) [3]. By itself, usnic acid is quite toxic. Modifications of its structure, developed by our team, provide a reduction in toxicity and effective binding to the active center Tdp1. The derivative with chemical modifications of rings A and C was synthesized on the basis of usnic acid. The introduction of cyanogroups attached via long aliphatic linkers ensures binding to the enzyme, and the introduction of a pyrazole fragment into the C ring reduces the toxicity of the compound compared to the original usnic acid (toxicity is absent in the whole range of studied concentrations to 100  $\mu$ M) [3]. According to the results of screening of the inhibitory activity of this compound with respect to purified recombinant Tdp1, it can be noted that it has a high inhibitory activity ( $IC_{50} = 2.9 \pm 0.8 \mu$ M) [3, 4].

*Methods and Algorithms:* In the research, we used female mice of the C57BL/6 line (35 individuals) weighing 19–21 g. We transplanted an experimental Lewis lung carcinoma tumor model into mice intramuscularly. The animals were divided into five

groups: the control group without treatment; the group that was injected with vehicle (DMSO + Tween-80); the group that was injected with olaparib; the group that was injected with olaparib and Tdp1 inhibitor; the group that was injected with Tdp1 inhibitor only. The effect of the drugs was evaluated at the end of the experiment by the weight and size of the tumor, the quantity of metastases, as well as by the weight of the liver and spleen. Moreover, we calculated erythrocyte count, leukocyte count and leukocytic formulas.

*Results:* The derivative of usnic acid has a more pronounced antitumor and antimetastatic effect compared to olaparib and their combination. The Tdp1 inhibitor had no toxic effect on the internal organs of mice. The injection of the Tdp1 inhibitor leads to the normalization of hematopoiesis. At the end of the experiment blood levels were close to normal values.

*Conclusion:* This usnic acid derivative is promising prototype for anticancer drugs development. It is a non-toxic compound and has pronounced antitumor, antimetastatic and hematoprotective activity.

*Acknowledgements:* The work was supported by the Russian Scientific Foundation grant No. 21-14-00105.

### 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(4):494-507.
2. Das B.B., Huang S.Y., Murai J., Rehman I., Amé J.C., Sengupta S., Das S.K., Majumdar P., Zhang H., Biard D. et al. PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage. *Nucleic Acids Res.* 2014;42:4435-4449.
3. Zakharenko L., Luzina O.A., Sokolov D.N., Zakharova O.D., Rakhmanova M.E., Chepanova A.A., Dyrkheeva N.S., Lavrik O.I., Salakhutdinov N.F. Usnic acid derivatives are effective Inhibitors of Tyrosyl-DNA phosphodiesterase I. *Russ J Bioorganic Chem.* 2017;43(1):84-90.
4. Luzina O.A., Salakhutdinov N.F. Biological activity of usnic acid and its derivatives: Part 2. effects on higher organisms. Molecular and physicochemical aspects. *Russ J Bioorganic Chem.* 2016;42:249-268.

## Non-thermal atmospheric plasma in cancer therapy: cell death mechanisms and immunity

Koval O.<sup>1,2\*</sup>, Birykov M.<sup>1,2</sup>, Patrakova E.<sup>1,2</sup>, Milakhina E.<sup>3</sup>, Gugin P.<sup>3</sup>, Zakrevsky D.<sup>3</sup>, Schweigert I.<sup>4</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *A.V. Rzhanov Institute of Semiconductor Physics, SB RAS, Novosibirsk, Russia*

<sup>4</sup> *Khrstianovich Institute of Theoretical and Applied Mechanics, SB RAS, Novosibirsk, Russia*

\* [o\\_koval@ngs.ru](mailto:o_koval@ngs.ru)

**Key words:** cold atmospheric plasma, RNA-Seq, immunogenic cell death, dendritic cells

*Motivation and Aim:* Along with developing efficacious chemotherapeutic agents to treat malignant neoplasms, much attention is currently being paid to physical methods. Cold atmospheric plasma (CAP) has been shown as potential anticancer tool. CAP is an ionized gas consisting of charged particles, active uncharged particles, an electric field and UV radiation. The CAP jet is visually composed by several millions of streamers per a minute propagating from the powered electrode to the target direction with the speed of 5–20~km/c. High energy plasma electrons in the streamers head produce chemically active oxygen (ROS) and nitrogen (RNS) species in the mixture of the inert gas with air in vicinity of the bio-target surface. These active compounds are known to be the main CAP cytotoxic affecters. We have shown previously that CAP induced death of cancer cells with the hallmarks of immunogenic cell death (ICD): calreticulin and heat shock protein 70 (HSP70) externalization and high-mobility group box 1 protein (HMGB1) release [1]. The present study aimed to address how CAP treatment changes the morphology of tumor cells, affects their transcription activity, and finally stimulates cell death or survival. Moreover, we analyzed if plasma-treated cells potent to stimulate dendritic cells maturation for the forming long-term antitumor immunity.

*Methods and Algorithms:* In our experiments, the source of the plasma jet was a coaxial dielectric channel with an inner diameter of 8 mm with a capillary nozzle with a diameter of 2.3 mm. The models used in this study were pulmonary-originated cell lines: lung adenocarcinoma cells A549 and Wi-38 normal lung fibroblasts. Differential gene expression was assessed by mass parallel sequencing on the Illumina NextSeq platform (RNA-Seq). Functional analyses were employed to reveal the difference in normal and cancer cell response. Validation of the differential expression of certain genes was performed by RT-PCR.

*Results:* The CAP irradiation conditions were optimized to kill A549 cancer cells and stay alive Wi-38 cells. Under these conditions the difference in genes expression was analyzed in A549 and Wi-38 irradiated cells. We identified universal indicator RNAs whose expression is modulated differently by CAP treatment for healthy and tumor cells. The substantial differences in expression was noted for genes of transcription factors, regulating RNA polymerase II activity. Despite the fact that CAP exposure led to the appearance of ICD markers in treated cancer cells *in vitro*, such cells have stimulated only weak maturation of dendritic cells.



*Conclusion:* The data obtained could be a basis for the development of selective CAP treatment of solid cancers in patients. This study partly explains the mechanisms of selective action of CAP towards tumor cells. The finding the conditions of CAP treatment under whose dying cells will stimulate the dendritic cell maturation can be considered an important task for future research

*Acknowledgements:* The study is supported by the RSF grant No. 19-19-00255-II.

#### *References*

1. Troitskaya O.S. et al. Non-thermal plasma application in tumor-bearing mice induces increase of serum HMGB1. *IJMS*. 2020;21:5128-5142.

## NER additional players at the repair late stages

Krasikova Y.S.<sup>1\*</sup>, Rechkunova N.I.<sup>1</sup>, Lavrik O.I.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *Y.Krasikova@gmail.com*

**Key words:** DNA repair, BER, NER, XPA, FEN1, XRCC1

*Motivation and Aim:* Nucleotide excision repair (NER) is the most versatile DNA repair pathway, which can remove diverse bulky DNA lesions destabilizing a DNA duplex. NER substrates are UV photoproducts, e.g., cyclobutane pyrimidine dimers (CPDs), pyrimidine-pyrimidone-(6-4)-photoproducts (6-4PPs), intrastrand crosslinks, and bulky adducts of DNA bases with reactive metabolites of some chemical carcinogens or chemotherapeutic agents. These kinds of lesions can be substrates for two NER sub-pathways—global genome NER (GG-NER) and transcription-coupled NER (TC-NER)—that overlap, except for the mode of DNA damage recognition. After the lesion has been recognized, a pre-incision complex is formed. Subsequent NER stages includes excision oligonucleotide containing lesion with ssDNA gap generating, gap filling and sealing the nick.

The DNA incision is first carried out by XPF–ERCC1 from the 5' side to the damage site. Formed DNA intermediate contains long gap (about 30nt) with a free 3'-OH group, long flap bearing a lesion and an intact template. Next, replication machinery loads to start repair synthesis. Possible sets of replication machines: (DNA polymerase  $\delta$ / PCNA/ RFC1-RFC-p66), (DNA polymerase  $\epsilon$ / CTF18/RFC) or (DNA polymerase  $\kappa$ / ubiquitinated PCNA/ XRCC1). Which set of replication factors are definitely recruited remains unknown, but generally it depends on the status of PCNA ubiquitination as well as on usage of distinct clamp loader complexes or XRCC1. Repair synthesis can proceed halfway through the gap. Then the XPG-made incision cleaves the flap and leaves a 5'-phosphate that is utilized in the nick-sealing reaction by DNA ligase I or by complex (DNA ligase III $\alpha$ / XRCC1).

We have shown previously that XPA – the key NER protein factor – could stay in NER complex at the post-incision stages through its interaction with RPA, the major eukaryotic ssDNA-binding protein [1]. These data supported by the existed information that XPA interacts with PCNA. To further understand possible XPA involvement to the late NER stages, we analyze this protein interactions with two proteins that also could form stable complexes with PCNA: XRCC1 and FEN1.

The XRCC1 (X-ray cross complementing group 1) also plays important roles in the different repair pathways outside the NER. In general, XRCC1, same as XPA, is a scaffold protein so it has ability to interact with multiple protein partners. Furthermore, same as XPA, XRCC1 has intrinsic DNA-binding activity with unknown functional meaning. FEN1 (flap endonuclease 1) belongs to the same as XPG structure-specific nuclease family and key component of all replication machines where flap DNA structures could arise. To further explore possible FEN1 and XRCC1 involvement in an incision/resynthesis stages coordination we analyze protein-protein interactions between

XPA, FEN1, XRCC1 and PCNA and these protein factors interactions with DNA mimicking the late NER stages DNA intermediates.

*Methods and Algorithms:* We have used 60-mer DNA duplexes containing a 31 nt flap with the fluorescein substituted dUMP residue mimicking the bulky lesion and a 26, 10 or 3 nt gap. The structure containing a damaged flap and a 26 nt gap imitates the DNA intermediate arising in the NER process after damaged strand cleavage by XPF-ERCC1; DNA with the same flap and a 10 nt gap can be attributed to the intermediate of the subsequent partial gap filling. Protein-DNA interactions were analyzed by gel retardation and Flu-dUMP residue were used as a label for detection. Two methods were used to detect protein-protein interactions: a dot-blot analysis and detection complexes formed by fluorescently labeled protein (FAM-RPA, FAM-XPA or FAM-XRCC1). XPA and XRCC1 influence on the FEN1 enzymatic activity also analyzed.

*Results:* XRCC1 binds to DNA containing damaged flap and gap and Y-type structures with nano molar range of Kd. Thus, we performed gel retardation experiments with these DNAs and bands corresponded to ternary complexes XRCC1-XPA-DNA or XRCC1-FEN1-DNA were detected. The XRCC1-XPA-DNA ternary complex detected with all DNA structures except one: duplex containing 5'-protruding 31nt ss tail. XPA-FEN1-DNA ternary complex also registered with all DNA structures used in experiments. As expected, FEN1 effectively binds and catalyze flap incision of DNA containing flaps and different size gaps. XRCC1 and XPA are slightly inhibit FEN1 catalysis at concentration corresponded to the ternary complex formation. Using dot blot analysis and experiments with fluorescently labeled proteins, XPA-FEN1 and XPA-XRCC1 complexes existence confirmed.

*Conclusion:* Our findings expand protein participants of the last NER stages and allow us to propose XPA-FEN1-XRCC1-PCNA complex formation. Functional role of this complex stay to be unknown and we assume that future experiments will shed the light on it.

*Acknowledgements:* The study is supported by the RSF (19-74-10056).

### *References*

1. Krasikova Y.S. et al. RPA and XPA interaction with DNA structures mimicking intermediates of the late stages in nucleotide excision repair. *PLoS One*. 2018;13(1):e0190782. doi: 10.1371/journal.pone.0190782.

## HPF1 can promote opposite effects on different stages of PARP1 and PARP2 autoPARylation and histone modification

Kurgina T.A.<sup>1, 2</sup>, Moor N.A.<sup>1</sup>, Kutuzov M.M.<sup>1</sup>, Naumenko K.N.<sup>1</sup>, Ukraintsev A.A.<sup>1</sup>, Lavrik O.I.<sup>1, 2</sup>

<sup>1</sup>*Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* *t.a.kurgina@gmail.com*

**Key words:** DNA repair, PARP1, PARP2, HPF1

Poly(ADP-ribose) (PAR) composed of linear and/or branched repeats of ADP-ribose is synthesized by enzymes called poly(ADP-ribose) polymerases (PARPs) that utilize NAD<sup>+</sup> as a substrate to modify themselves (automodification) or target molecules (heteromodification) [1, 2]. Poly(ADP-ribosyl)ation (PARylation) catalyzed by polymerases (PARPs) is one of the immediate cellular responses to DNA damage.

The best studied PARP family member, PARP1, is the abundant nuclear protein involved in multiple cellular processes, among them, DNA repair regulation. PARP1 serves as a sensor of the DNA lesions usually induced by ionizing irradiation and oxidative stress and signals to recruit the appropriate proteins to the sites of damage [3, 4]. PARP2 was also discovered as an enzyme that catalyzes the synthesis of PAR [5]. PARP2 shares significant homology with PARP1 in the catalytic domain structure, but differs in the domain architecture due to the absence of zinc fingers and BRCT domain<sup>8</sup>. Roles of PARP1 and PARP2 in base excision repair and single-strand break repair were intensively studied. Overlapping functions of PARP1 and PARP2 in regulation of these processes were shown by using model DNA duplexes as well as nucleosomes [6, 7]. PARP2 synthesizes shorter PAR chains than PARP1 does and also inhibits PARP1-catalyzed PAR synthesis via hetero-oligomerization of the two PARPs [8]. It has been shown that removal of histone H3 is largely suppressed in PARP2-deficient cells<sup>21</sup>. Thus, PARP1 and PARP2 can perform both overlapping and specific functions in maintaining genome stability, and these enzymes cooperate in the regulation of cellular processes.

A new histone PARylation factor (HPF1) which modulates PARP1/2 activity was discovered recently. HPF1 switches the PARP1/2 PARylation specificity to serine residues by completing their active sites. This joint active site of HPF1-PARP complex involved in the PARylation of histones [9]. It was shown that HPF1 enhances NAD<sup>+</sup>-hydrolysis catalyzed by PARP1 and leads to the shortening of synthesized PAR. However, the general picture of HPF1 interaction with PARP only begins to clear up. In the presented study, we demonstrate for the first time that HPF1 can stimulate the autoPARylation of PARP1/2 and the heteroPARylation of histones in the context of nucleosomes. Interestingly, stimulation is promoted by the incomplete serine-specificity switch. We have shown that only a large excess of HPF1 over PARPs can switch the activities to NAD<sup>+</sup> hydrolysis. HPF1 more efficiently stimulates PARP2, compared with PARP1 and the effect of stimulation is dependent on the structure of DNA damage. We suggest a specific role of PARP2 in the ADP-ribosylation-dependent modulation of chromatin structure. We provide evidence that HPF1 can promote opposite effects on

different stages of the PARP1/2-catalyzed reaction: it stimulates early stages and inhibits elongation by shielding of amino acid residues important for PAR chain elongation [9]. The dual regulatory function of HPF1 may contribute to maintenance of PARP activity at the level required for its function in the DNA damage response signaling and repair, independently on the NAD<sup>+</sup> concentrations. The suppression is most likely predominant function of HPF1 at the normal NAD<sup>+</sup> concentration in the nucleus (100 μM) to prevent negative consequences of PARP hyperactivation. The stimulatory action of HPF1 might regulate PARP activity under stressed conditions, when the sustained PARP activity leads to NAD<sup>+</sup> depletion [10].

*Acknowledgements:* The work was supported by grants from RSF 21-64-00017 and RFBR 20-34-90095.

### *References*

1. Crawford K., Bonfiglio J.J., Mikoc A., Matic I., Ahel I. Specificity of reversible ADP-ribosylation and regulation of cellular processes. *Crit Rev Biochem Mol Biol.* 2018.
2. Spiegel J.O., Van Houten B., Durrant J. D. PARP1: Structural insights and pharmacological targets for inhibition. *DNA Repair (Amst).* 2021.
3. Schreiber V., Dantzer F., Ame J.C., de Murcia G. Poly(ADP-ribose): novel functions for an old molecule. *Nat Rev Mol Cell Biol.* 2006.
4. Pascal J., Ellenberger T. The rise and fall of poly(ADP-ribose): An enzymatic perspective. *DNA Repair (Amst).* 2015.
5. Amé J. et al. PARP-2, a novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. *J Biol Chem.* 1999.
6. Sukhanova M., Khodyreva S., Lavrik O. Poly(ADP-ribose) polymerase 1 regulates activity of DNA polymerase beta in long patch base excision repair. *Mutat Res.* 2010.
7. Kutuzov M.M. et al. The contribution of PARP1, PARP2 and poly(ADP-ribosylation) to base excision repair in the nucleosomal context. *Sci Rep.* 2021.
8. Vasil'eva I., Moor N., Anarbaev R., Kutuzov M., Lavrik O. Functional roles of PARP2 in assembling protein-protein complexes involved in base excision DNA repair. *Int J Mol Sci.* 2021.
9. Suskiewicz M.J. et al. HPF1 completes the PARP active site for DNA damage-induced ADP-ribosylation. *Nature.* 2020.
10. Kurgina T.A. et al. Dual function of HPF1 in the modulation of PARP1 and PARP2 activities. *Commun Biol.* 2021.

## PARP2 and BER in the nucleosomal context

Kutuzov M.\*, Belousova E., Kurgina T., Ukraintsev A., Vasil'eva I., Khodyreva S., Lavrik O.

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

\* *kutuzov.mm@mail.ru*

**Key words:** PARP2, PARP1, nucleosome, base excision repair, poly(ADP-ribosyl)ation, APE1, DNA polymerase  $\beta$ , DNA ligase III $\alpha$ , XRCC1

*Motivation and Aim:* The most often existing damage types such as oxidative DNA lesions or spontaneously appeared AP sites are repaired via base excision repair, or BER. This system includes the following main stages: DNA damage recognition, excision of the damaged base, incision of the sugar-phosphate backbone, gap processing (including the incorporation of dNMP) and DNA ligation, which catalysed by specific glycosylases, AP endonuclease 1, DNA polymerase  $\beta$  (Pol $\beta$ ) and DNA ligase III $\alpha$  (LigIII $\alpha$ ) in complex with XRCC1 respectively.

In eukaryotic cell the metabolism of poly(ADP-ribose) (PAR) in the nucleus is carried out by nuclear PARPs – PARP1 and PARP2. PARP1 is the most abundant and active PARP; it's strongly activates in response to DNA damage and synthesises most of PAR after genotoxic or oxidative stress. Mice devoid of PARP1 or PARP2 are viable but exhibit sensitivity to genotoxic agents. Nonetheless, double deletion of PARP1 and PARP2 causes early embryonic lethality in mice, indicating that each of these two proteins is involve in DNA repair and can partly compensate for the absence of the other one. As for biochemical data, the protein-protein interaction pattern of PARP1 has been well estimated quantitatively. In particular, it efficient interacts with main repair enzymes, including BER system, and with histone and non-histone chromatin proteins and XRCC1 by itself or through the PAR molecule.

As for PARP2, its role in the repair processes is described to a much lesser extent and the interactions of PARP2 with the BER proteins have not been estimated in details. So, the final function of PARP2 in this process is not clear. Additionally, it is interesting how BER is regulated by the nuclear PARPs and how the roles are distributed between different PARPs.

The most part of these data were obtained using naked DNA substrates, while human genome is presented by the chromatin with nucleosomes as a lowest level of DNA compaction. Our work aimed to obtain a biochemical data set concerning the influence of PARP2 and PARylation in comparison with PARP1 at different stages of BER during the repair process in the context of NCP. Based on all of the data, we proposed a model of BER regulation by PAR in the nucleosomal context.

*Methods and Algorithms:* To determine the contribution of the interaction of PARPs with DNA or histone components of an NCP in the BER process, we evaluated the affinity of the proteins for native, AP-NCP or gap-NCP by fluorescence measurements and EMSA. Additionally, we evaluated the activation efficacy of both proteins under PARylation. Then, we analyzed the activity of main BER enzymes – APE1, Pol $\beta$ , LigIII $\alpha$  in complex with XRCC1, in the presence of PARPs and under PARylation in the nucleosomal context.

*Results:* It was found that whereas PARP1 suppresses the activity of APE1, Pol $\beta$  and to a smaller extent LigIII $\alpha$ , PARP2 exerts a significant effect on the activity of LigIII $\alpha$  only. The decreases in the activities of APE1 and Pol $\beta$  are reversed in the presence of the PARylation catalyzed by PARP1 and to a lesser extent by PARP2. Meanwhile, the complementation with XRCC1 protects the dNMP transferase activity of Pol $\beta$  from the inhibition by PARP1 and PARP2. Moreover, this complementation slightly affects the intensity of DNA synthesis under the PARylation catalysed by PARP1 and leads to a decrease of Pol $\beta$  activity in case of PARylation by PARP2, in a NAD<sup>+</sup>-dependent manner. A remarkable result was obtained in the assay of DNA sealing by LigIII $\alpha$ . This reaction is stimulated by the PARylation catalysed by PARP2. It is worth mentioning that outward-oriented lesions are more accessible to all the enzymes, and their processing requires a lower enzyme concentration and PAR level. More PAR is needed for the repair of more hidden damage sites.

*Conclusion:* Therefore, all obtained results and literature data indicate functional dependence of various repair enzymes' activities on PARylation and PAR synthesis and can be described by the following hypothetical model of the regulation of the basal BER process in the absence of massive DNA damaging exposure. DNA damage triggers the activation of a signalling pathway through PARP1. In this process, some regions of compacted chromatin could be converted to a partially de-compacted state. PARylation enables the dissociation of histone and non-histone chromatin proteins and involves specific repair proteins. The main role of PAR at this stage is to loosen chromatin structure and to attract BER proteins such as scaffold protein XRCC1 and possibly some key proteins such as APE1 and others. The cleavage of an AP site by APE1 leads to additional activation of PARP1 and PARylation resulting finally in the dissociation of PARP1. It makes way for subsequent stimulation of the BER process by the removal of APE1 from its product by Pol $\beta$ . The weight of evidence suggests that the functions of PARP1 and PARP2 probably overlap partially after the AP site cleavage and supports a role of PARP2 in the regulation of BER activity. We believe that the contribution of PARP2 to BER becomes more specific starting from this point and begins to increase. The presence of a specific single-strand break DNA structure as well as PAR synthesised by PARP1 attracts PARP2, which in turn synthesises additional PAR molecules. PAR can regulate the turnover of BER proteins at various stages of the process. PARP1 and PARP2 regulate Pol $\beta$  activity, whereas PARP2 stimulates LigIII $\alpha$  activity to complete the repair process, indirectly through their interactions with XRCC1 and PAR. The synthesis of PAR that is catalysed by PARP1 and PARP2 during the interaction with damaged DNA helps the Pol $\beta$ -XRCC1 complex to complete its action, and PARP2 providing PARylation stimulates the activity of LigIII $\alpha$ -XRCC1 towards the NCP. The higher PAR level leads to a breakdown of the specific ligase complex, and repair is thus accomplished. We propose that the PAR synthesised at the initial stages by PARP1, and at the final steps also by PARP2, can act as a regulatory factor for the step-by-step progression of the BER process.

*Acknowledgements:* The study was supported by the Russian Foundation for Basic Research (No. 20-04-00674).

# **A conserved mechanism of damage recognition by apurinic/apyrimidinic endonucleases from different structural families**

Kuznetsov N.A.

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

*Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

*Nikita.Kuznetsov@niboch.nsc.ru*

Apurinic/apyrimidinic (AP) endonucleases process intermediate DNA products generated by DNA glycosylases in the course of base excision repair (BER) and cleave an AP site. Data obtained over last two decades clearly indicate that members of the different structural families of AP endonucleases can recognise as a substrate not only an AP site but also various damaged nucleotides. Molecular origin of the wide substrate specificity of AP endonucleases, which can cleave damaged and undamaged DNA in endo- and exonuclease types of interaction, is not clear at present. Significant differences in the structure of target nucleotides support the idea that the recognition of a specific site by the enzyme is determined by some conformational properties of DNA at the lesion site.

Here, the mechanism of target nucleotide recognition by Nfo enzyme from *Escherichia coli* was analysed by pulsed electron–electron double resonance spectroscopy and pre–steady-state kinetic analysis with Förster resonance energy transfer detection of DNA conformational changes during DNA binding. To elucidate the stages of catalytic-complex formation and to reveal key factors responsible for inner-strand damage recognition, for discrimination of structurally different damaged nucleotides and for cleavage efficacy, a set of model DNA substrates was tested. Obtained results strongly support the notion that the recognition of diverse lesions by AP endonuclease Nfo depends on the ability of a damaged nucleotide to change the enzyme’s DNA-bending and DNA-unwinding ability. Comparison of the present data and data obtained previously for human APE1 indicates that this strategy of damage recognition is suitable for AP endonucleases from both structural families.

The results for the first time show that AP endonucleases from different structural families utilise a common strategy of damage recognition, which globally may be integrated with the mechanism of searching for specific sites in DNA by other enzymes.

*Acknowledgements:* This work was supported by Russian Science Foundation Grant No. 21-64-00017.



# The pre-steady state analysis of human terminal deoxynucleotidyltransferase conformational dynamics under DNA synthesis

Kuznetsova A.A.\* , Fedorova O.S., Kuznetsov N.A.

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

\* *sandra-k@niboch.nsc.ru*

Terminal deoxynucleotidyltransferase (TdT) is a member of the DNA polymerase X family that is responsible for random addition of nucleotides to single-stranded DNA at V-D and D-J junctions in genes of heavy chains of immunoglobulins and T-cell receptors. We examined pre-steady-state kinetics and MD simulations of nucleotide incorporation by TdT. It was demonstrated that the incorporation efficiency decreases in the order  $dGTP > dTTP \approx dATP > dCTP$  in the presence of  $Mg^{2+}$ . It was shown that the elongation mode, namely the ratio of translocation and dissociation rate constants, as well as the catalytic rate constant are dependent on the nature of the nucleobase. The MD data indicate that during the recognition of various nucleotides, TdT comes into specific contacts with the nucleotide being incorporated. It turned out that guanine forms a broad network of contacts with amino acid residues in the active-site pocket, which allow dGTP to get efficiently accommodated at the correct position and facilitate the catalytic reaction. According to the MD simulations, it is possible that interactions with Asp395, Val394 and Arg453 are responsible for efficient processive dGTP incorporation by TdT. We have applied stopped-flow kinetics using intrinsic tryptophan fluorescence of the enzyme to analyze the conformational dynamics of TdT during processing of DNA synthesis. Similar to other DNA polymerases, TdT seems to undergo a multistep conformational transition during substrate binding and further conformational changes during catalysis. Such behavior illustrates the concept of the dynamic model of enzyme catalysis, in which conformational changes in the enzyme molecule drive the substrate-to-product conversion.

*Acknowledgements:* This work was supported by Russian Science Foundation grant No. 21-64-00017.

## **Delineating molecular mechanisms of c-FLIP isoforms as control checkpoints of DED filament assembly and caspase-8 activation**

Lavrik I.

*Translational Inflammation Research, Medical Faculty, Otto von Guericke University Magdeburg, Magdeburg, Germany*

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

*lavrik@bionet.nsc.ru*

Induction of apoptosis via CD95 and TRAILR1/R2 is triggered by the formation of Death-Inducing Signaling Complex (DISC) and death effector domain (DED) filaments. Caspase-8 activation at the DISC is largely controlled by c-FLIP proteins. Despite apparent progress in understanding the assembly of CD95-activated platforms and DED filaments, the detailed molecular mechanism of c-FLIP action remains elusive. To get more insight into this question we have created several mathematical models of CD95 network. The modeling has pointed out that c-FLIP is a key target in the network. Subsequently, we have rationally designed a chemical probe targeting c-FLIPL in the heterodimer caspase-8/c-FLIPL. This rationally designed small molecule was aimed to imitate the closed conformation of the caspase-8 L2' loop and thereby increase caspase-8 activity after initial processing of the heterodimer. In accordance with *in silico* predictions, the small molecule enhanced caspase-8 activity in the DISC, CD95L/TRAIL-induced caspase activation and subsequent apoptosis. A generated computational model provided further evidence for the proposed effects of the small molecule on the heterodimer caspase-8/c-FLIPL at the DED filaments. In particular, the model demonstrated that boosting caspase-8 activity at early time points after DISC assembly by caspase-8/c-FLIPL heterodimer is crucial to promote apoptosis induction. To further address the mechanism of action of c-FLIP isoforms, we have investigated their role at the DED filaments using quantitative mass spectrometry and structural modeling. These analyses uncovered that the overexpression of both c-FLIP isoforms leads to the formation of shorter DED filaments and an increase in the DISC quantities. Our data strongly indicate that c-FLIP can bind to both FADD and procaspase-8 at the DED filament. Moreover, the constructed *in silico* model shows that c-FLIP proteins can serve as bridging motifs for building a cooperative DISC network, in which adjacent CD95 DISCs are connected by DED filaments. These findings provide new insights into DISC and DED filament regulation and open innovative possibilities for targeting the extrinsic apoptosis pathway.

## Regulation of PARP1 activity by its protein partners

Lavrik O.

*Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia  
Novosibirsk State University, Novosibirsk, Russia  
lavrik@niboch.nsc.ru*

**Key words:** DNA repair, PARP1, PAR, YB-1, FUS, dynamic compartments, liquid-liquid phase separation, liquid demixing

Poly (ADP-ribose) polymerase 1 (PARP1) is catalysis poly(ADP-ribose) (PAR) at DNA damage sites to recruit DNA repair factors. PARP1 is playing key role in regulation of multiple processes in high eukaryotes and interacts with DNA and RNA binding proteins. Among proteins relocated on damaged DNA, the nuclear RNA-binding protein FUS is one of the most abundant, raising the issue about its involvement in DNA repair. The reconstituted PARP-1/PAR/damaged DNA system was analyzed by atomic force microscopy in the presence of FUS. We demonstrated FUS recruitment at DNA damage sites via PAR binding, and observed assembly of dynamic compartments in which damaged DNA is concentrated (Singatulina et al., Cell Rep, 2019; Sukhanova et al., IJMS, 2021). FUS contains disordered prion-like domains and its interaction with PAR is a driving force in formation of membrane less compartments by liquid-liquid phase separation. We anticipate that FUS facilitates DNA repair through the transient compartmentalization of damaged DNA. This process permits to select damaged DNA from undamaged one and attracts DNA repair proteins. The hydrolysis of PAR by PARG dissociates these compartments, which in turn promotes the turnover of DNA repair in cell. Y-box-binding protein 1 (YB-1) is a multifunctional positively charged protein that interacts with DNA or RNA and poly(ADP-ribose) (PAR). YB-1 contains a disordered Ala/Pro-rich N-terminal domain (A/P domain), a cold shock domain (CSD) and an intrinsically disordered C-terminal domain (CTD) containing several clusters of positively and negatively charged amino acid residues. The C-terminal domain plays a critical role in YB-1 binding to nucleic acids, including polymer poly(ADP-ribose). YB-1 was shown functionally interacts with poly(ADP-ribose) polymerase 1 (PARP1) which catalyzes the synthesis of a poly(ADP-ribose) and protein poly(ADP-ribosylation), including auto-poly(ADP-ribosylation), during genotoxic stress. It was shown by us that Yb-1 stimulates PARP1 activity (Naumenko et al., Biomolecules, 2020). We studied the functional role of disordered C-terminal domain of YB-1 in the regulation of poly(ADP-ribose) synthesis by PARP1 *in vitro*. We demonstrate that the PARP1-dependent poly(ADP-ribose) synthesis rate is increased in the presence of YB-1 and is strongly controlled by the interaction of CTD (YB-1) with poly(ADP-ribose). Moreover, we show that both decrease of chain length or increase of branching frequency of poly(ADP-ribose) have an effect on the poly(ADP-ribose) binding affinity of YB-1 and YB-1-mediated stimulation of PARP1 enzymatic activity. These findings provide interesting insights into the regulation of PARP1 activity with poly(ADP-ribose) binding proteins containing charged amino-acid residues in disordered domains (Naumenko, Biomolecules, 2021). YB-1 is expressed in malignant tumors and may influence sensitivity to anticancer drugs. Therefore, our study shed light on the role of PARP1 and RNA binding proteins containing prion-like domains in stimulation of PARP1 activity and in the formation of dynamic liquid compartments, which can facilitate DNA repair by concentrating damaged DNA and repair proteins.

*Acknowledgements:* This work was supported by RSF, No. 22-14-00112.

# Mechanism of primase activity human primase polymerase PrimPol

Makarova A.V.<sup>1\*</sup>, Boldinova E.O.<sup>1</sup>, Baranovskiy A.G.<sup>2</sup>, Gagarinskaya D.I.<sup>1,2</sup>, Tahirov T.H.<sup>2</sup>

<sup>1</sup> Institute of Molecular Genetics, National Research Center “Kurchatov Institute”, Moscow, Russia

<sup>2</sup> Eppley Institute for Research in Cancer and Allied Diseases, Fred & Pamela Buffett Cancer Center, University of Nebraska Medical Center, Omaha, Nebraska, United States

\* amakarova-img@yandex.ru

**Key words:** PrimPol, primase, *de novo* DNA synthesis

**Motivation and Aim:** Human primase and DNA polymerase PrimPol belongs to the archaeo-eukaryotic superfamily and synthesizes primers *de novo* using deoxyribonucleotides. PrimPol plays a role in replication both the nucleus and mitochondria by re-initiation of DNA synthesis at sites containing DNA damage or secondary structures that block high-fidelity DNA polymerases.

PrimPol is composed of two domains: an N-terminal AEP-like catalytic domain (NTD) harboring catalytic amino acid residues Asp114/Glu116 and Asp280 involved in the coordination of Me<sup>2+</sup> ions and a C-terminal domain (CTD) containing a conserved C-H-C-C zinc finger (ZnF) motif that coordinates the Zn<sup>2+</sup> ion (Cys419, His426, Cys446, Cys451). Deletion of ZnF and mutations of Cys419 and His426 disrupt PrimPol primase activity while retaining DNA polymerase activity. PrimPol prefers to initiate DNA synthesis on the 3'-GTC-5' sequence with cryptic G and its activity is stimulated by Mn<sup>2+</sup> ions. It is assumed that binding of the first two nucleotides occurs at the 3'-elongation site and the 5'-initiation site (preferably ATP), followed by catalysis and formation of the rA-dG dinucleotide.

The detailed mechanism of PrimPol primase activity is not fully understood. Structures of the full-length PrimPol and the initiation complex of PrimPol with a DNA template and a dinucleotide have not been deciphered. However important information on the mechanism of PrimPol primase activity has been obtained in biochemical studies.

**Methods and Algorithms:** We investigated the mechanism of PrimPol primase activity using its variants that selectively disrupt any catalytic activity (D114A) or only primase activity (ZnF mutations R417A and R424A), as well as individual NTD and CTD domains of PrimPol. To analyze protein interaction with DNA, a DNA oligonucleotide containing the 5'-terminal adenosine triphosphate was synthesized enzymatically *de novo* using the wild-type PrimPol.

**Results:** Due to bi-modal organization, primases can operate in the *cis*- or *trans*-orientation. We show that PrimPol performs primer synthesis in the *cis*-orientation, when the NTD catalytic and CTD regulatory domains of one molecule take part in catalysis. According to the suggested model, the catalytic NTD is responsible for catalysis and can initiate *de novo* DNA synthesis alone but the CTD plays a key regulatory role and is involved in template-primer binding by PrimPol. We showed that the interaction of ZnF of the CTD involves the terminal 5'-triphosphate of ATP. During *de novo* DNA synthesis, PrimPol makes pauses.

**Acknowledgements:** This work was supported by the RSF grant No. 18-14-00354.

# Nucleosome reorganization by PARP-1

Maluchenko N.<sup>1\*</sup>, Nilov D.<sup>2</sup>, Koshkina D.<sup>1,3</sup>, Lys A.<sup>1,3</sup>, Geraskina O.<sup>1,3</sup>, Korovina A.<sup>1</sup>, Feofanov A.<sup>1,4</sup>, Studitsky V.<sup>1,5</sup>

<sup>1</sup> Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> Institute of Gene of Biology, RAS, Moscow, Russia

<sup>4</sup> Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>5</sup> Fox Chase Cancer Center, Philadelphia, USA

\* mal\_nat@mail.ru or malyuchenkov@my.msu.ru

**Key words:** nucleosome; poly(ADP-ribose) polymerase 1; spFRET microscopy

*Introduction:* Human poly (ADP-ribose) polymerase (PARP) family consists of 17 proteins that are involved in many cellular processes, including transcription, modulation of chromatin structure, genome maintenance, replication, recombination, and DNA damage repair [1]. PARP-1 is the best structurally and functionally characterized protein of this family, since it is a key player in almost all nuclear and some cytoplasmic events in the cell. The diverse spectrum of PARP-1 activity in the cell is due to its enzymatic and DNA-binding activities. The multidomain structure of PARP-1 includes a DNA-binding region with three zinc fingers localized at the N-terminal end and an enzymatic active site localized at the C-terminal region of the protein. PARP-1 has distinct catalytic activity-dependent and independent functions during DNA transactions [2]. First, upon binding to a DNA lesion, PARP-1 becomes catalytically active, thus acting as a DNA damage sensor enzyme, which hydrolyzes NAD<sup>+</sup> and forms long, branched chains of negatively charged polyADP-ribose (PAR) on a variety of nuclear proteins, including histones and PARP-1 itself [3]. PARP-1 is responsible for the production of approximately 90 % of PAR in the cell [4]. The catalytic activity-independent function is realized during interaction of PARP-1 with intractable chromatin regions where it makes protein-binding sites on DNA available for interactions with pioneer factors [2]. Impaired PARP-1 metabolism is associated with the development of various cancers, as well as cardiovascular and neurodegenerative diseases [5]. Accordingly, PARP-1 inhibitors are used as antitumor drugs, and the applicability to the treatment of other metabolic diseases is being investigated [6]. Despite the intensive study of PARP-1, many aspects of its interaction with chromatin at the nucleosomal level are not fully understood; in particular, it is not clear how PARP-1 contributes to the accessibility of DNA regions hidden in nucleosomes to pioneer factors. How is the catalytically independent mode of action of PARP-1 realized on nucleosomes? To answer these questions, we developed sensitive sensors using fluorescent nucleosomes to probe the conformational changes in nucleosomal DNA induced by PARP-1.

*Results:* Uniquely positioned nucleosomes were reconstituted on the 603 DNA templates *in vitro*. To make the nucleosomes suitable for analysis by single-particle Förster resonance energy transfer (spFRET) microscopy, a single pair of fluorescent dyes Cy3 (donor) and Cy5 (acceptor) was introduced in different, close positioned regions of the nucleosomal DNA [7]. The efficient FRET was observed between Cy3 and Cy5 without disturbance of histone-DNA contacts in the assembled nucleosomes. Labeled DNA

regions localized 13 and 135 bp from the nucleosome boundary are involved in so called DNA “breathing” in the nucleosome. Labels at 35 and 112 bp are located near the DNA contacts with H2A/H2B dimers, whereas labels at 57 and 91 bp are near the H3/H4 tetramer. To study PARP-1 induced changes in the linker DNA region, both linker DNA helixes were labeled at positions of 18 bp from the boundaries of the core nucleosome. Electrophoretic mobility shift assay revealed that PARP-1 can form three distinct complexes with linker nucleosomes. It was found that these complexes differed in a number of PARP-1 molecules bound to a nucleosome. First PARP1-nucleosome complex is observed at a low PARP1 concentration (10 nM). Second and third complexes having a lower mobility in the gel are formed at 20 nM and 50 nM PARP1 respectively. A similar concentration-dependent effect was observed during nucleosome reorganization by PARP-1. After binding of the first PARP-1 molecule, no pronounced structural changes were observed in nucleosomes, probably because of PARP-1 binding at the DNA ends localized at a distance from the nucleosome. With an increase in the concentration of PARP-1 to 20 and 50 nM, the distance between nucleosomal DNA gyres increased along the entire DNA length, while linker DNA helixes converged. Last observation indicates that one of the PARP-1 molecules can interact with the nucleosome in the H1-like manner by binding near the nucleosome dyad. Moreover, we have shown that PARP-1 was able to displace H1.0 from nucleosomes in the absence of polyADP-ribosylation. To understand the effect of PARP-1 on nucleosome structure, 25 ns molecular dynamics of nucleosome complexes with PARP-1 was performed. Analysis showed that two PARP-1 molecules bound to DNA ends can form intermolecular bonds between the catalytic domains. This interaction, in turn, induced a distortion of the linker DNA accompanied by an increase in the distance between the gyres of nucleosomal DNA.

*Conclusion:* Our data suggest that PARP-1 can induce formation of an alternative nucleosome state that is likely involved in gene regulation and DNA repair.

*Acknowledgements:* Studies were supported by Russian Science Foundation No. 21-64-00001.

### *References*

1. Richard I.A. et al. Beyond PARP1: The potential of other members of the poly (ADP-ribose polymerase family in DNA repair and cancer therapeutics. *Front Cell Dev Biol.* 2022;9:801200.
2. Liu Z., Kraus W.L. Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol Cell.* 2017;65(4):589-603.e9.
3. Hottiger M. Nuclear ADP-ribosylation and its role in chromatin plasticity, cell differentiation, and epigenetics. *Ann Rev Biochemistry.* 2015;84:227-63.
4. D'Amours D. et al. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem J.* 1999;342:249-68.
5. Maluchenko N. et al. PARP-1-associated pathological processes: inhibition by natural polyphenols. *Int J Mol Sci.* 2021;22(21): 11441.
6. Mateo J. et al. A decade of clinical development of PARP inhibitors in perspective. *Ann Oncol* 2019;30:1437-1447.
7. Maluchenko N. et al. Mechanisms of Nucleosome Reorganization by PARP1. *Int J Mol Sci.* 2021;9(22):12127.

## **Focal adhesion kinase plays a dual role in TRAIL resistance and metastatic outgrowth of malignant melanoma**

Mistro G.<sup>1</sup>, Riemann S.<sup>1</sup>, Schindler S.<sup>1,2</sup>, Beissert S.<sup>1</sup>, Kontermann R.<sup>3</sup>, Ginolhac A.<sup>4</sup>, Halder R.<sup>5</sup>, Presta L.<sup>4</sup>, Sinkkonen L.<sup>4</sup>, Sauter T.<sup>4</sup>, Kulms D.<sup>1,2</sup>

<sup>1</sup> *Experimental Dermatology, Department of Dermatology, Dresden, Germany*

<sup>2</sup> *National Center for Tumor Diseases Dresden, Dresden, Germany*

<sup>3</sup> *Institute of Cell Biology and Immunology and Stuttgart Research Centre Systems Biology, University of Stuttgart, Stuttgart, Germany*

<sup>4</sup> *Department of Life Sciences and Medicine, University of Luxembourg, Luxembourg*

<sup>5</sup> *Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg*

Despite remarkable advances in therapeutic interventions, malignant melanoma (MM) remains a life-threatening disease. Following high initial response rates to targeted kinase inhibition metastases quickly acquire resistance and present with enhanced tumor progression and invasion, demanding for alternative treatment options. We show 2nd generation hexameric TRAIL receptor agonist IZI1551 (IZI) to effectively induce apoptosis in MM cells irrespective of the intrinsic BRAF/NRAS mutation status. Conditioning to the EC50 dose of IZI converted the phenotype of IZI-sensitive parental MM cells into a fast proliferating and invasive, IZI-resistant metastasis. Mechanistically, we identified focal adhesion kinase (FAK) to play a dual role in phenotype-switching. In the cytosol, activated FAK triggers survival pathways in a PI3K- and MAPK-dependent manner. In the nucleus, the FERM domain of FAK prevents activation of wtp53, as being expressed in the majority of MM, and consequently intrinsic apoptosis. Caspase-8-mediated cleavage of FAK as well as FAK knock down, and pharmacological inhibition respectively, reverted the metastatic phenotype-switch and restored IZI responsiveness. FAK inhibition also re-sensitized MM cells isolated from patient metastasis that had relapsed from targeted kinase inhibition to cell death, irrespective of the intrinsic BRAF/NRAS mutation status. Hence, FAK-inhibition alone or in combination with 2nd generation TRAIL-receptor agonists may be recommended for treatment of initially resistant and relapsed MM respectively.

## Regulation of poly(ADP-ribose) polymerase 1 activity by Y-box-binding protein 1

Naumenko K.\*, Sukhanova M., Kurgina T., Anarbaev R., Alemasova E., Kutuzov M., Lavrik O.

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

\* *k-naumenko@mail.ru*

**Key words:** Y-box-binding protein 1, PARP1, trans-poly(ADP-ribosyl)ation, poly(ADP-ribose), disordered C-terminal domain

*Motivation and Aim:* The poly (ADP-ribose) polymerase 1 (PARP1) is considered as one of the key regulators of mammalian DNA damage response. PARP1 catalyzes synthesis of poly(ADP-ribose) (PAR) by transferring of ADP-ribose units from NAD<sup>+</sup> onto target proteins resulting in their mono- or poly(ADP-ribosyl)ation (MARylation or PARylation). It is currently believed that RNA-binding proteins can be involved in the DNA repair and regulation of PARP1's functions in cell responses to genotoxic stress. One of these proteins is the multifunctional Y-box binding protein (YB-1). YB-1 is considered as PARP1-interacting partner and is a target for the covalent PARylation both *in vitro* and *in vivo*.

*Methods and Algorithms:* We studied the YB-1-dependent regulation of PARP1 activity *in vitro* using biochemical approaches and single molecular method, such as atomic force microscopy.

*Results and Conclusion:* In the present study, we analyzed the regulation of PARP1 activity with YB-1 via its ability to bind PAR (damaged DNA) and be trans-PARylated. YB-1 consists of an alanine/proline-rich (AP) domain, a cold-shock domain (CSD), and an unstructured C-terminal domain (CTD) carrying clusters of negatively and positively charged residues. An unstructured C-terminal part of YB-1 involved in an interaction with PAR behaves similarly to full-length YB-1 in modulation of PAR synthesis, indicating that both DNA and PAR binding are involved in the regulation of PARP1 activity by YB-1. To determine molecular mechanism by which YB-1 regulates PARP1 activity, we generated YB-1 mutants which lacks clusters of positively charged amino acids in the C-terminal domain and PARP1 mutants synthesizing highly branched or short PAR polymers. We found a correlation between the positive charge of the YB-1 CTD and its ability to interact with PAR and damaged DNA and to be trans-PARylated. Furthermore, PAR structural features such as frequency of branching and chain length were observed to influence noncovalent binding of YB-1 and to have a strong influence on YB-1 trans-PARylation.

Thus, the CTD can be considered a specific PAR-interacting module that plays a critical role for binding of YB-1 to PAR.

*Acknowledgements:* The study is supported by RSF No. 20-14-00086.



## Inter-subunit crosstalk synergistically regulates allosteric activation of proapoptotic serine protease HtrA2

Parui A.<sup>1,2</sup>, Mishra V.<sup>3</sup>, Dutta S.<sup>1</sup>, Bhaumik P.<sup>3</sup>, Bose K.<sup>1,2\*</sup>

<sup>1</sup> Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Navi Mumbai, India

<sup>2</sup> Homi Bhabha National Institute, Anushaktinagar, Mumbai, India

<sup>3</sup> Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India

\* [kbose@actrec.gov.in](mailto:kbose@actrec.gov.in)

**Key words:** HtrA2, protease, crosstalk, PDZ domain, apoptosis

High-temperature requirement protease A2 (HtrA2) is a complex trimeric mitochondrial serine protease that primarily acts as a key player in apoptosis. It belongs to a large family of multi-domain serine proteases (S1, chymotrypsin family) that is found to be conserved from prokaryotes to humans [1]. There exist four human homologues (HtrA1-HtrA4), which reveal a common evolutionarily conserved structural architecture. They share an overall common polypeptide fold with an active site containing a catalytic triad, oxyanion hole and substrate specificity pocket. Despite sharing such structural integrity, low sequence identity exists between these serine protease homologues and the subtle conformational variations induced upon ligand binding provide the basis for their specificity and functional diversity. The presence of an N-terminal tetrapeptide AVPS motif in mature trimeric HtrA2 makes it the only unique IAP-binding protein amongst all the HtrA homologues that complements the proapoptotic activity of SMAC protein. Apart from mediating caspase-dependent apoptosis by deactivating XIAP (X-linked Inhibitor of Apoptosis Protein), the structural organization and domain plasticity have also enabled HtrA2 to promote caspase-independent apoptotic pathways via its intricate serine protease activity, the complex mechanisms of which are yet to be delineated [2, 3]. Deregulation in the functioning of a multitasking protein such as HtrA2 lays the foundation for a number of pathophysiological conditions, such as cancer, arthritis and neurodegenerative disorders [4]. However, the complex mechanism through which it promotes apoptotic cascades or regulates protein quality control still remains elusive.

HtrA2 comprises three important regions: a short N-terminal trimerization motif, a serine protease domain (SPD) and a C-terminal protein-protein interaction or PDZ domain. The available crystallographic data of HtrA2 provides a broad overview of the overall structural organization. However, the structural data is acquired from the mature unbound form of the inactive protease and hence fails to elucidate the intricate structural dynamics of PDZ-protease crosstalk that take place upon ligand binding. Previous studies from our lab have shown that activation of HtrA2 occurs allosterically, where the signal is relayed via a distal non-canonical substrate-binding site with subsequent opening up of the peptide-binding groove in PDZ.<sup>5</sup> Thus, the transition from basal to proteolytically active state seems to be predominantly governed by PDZs, owing to its specific interactions with numerous C-terminal binding partners. To expound on the role of regulatory PDZ domains in promoting synergistic coordination between HtrA2 subunits, we generated heterotrimeric HtrA2 variants comprising different numbers of PDZs and/or active-site mutations. Sequential deletion of PDZs from the trimeric

ensemble significantly affected its residual activity in a way that proffered a hypothesis advocating *intermolecular* allosteric crosstalk via PDZ domains in trimeric HtrA2. Furthermore, structural and computational snapshots affirmed the role of PDZs in secondary structural element formation and coordinated reorganization of the N-terminal region and regulatory loops. Therefore, apart from providing cues for devising structure-guided therapeutic strategies, this study establishes a working model of complex allosteric regulation through a multifaceted *trans*-mediated cooperatively-shared energy landscape.

*Acknowledgements:* The authors thank the Department of Biotechnology (DBT), Govt. of India (grant number BT/HRD/NWBA/37/01/2015) for funding this project.

### *References*

1. Singh N., Kuppli R.R., Bose K. *Arch Biochem Biophys.* 2011;516:85-96.
2. Singh N. et al. *FEBS J.* 2014;281:2456-2470.
3. Chaganti L.K., Kuppli R.R., Bose K. *FASEB J.* 2013;27:3054-3066.
4. Merski M. et al. *Cell Death & Disease.* 2017;8:e3119-e3119.
5. Bejugam P.R. et al. *PLoS One.* 2013;8:e55416.

# G-quadruplex formed by the promoter region of the *hTERT* gene: structure-driven effects on DNA mismatch repair functions

Pavlova A.<sup>1</sup>, Savitskaya V.<sup>1</sup>, Dolinnaya N.<sup>1</sup>, Monakhova M.<sup>2</sup>, Litvinova A.<sup>3</sup>, Zvereva M.<sup>1</sup>, Kubareva E.<sup>2\*</sup>

<sup>1</sup> Chemistry Department, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup> Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

\* [kubarevaea@gmail.com](mailto:kubarevaea@gmail.com)

**Key words:** DNA mismatch repair, MutS, MutL, G-quadruplexes, *hTERT* promoter

**Motivation and Aim:** G-quadruplexes (G4s) are a unique class of stable four-stranded noncanonical DNAs that play a key role in cellular processes. In particular, G4s have been extensively studied as promising anti-cancer targets for clinical application. Here we have focused on the G4-forming 68-nt promoter region of the human telomerase reverse transcriptase (*hTERT*) oncogene. *hTERT* is the catalytic subunit of telomerase, an enzyme primarily responsible for the immortality of cancer cells. Genetic studies have identified point mutations in the *hTERT* promoter sequence that are directly associated with increased *hTERT* expression. Also known is the ability of G4s to modulate the exposure of some DNA promoter sites to various repair systems, thus influencing their efficiency. It was shown that G4s have a diverse effect on the functioning of the base and nucleotide excision repair systems. We focused on studying the impact of *hTERT* G4s, including mutated forms of quadruplex structures, on the functioning of various DNA mismatch repair (MMR) pathways.

**Methods and Algorithms:** Using chemical probing and spectroscopic methods, the effect of four clinically significant nucleotide substitutions in the G4 structure formed by the single-stranded (ss) 68-nt *hTERT* promoter region on the G4 folding was studied. Mutually exclusive mutations occur most frequently at positions –124 bp (position 1295228 of chromosome 5, the mutation is designated as G228A) and –146 bp (position 1295250 – G250A) relative to the ATG start codon of the *hTERT* gene. A less common is double substitution at positions –138/–139 bp from the start codon (G242A/G243A). We elucidated how ss *hTERT* multiG4 structure and its mutated forms affect the binding affinities and functional responses of key proteins, MutS and MutL, involved in the initial stage of DNA mismatch repair (MMR) from two bacterial species, *E. coli* and *Neisserria gonorrhoeae*. The latter have different repair mechanism, but some common features with the human MMR system.

**Results:** Characteristic features of the temperature dependence of UV absorption at 295 nm indicate G4 unfolding with a melting temperature of 58°C for the ss 96-nt WT *hTERT* promoter region (ss96-WT). The CD spectra displayed characteristics of a parallel G4 topology of ss96-WT. The DMS probing results suggest that formation of stacked three-G4 structure actually occurs in WT and mutant 96-nt DNA fragments of the *hTERT* promoter. No significant difference was observed in the binding of both MutS and MutL from *E. coli* to WT and mutant *hTERT* G4s, suggesting these MMR proteins recognize

the G4 conformation through a structure-based mechanism. We have shown that *N. gonorrhoeae* MutL (ngMutL) binding to G4 has a 3-fold higher affinity than to double-stranded DNA of the same length. The introduction of mutations in the G4 *hTERT* sequence had no significant effect on complex formation with ngMutL as in the case of *E. coli* MutL. For the first time, the hydrolysis of linear ss *hTERT* substrates was carried out by ngMutL. Protein cleaves ss96-WT between G4s, but not within quadruplex motifs. Mutations in *hTERT* region did not significantly affect the endonuclease function of ngMutL.

*Conclusion:* It has been shown for the first time that MutS and MutL proteins from MMR system effectively interact with three tandem parallel quadruplexes of the G-rich strand of *hTERT* promoter despite single nucleotide substitutions in it. Formation of G4 in the *hTERT* promoter prevents the introduction of a nick by the ngMutL, thereby suppressing MMR.

*Acknowledgements:* The study is supported by Russian Science Foundation (grant 21-14-00161).

# Assessment of the nucleotide excision repair system activity *ex vivo*

Popov A.A. \*, Petrusseva I., Lavrik O.

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

\* [aa-popov1994@yandex.ru](mailto:aa-popov1994@yandex.ru)

**Key words:** nucleotide excision repair, *ex vivo* methods, DNA damages

**Motivation and Aim:** Nucleotide excision repair (NER) is one of the main repair systems in cells of living organisms, which is responsible for removal of a wide range of bulky DNA lesions. The use of approaches focused on exploring the structure and functions of proteins involved in NER would enable elucidation of the process mechanism and identification of the main stages affecting its efficiency as well as the composition and structure of multiprotein complexes that arise and function at different NER stages [1, 2]. Nevertheless, the development of approaches to investigating and comparing the efficiency of bulky lesion repair in living cells (*ex vivo*) remains topical for both fundamental and applied research.

**Methods and Algorithms:** Recombinant plasmids containing bulky nFlu and nAnt lesions (hereinafter referred to as nFlu and nAnt DNA respectively) were synthesized. Transfection of cells with the plasmid was performed using Lipofectamine™ 2000 (Invitrogen) according to the manufacturer's protocol. To analyze the efficiency of NER of nFlu- and nAnt-DNA in HEK 293T human embryonic kidney cells we assessed the time of emergence of cells whose fluorescence indicated recovery of TagRFP protein expression using the Cell-IQ MLF system (Chip-Man Technologies, Finland) for long-term intravital monitoring of cells at the Common Use Center of the Institute of Cytology and Genetics, SB RAS. Cells were pictured at 10 min intervals in phase contrast and fluorescence modes using a Nikon CFI Plan Fluorescence DL ×10 objective. The resulting images were analyzed using the ImageJ and Cell-IQ Analyzer software.

**Results:** Our constructed plasmids containing bulky nFlu or nAnt lesions near the *tagrfp* gene promoter were shown to undergo repair in eukaryotic cells (HEK 293T) and that they can be used to analyze the efficiency of NER *ex vivo*. A comparative analysis of the time dependence of fluorescent cells accumulation after transfection with nFlu- and nAnt-DNA revealed that there are differences in how efficient their repair by the NER system of HEK 293T cells can be.

**Conclusion:** The proposed method enables assessing the efficiency of removal of bulky nAnt and nFlu lesions from model plasmids by the NER system of HEK 293T cells. The method is a promising tool for studying NER; it enables comparing both the repair status of various cells and the efficiency of repair of various structural lesions.

**Acknowledgements:** This study was supported by the Russian Science Foundation (project No. 19-74-10056).

## References

1. Schärer O.D. Nucleotide Excision Repair in Eukaryotes. *Cold Spring Harb Perspect Biol.* 2013;5(10):1-20.
2. Luijsterburg M.S. et al. Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. *J Cell Biol.* 2010;189(3):445-463.

## The novel radioprotective agent. Characterization of the substance and the mechanism of its action

Ritter G. \*, Bogachev S.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [dilkwe@yandex.ru](mailto:dilkwe@yandex.ru)

**Key words:** RNA internalization, radioprotection, double-stranded RNA, homologous recombination

*Motivation and Aim:* Nucleic acids are known to exert radioprotective properties, which are believed to be associated with the restoration of the integrity of nuclear DNA [1]. Nevertheless, no studies on determining the cellular and molecular mechanisms of the radioprotective capabilities of nucleic acids have been undertaken yet. In the current report, a stage of the research aimed at clarifying the mechanisms mediating the radioprotective effect of RNA is described.

*Methods and Algorithms:* Mice were intravenously injected with yeast total RNA and than exposed to an absolutely lethal dose of  $\gamma$ -radiation. The effective parameters of the radioprotective effect of the compound were established. Pathomorphological examination of spleens, as well as analysis of the rate of post-irradiation recovery of hematopoiesis, proved the RNA preparation to facilitate the reconstitution of the hematopoietic and immune systems. Using hydroxyapatite chromatography, it was found that it is double-stranded RNA (dsRNA) that mediates the radioprotective effect of the yeast RNA preparation. Based on two sequenced molecules as a model structure, we have designed and synthesized the fluorescent dsRNA probe. Using the double fluorescent-radioisotope labeling, we have demonstrated the native internalization of dsRNA into hematopoietic stem cells. Molecules of dsRNA were show to be resistant to blood nucleases and, being administered intravenously, accumulate in some “critical” organs. Using bone marrow cells pre-incubated with the radioactive dsRNA probe, we have identified the organs predominantly populated by circulating (i.e. mobilized) hematopoietic stem cells.

*Results:* Intravenous injection of 7 mg of total RNA 30 minutes prior to exposure to an absolutely lethal dose (9.4 Gy) of  $\gamma$ -radiation increases the survival rate of irradiated mice up to 80–100 % [2]. The acting component, dsRNA, was isolated from *S. cerevisiae* total RNA. DsRNA was established to exert the radioprotective effect by facilitating the post-irradiation reconstitution of the hematopoietic and immune systems [3]. It was found that radioprotective properties do not depend on the primary structure of dsRNA, but unequivocally depend on the presence of free double-stranded ends. It was demonstrated that dsRNA can be natively internalized into hematopoietic stem cells [4]. P<sup>32</sup>-labeled dsRNA was shown to remain intact after at least one hour of circulation in blood, after which it can be detected in the bone marrow of the recipient animal. Intravenous injection of bone marrow cells exposed to P<sup>32</sup>-labeled dsRNA results in detection of P<sup>32</sup> predominantly in the bone marrow and spleen.

*Conclusion:* Based on the results obtained, the mechanism for the radioprotective capability of dsRNA, which implies the internalization of intravenously administered dsRNA by hematopoietic stem cells, where it further participates in the post-irradiation repairing of double-stranded DNA breaks, has been hypothesized. Once the repair

process is over, the preserved hematopoietic stem cells exit to the circulatory system and then repopulate the emptied niches in hematopoietic organs, where they proliferate and reconstitute the hematopoietic and immune systems. Due to the greater number of the preserved hematopoietic stem cells, such a reconstitution goes much faster and thus increases the survival rate of irradiated mice.

*Acknowledgements:* This research was funded by the Russian Ministry of Science and High Education via the Institute of Cytology and Genetics (State Budget Project No. FWNR-2022-0016), LLC “KARANAHAN” and LLC “ACTIVATOR MAF”.

### *References*

1. Likhacheva A.S. et al. Exogenous DNA can be captured by stem cells and be involved in their rescue from death after lethal-dose  $\gamma$ -radiation. *Gene Ther Mol Biol.* 2007;11(2):305-314.
2. Ritter G.S. et al. Characterization of biological peculiarities of the radioprotective activity of double-stranded RNA isolated from *Saccharomyces cerevisiae*. *Int J Radiat Biol.* 2020;96(9):1173-1191.
3. Dubatolova T.D. et al. Changes in the number and morphology of blood cells in mice pretreated with RNA preparations and exposed to the sublethal dose of gamma radiation of 8 Gy. *Int J Radiat Res.* 2021. (In press).
4. Ritter G.S. et al. Characteristic of the active substance of the *Saccharomyces cerevisiae* preparation having radioprotective properties. *Vavilov J Genet Breed.* 2020;24(6):643-652.

## Targeting urokinase receptor uPAR in cancer: high expectations, modest achievements

Semina E.<sup>1,2\*</sup>, Rysenkova K.<sup>1,2</sup>, Klimovich P.<sup>1,2</sup>, Shmakova A.<sup>1,2</sup>, Karagyaur M.<sup>2</sup>, Rubina K.<sup>2</sup>

<sup>1</sup>*Institute of Experimental Cardiology, National Cardiology Research Center of the Ministry of Health of the Russian Federation, Moscow, Russia*

<sup>2</sup>*Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia*

\* [e-semina@yandex.ru](mailto:e-semina@yandex.ru)

**Key words:** urokinase receptor, uPAR, urokinase, uPA, neuroblastoma, chemoresistance, apoptosis, CRISPR/Cas9

Elevated levels of urokinase (uPA) and its receptor (uPAR) are detected in various malignancies and urokinase system is considered to be a potential therapeutic target in cancer. Binding of uPA to uPAR triggers the activation of proteinases, thus mediating matrix degradation and facilitating cancer cell invasion/metastasis and neoangiogenesis. In our previous studies using CRISPR/Cas9 technology in Neuro2a neuroblastoma cells, we have demonstrated that uPAR-knockout results in the reduced proliferation, suppressed uPA-dependent invasion and in the activation of caspase-dependent apoptosis. Strikingly, uPAR-deficient cells exhibit increased migration which is independent of the urokinase uPA. In search for the mechanisms of this phenomenon, we have found that urokinase in these cells can penetrate into the nucleus and can affect the expression of oncogenes (such as NF- $\kappa$ B, p53) triggering the onset of the epithelial-mesenchymal transition and selection/proliferation of chemoresistant cancer cells. Although underlying mechanisms are yet to be fully elucidated, we surmise that a decrease in uPAR expression in certain cancer cells can lead to the selection of low-proliferating cells that have a higher migratory potential and are insensitive to antitumor therapy, possibly resulting in the formation of dormant metastases.

The increased attention of scientists to the urokinase system and its role in carcinogenesis seems to be quite justified, since a significant amount of data has been accumulated to date, convincingly proving the direct involvement of the urokinase system in tumor progression. Urokinase uPA and urokinase receptor uPAR in tumor tissue promote extracellular matrix remodeling, invasion, migration and metastasis of tumor cells, tumor angiogenesis, they support the proliferation and survival of tumor cells. High expression of uPA and uPAR is often associated with tumor progression, with a poor prognosis of the course of the disease, and with reduced survival, and targeted suppression of the components of the urokinase system has been effective in a number of experimental models *in vitro* and *in vivo*. Despite this, therapeutic approaches aimed at the uPA/uPAR system did not live up to expectations and did not give convincing results in clinical trials, which left the question of the functional significance of the uPA/uPAR system in oncogenesis unresolved. In addition to insufficient knowledge, the relevance of research in this area is determined by the high prevalence of neoplasms in the population, which also applies to neuroblastoma, the tumor studied in this study. Our data postulate that suppression of uPAR reduces ERK-dependent proliferation of Neuro2a neuroblastoma cells *in vitro* and tumor growth *in vivo*. At the same time, we



found that uPAR expression is important for maintaining the epithelial phenotype in Neuro2a cells and that uPAR suppression promotes the epithelial-mesenchymal transition, increased cell migration *in vitro*, and metastasis *in vivo*. It turned out that the suppression of uPAR leads to an increase in the expression of mRNA of *Snail*, which is a transcription factor that directly represses E-cadherin, and to an increase in the expression of IL-6, a cytokine that stimulates cell migration and maintains the mesenchymal phenotype of tumors, which triggers the epithelial-mesenchymal transition in neuroblastoma cells. In addition, the suppression of uPAR in Neuro2a neuroblastoma cells led to the activation of p38 kinase, which inhibits cell proliferation, and an increase in the expression of p21, an inhibitor of cyclin-dependent kinase 1A, which indicated the dormant phenotype of Neuro2a neuroblastoma cells.

Our study is the first to show that uPAR-deficient cells are less sensitive to chemotherapy and demonstrate lower activation of the p53 protein, the main protein that regulates the cell's response to DNA damage and cell death/survival. Finally, by analyzing transcriptome data from human neuroblastomas, we showed for the first time that initially high *PLAUR* expression predicts poor survival in human neuroblastoma, but relapsed neuroblastomas have significantly reduced *PLAUR* expression. From the analysis of scientific sources, it can be concluded that this is the first study that demonstrated the pathological role of uPAR-deficient tumor cells with a chemoresistant and mobile phenotype.

Mining Transcriptome Data for Neuroblastoma. The microarray-based gene expression data for the Versteeg cohort (NCBI GEO accession GSE16476) were downloaded and analyzed in the R2 database (<http://r2.amc.nl>, accessed on 12 July 2021). The microarray-based gene expression data for primary and relapsed neuroblastoma, published by Schramm et al. (NCBI GEO accession GSE65303), were downloaded and compared using the R2 database (<http://r2.amc.nl>, accessed on 12 July 2021).

*Acknowledgements:* This research was supported by the Russian Foundation of Basic Research (Project No. 20-015-00186).

## The role of active-site amino acids in template-independent DNA synthesis by human terminal deoxynucleotidyl transferase

Senchurova S.I.<sup>1\*</sup>, Kuznetsova A.A.<sup>1</sup>, Laprina D.S.<sup>1</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* *svetasenchurova@gmail.com*

Terminal deoxynucleotidyl transferase (Tdt) is a template-independent DNA polymerase belonging to the DNA polymerase X family. Tdt is only present in vertebrate cells and responsible for adding nucleotides to the 3'-end of single strand DNA at V-D and J-D junction of immunoglobulins and TCR receptor genes during V(D)J recombination. Being isolated and described in 1960 it remains one of the least studied enzyme catalyzing DNA synthesis. Although kinetic mechanisms have been identified for many template-dependent polymerases, it is clear that the absence of a template chain should have a significant impact on the kinetic parameters of DNA and nucleotide binding, catalysis and product dissociation. Herein, we aimed to gain a better understanding of the role of active-site amino acids R336, H342, D345 and L397. According to X-ray structural data, H342 and L397 form contacts with primer, R336 and H342 form contacts with deoxynucleoside triphosphates and D345 forms contacts with catalytic divalent metal ion. The studies were carried out under single-turnover kinetic conditions. The kinetic and thermodynamic parameters of the untemplated DNA synthesis by R336Q, H342A, D345E, L397A mutant forms were obtained. It was shown that substitution of these residues lead to decrease in Tdt activity. Highest efficiency of nucleotide incorporation was observed in the presence of  $Mn^{2+}$  compared to  $Co^{2+}$  and  $Mg^{2+}$  ( $Mn^{2+} > Co^{2+} \gg Mg^{2+}$ ). In the case of mutant D345E, a significant decrease in the binding constant with nucleoside triphosphate compared to WT Tdt was shown.

*Acknowledgements:* This work was supported by Russian Science Foundation grant No. 21-64-00017.

# PARP-1 activation regulates recruitment of FUS to SGs and their assembly without additional recruitment of PAR to SGs

Singatulina A.<sup>1\*</sup>, Desforges B.<sup>2</sup>, Sukhanova M.<sup>1</sup>, Pastré D.<sup>2</sup>, Lavrik O.<sup>1</sup>

<sup>1</sup>LBCE, Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

<sup>2</sup>SABNP, INSERM U1204, Université Paris-Saclay, Evry, France

\* [lasy@ngs.ru](mailto:lasy@ngs.ru)

**Key words:** poly(ADP-ribose) polymerases, DNA damage, stress granules, Fused in Sarcoma, neurodegeneration, mRNA-binding proteins

*Motivation and Aim:* Synthesis of poly(ADP-ribose) (PAR) catalysed by PARP1 and PARP2 is one of the earliest events in the cellular response to DNA damage in higher eukaryotes. Upon binding to DNA breaks, PARPs catalyze the synthesis of PAR covalently attached to PARPs itself or an acceptor protein using NAD<sup>+</sup> as substrate for the post-translation modification [1]. PAR covalently attached to acceptor proteins can be subsequently hydrolysed by PAR glycohydrolase (PARG) thereby releasing free mono(ADP-ribose) or oligo(ADP-ribose) [2]. Recently, several reports have indicated that nuclear poly(ADP-ribose) polymerases (PARP1 and PARP2) play a critical role in the regulation of stress granules (SGs) assembly/disassembly [3–5]. SGs are ribonucleoprotein complexes containing untranslated mRNA and RNA-binding proteins (RBPs) that appear in the cytoplasm of mammalian cells upon the dissociation of polysomes, which occurs generally after various types of cellular stress [6]. There is real interest in deciphering the contribution of nuclear PARPs to SGs formation in cells under oxidative stress conditions, because SGs contain unstructured RBPs, such as TDP-43 and FUS that are associated with neurodegenerative diseases and can form insoluble cytoplasmic inclusions in the motor neurons of patients with amyotrophic lateral sclerosis [7]. We focused mainly on analysis of SGs recruitment of FUS (Fused in Sarcoma). FUS is one of the most abundant nuclear RBPs which interact with PAR and can be PARylated. In addition, we previously found that FUS is translocated to the cytoplasm after hydrogen peroxide treatment in a PARP1-dependent manner [8].

*Methods and Algorithms:* To analyze the formation of SGs and their composition under oxidative stress conditions we used HeLa cell line. The cells were treated with hydrogen peroxide and (or) arsenite, SGs formation in fixed HeLa cells was traced by immunofluorescence of protein/PAR components or in situ hybridization of cy3-labelled poly(dT) with poly(A)mRNA included into SGs. Statistical analysis of SGs number in cell or enrichment of proteins/PAR in SGs were performed using Cell Profiler software. PARP1 and PARP2 activities in cells were inhibited by chemical inhibitors (Olaparib or Talazoparib) or by using PARP1 siRNA. To analyze polysomes formation and translation regulation during PARP1 activation, we used polysome fractionation by sucrose density gradient centrifugation for comparison of polysome profiles in HeLa cells treated with arsenite and (or) hydrogen peroxide in the presence or absence of PARP1 inhibitor olaparib.

*Results:* By analysing SGs assembly in the cytoplasm of HeLa cells, we found that hydrogen peroxide, a potent nuclear PARP's activator, contrary to what might be expected, prevents SGs formation in the cytoplasm in a nuclear PARP1-dependent

manner [9]. Polysome profile analysis also showed that PARP1 activation upon DNA damage limits polysome dissociation, which is consistent with inhibition of SGs assembly. However, when SGs were formed in the cytoplasm, activation of nuclear PARPs positively regulates SGs production and makes them persistent in the cytoplasm. However, we failed to detect an increase of PAR enrichment in SGs, suggesting that PAR synthesized in the nucleus upon PARP1 activation is not recruited in SGs. This finding is very important for the function under stress conditions of many self-associated RNA-binding proteins that interact with PAR and/or undergo post-translational modification via poly(ADP-ribosyl)ation. Among them, FUS has been found to act in response to DNA damage through a PARP1-dependent mechanism [10]. In the present study, we show that PARP1 activation leads to FUS translocation from the nucleus to the cytoplasm and its accumulation in SGs of HeLa cells following hydrogen peroxide-induced oxidative stress, which in turn may promote SGs assembly, regulate an adapted translational response to stress and the aggregation of RNA-binding proteins in neurodegenerative diseases.

**Conclusion:** Overall, our data suggest that PARP1 activation after genotoxic stress can either block the formation of SGs if they were not preformed, but also can represent an aggravating factor in neurodegeneration diseases by preventing the dissociation of preformed FUS condensates.

**Acknowledgements:** The work was supported by RSF (20-14-00086).

### References

1. d'Amours D., Desnoyers S., d'Silva I., Poirier G.G. Poly (ADP-ribosyl) ation reactions in the regulation of nuclear functions. *Biochemical J.* 1999;342(2):249-268.
2. Bonicalzi M.E. et al. Regulation of poly (ADP-ribose) metabolism by poly (ADP-ribose) glycohydrolase: where and when? *Cell Mol Life Scie.* 2005;62(7-8):739-750.
3. Leung A.K.L. et al. Poly (ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Mol Cell.* 2011;42(4):489-499.
4. McGurk L. et al. Poly (ADP-ribose) prevents pathological phase separation of TDP-43 by promoting liquid demixing and stress granule localization. *Mol Cell.* 2018;71(5):703-717.
5. Duan Y. et al. PARYlation regulates stress granule dynamics, phase separation, and neurotoxicity of disease-related RNA-binding proteins. *Cell Res.* 29.3 (2019):233-247.
6. Buchan J.R., Parker R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell.* 2009;36(6):932-941.
7. Leung A.K.L. Poly (ADP-ribose): an organizer of cellular architecture. *J Cell Biol.* 2014;205(5):613-619.
8. Sama R.R.K., Ward C.L., Bosco D.A. Functions of FUS/TLS from DNA repair to stress response: implications for ALS. *ASN Neuro.* 2014;6(4):1759091414544472.
9. Singatulina, Anastasia 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(6):1809-1821.
10. 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;42(1):307-314.

## ***In vitro* characterization of human REV1 low-fidelity DNA polymerase**

Stolyarenko A.<sup>1\*</sup>, Novikova A.<sup>1,2</sup>, Shilkin E.<sup>1</sup>, Poltorachenko V.<sup>1</sup>, Makarova A.V.<sup>1</sup>

<sup>1</sup> *Institute of Molecular Genetics, National Research Center “Kurchatov Institute”, Moscow, Russia*

<sup>2</sup> *MIREA Technological University, Moscow, Russia*

\* *stol@img.ras.ru*

**Key words:** DNA damage, translesion synthesis, DNA polymerases

Whenever DNA damage fails to be repaired, it is encountered by translesion synthesis machinery in order to maintain genomic stability. The main players of translesion synthesis are low-fidelity DNA polymerases. Their functioning can lead to mutagenesis and chemotherapy resistance, and alterations in their activity or expression can lead to various cancers.

REV1 is perhaps the most important of translesion polymerases since it possesses not only a catalytic function, but also a structural function. The former consists in REV1 being primarily a deoxycytidyl transferase enabled by its catalytic domain forming special interactions with the incoming cytidine and located approximately in the middle of the protein. The structural function is performed by the C-terminal domain of REV1 allowing it to serve as a platform for the assembly of other low-fidelity polymerases of translesion synthesis. Studies of human REV1 are hampered by the general difficulty of the isolation of the protein, as well as it being prone to the cleavage of the C-terminal domain.

This work introduces a new method for isolation of intact and highly-active hREV1 from *E. coli* and *S. cerevisiae*. The *S. cerevisiae* preparation was chosen for in-depth analysis, as it is the first full-sized hREV1 protein without a tag isolated from a eukaryotic expression system.

The purified protein demonstrated primarily deoxycytidyl transferase activity when bypassing undamaged nucleotides and the most wide-spread lesions AP-site and 8-oxoguanine. It was shown that the G-G intrastrand crosslink caused by a popular chemotherapy agent cisplatin could not be efficiently bypassed by REV1 *in vitro*. Analysis of DNA lesions that allow to elucidate the mechanism of REV1 (1,N6-etenoadenine and 7-deazaadenine/7-deazaguanine) demonstrated that disrupting Watson-Crick base pairing of the template nucleotide did not inhibit REV1. However, disrupting Hoogsteen base pairing leads to a significant decrease of its activity, which presents biochemical evidence for a unique mechanism of hREV1 active site functioning, in line with previously obtained structural data.

*Acknowledgements:* This work is supported by RSCF grants No. 18-14-00354-P and No. 22-24-20155.

# Molecular dynamics insight into mechanism of nucleotide incorporation specificity of human terminal deoxynucleotidyltransferase

Tyugashev T.E.<sup>1\*</sup>, Fedorova O.S.<sup>1</sup>, Kuznetsov N.A.<sup>1,2</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

\* [tyugashev@niboch.nsc.ru](mailto:tyugashev@niboch.nsc.ru)

Terminal deoxynucleotidyltransferase (TdT), a unique eukaryotic X-family DNA polymerase, displays a clear deoxynucleotide triphosphate preference in its activity template-independent oligonucleotide primer elongation. The enzyme properties have been extensively studied by biochemical, mutational and structural approaches for decades. Multiple kinetic and crystal structure data are available, mostly for human or murine TdT orthologs. In order to complement experimental results known up today, molecular dynamics calculations of several complexes of human TdT with DNA and dNTPs have been carried out in the present study.

The MD simulations of two types of complexes were performed: i) pre-catalytic ternary protein/primer/dNTP complex and ii) post-catalytic binary protein/primer-dN oligonucleotide complex. Both types of complexes were constructed based on the available crystal structures of murine TdT with partially-resolved active site atom positions and homology modelling of human TdT investigated with classical MD approach.

Results of the simulations allow to elucidate the substrate preferences of TdT, with the numbers and ratios of hydrogen bonds between the base of incoming nucleotide and the protein in both pre-catalytic and post-catalytic models. Thus dGTP enjoys both experimentally the highest incorporation rate and theoretically the largest number of contacts in the models, with slight decrease in the post-catalytic state, while dCTP, the slowest to incorporate *in vitro*, also displays the minimal amount of contacts in the pre-catalytic state, with noticeable increase in the post-catalytic state *in silico*. Both dTTP and dATP share middling experimental elongation performance and similar pattern of bonds without drastic change in number between the two theoretically investigated states.

Obtained data shed light on the network contacts between dNTP and enzyme active site in the course nucleotide recognition and primer elongation. These findings expand the understanding of the mechanism behind specific interactions of human TdT with a DNA primer and dNTP.

*Acknowledgements:* This work was supported by Russian Science Foundation Grant No. 21-64-00017.

## Study of interaction of the PARP family DNA-dependent proteins with nucleosomes by atomic force microscopy

Ukraintsev A.A.\*, Belousova E.A., Kutuzov M.M., Lavrik O.I.

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

\* [alan.ukraintsev@gmail.com](mailto:alan.ukraintsev@gmail.com)

**Key words:** nucleosome, structure, PARP1, PARP2, PARP3, AFM

*Motivation and Aim:* It is well known that during the life of a cell, numerous damages accumulate in its genetic material. The cellular signal in response to DNA damage is the ADP-ribosylation reaction catalyzed by DNA-dependent proteins of the poly(ADP-ribose)polymerase (PARP) family, PARP1, PARP2, and PARP3. Studies of the activities of these three proteins to the DNA damage response were carried out mainly on the synthetic DNA substrates – free DNA molecules. However, in the cells of the DNA molecule is a part of chromatin – a protein-nucleic acid complex. The first level of chromatin compaction is a nucleosome. The nucleosome is a complex of DNA and a histone octamer, in which the DNA is wound around a spool of a histone disk like a thread. Today, the effect of DNA-dependent proteins of the PARP family on the structure of the nucleosomal complex has been little studied. In the present work, we investigated this question using an atomic force microscope.

*Methods and Algorithms:* We utilized the model nucleosome to study the effect of DNA-dependent PARP's family proteins on the stability of the nucleosomal complex. The nucleosome was reconstituted according to the work [1]. For this, DNA containing the "Clone 603" sequence of 147 bp long with linker regions of 79 bp and 120 bp long located at the ends was used. To study the effect of DNA-dependent PARP proteins on the stability of the nucleosomal complex, the dissociation constants of the PARP-nucleosome complexes were determined as described in [2]. The formation of PARP-nucleosome complexes was carried out at 37 °C for 15 minutes appropriate concentrations. Next, the reaction mixture was applied to a mica surface previously modified with APS [3]. The mica surface was scanned on a Scanning Probe Microscope (Multimode 8, Veeco-Bruker, USA) in TappingMode. The scanning area was 2\*2 μm with a resolution of 1024 points. The resulting images were processed using the Gwyddion software (v. 2.58). Measurement of various geometric parameters of the structures under study was carried out using the ImageJ software (v. 1.53e).

*Results:* All three DNA-dependent proteins of the PARP family were found to be able to interact with the substrates used. At the same time, most parts of the protein molecules were localized at the entry–exit site of the linker DNA into the nucleosomal disk (70 % of PARP1 molecules, 77 % of PARP2 molecules, and 86 % of PARP3 molecules). In addition, the stabilization of the nucleosomal complex in the presence of PARP1 and PARP3 proteins was observed: the data scattering of the angles' average value between the entry-exit sites of the linker DNA into the nucleosomal disk in the presence of proteins was smaller. PARP2 had no such significant effect on nucleosome stability. Moreover, it was found that the binding of PARP3 protein to the substrate leads to compaction of the nucleosome: the angle between the entry-exit sites was decreased

from 119 degrees for the native substrate to 103 degrees for PARP3-nucleosome complexes. No such results were found for PARP1/2 proteins.

*Conclusion:* The results obtained using model substrates disclosed different effects upon the binding of the DNA-dependent PARP proteins with the nucleosome. It can be assumed that the obtained differences are explained by the variation of the proteins' preferences to the chemical and structural nature of nucleosomal substrates. Nevertheless, the obtained results allow us to better understand the regulatory aspects of the biochemical mechanisms in which DNA-dependent PARP family proteins are involved.

*Acknowledgements:* The study are supported by the RFBR (No. 20-04-00674) and RSF (No. 21-64-00017).

### *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.A., Belousova E.A., Kutuzov M.M., Lavrik O.I. Study of Interaction of the PARP Family DNA-Dependent Proteins with Nucleosomes Containing DNA Intermediates of the Initial Stages of BER Process. *Biochemistry (Moscow)*. 2022;87:331-345.
3. Shlyakhtenko L.S., Gall A.A., Lyubchenko Y.L. Mica Functionalization for Imaging of DNA and Protein-DNA Complexes with Atomic Force Microscopy. *Methods Mol Biol*. 2013;931:295-312.



## Recombinant lactaptin analogon RL2 inhibits TRAIL-induced cell death in breast cancer cell lines by mitophagy

Wohlfromm F.<sup>1\*</sup>, Richter M.<sup>1</sup>, Koval O.<sup>2</sup>, Richter V.<sup>2</sup>, Lavrik I.<sup>1\*</sup>

<sup>1</sup> *Translational Inflammation Research, Medical Faculty, Otto von Guericke University, Magdeburg, Germany*

<sup>2</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

\* *fabian.wohlfrom@med.ovgu.de; inna.lavrik@med.ovgu.de*

**Key words:** breast cancer, mitochondria, lactaptin, RL2, cell death, TRAIL, DISC

The recombinant analogon of lactaptin (RL2) induces cell death of breast cancer cells, however, the molecular mechanism of RL2 action in breast cancer cells remains unknown. RL2 presents a promising therapeutic for treating breast cancer as a single drug or in a co-treatment with well-known therapeutics. In this study, we have investigated the effects of RL2 on cell death in a co-treatment with TRAIL in MDA-MB 231 and MCF-7 breast carcinoma cells. We have found that RL2 inhibits TRAIL-induced caspase-3/7 and caspase-8 activity as well as cell death upon short-term treatment. Further, RL2 reduced TRAIL-induced DISC (death-inducing signaling complex) formation. Moreover, administration of RL2 resulted in mitophagy induction at the early time intervals as well as the upregulation of anti-apoptotic proteins Bcl-2 and Bcl-xL. Additionally, RL2 treatment led to downregulation of the pro-apoptotic protein Bax and components of the TRAIL DISC, in particular procaspase-8 and FADD. Observed effects might contribute to down modulation of the TRAIL response upon the short-term treatment. Furthermore, our study has demonstrated two opposite effects of TRAIL and RL2 co-treatment: it blocks cell death upon short-term co-stimulation and increases cell death upon long-term treatment. Our findings have important implications for cancer therapy, especially for combinatorial treatments.

# Interaction of DNA replication and DNA repair systems with AP sites adducts with proteins, peptides and low-molecular-weight compounds

Yudkina A.<sup>1,3\*</sup>, Shilkin E.<sup>2</sup>, Boldinova E.<sup>2</sup>, Makarova A.<sup>2</sup>, Dvornikova A.<sup>1,3</sup>, Bulgakov N.<sup>1</sup>, Kim D.<sup>1,3</sup>, Zharkov D.<sup>3</sup>

<sup>1</sup> *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Molecular Genetics, National Research Center "Kurchatov Institute", Moscow, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *ayudkina@niboch.nsc.ru*

**Key words:** AP sites, DNA polymerases, AP endonucleases, DNA-protein cross-links, DNA-peptide cross-links, methoxyamine, polyamines

*Motivation and Aim:* Apurinic/apyrimidinic sites (AP sites) are the most frequent type of spontaneous DNA damage that occurs in the human genome at the steady-state level of ~20,000 per cell under normal conditions, with their number increasing significantly when cells are exposed to genotoxic stress. AP sites arise from the loss of DNA bases through spontaneous hydrolysis of the *N*-glycosidic bond, under the influence of genotoxic factors and as intermediates in DNA repair. AP sites and their derivatives readily react with nucleophilic groups. Since DNA in the cell is always associated with many proteins, this often leads to the formation of DNA-protein cross-links (DPCs). However, the mechanism of interaction of replication machinery with DPCs is not fully understood. The repair of DPCs begins with proteolysis of the protein part to a small peptide by the proteasome or one of dedicated proteases. However, at the moment it is completely unclear what happens with the resulting DNA-peptide cross-links. Moreover, AP sites are surrounded by low-molecular weight nucleophilic compounds such as polyamines. Methoxyamine is another small molecule that specifically inhibits BER by reacting with natural apurinic/AP sites, which prevents their cleavage by APEX1 endonuclease, however its miscoding properties remains unclear.

*Methods and Algorithms:* Here we used classical biochemical approaches to obtain and to study models of chemically determined AP site conjugated with nucleophilic molecules, namely proteins, peptides, polyamines, and methoxyamine. The nature of the adducted AP sites was confirmed by MALDI mass spectrometry.

*Results:* We have proposed the mechanism of interaction of DNA polymerases belonging to different families with DPCs in terms of protein surface contacts. We have studied the ability of DNA polymerases to carry out synthesis in the presence of peptide cross-links with AP sites in DNA, polyamine conjugated AP sites, methoxyamine-conjugated AP sites and the ability of nucleases involved in DNA repair to hydrolyze these lesions. *Conclusion:* AP sites readily react with lots of nucleophilic cellular compounds. The adducted AP sites are common phenomenon in the living cell, which necessitates the presence of dedicated pathways for their tolerance and repair.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant 21-74-00061 to A.V.Y.). A.V.Y. is a recipient of the Russian Presidential Fellowship for Young Scientists (SP-174.2021.4).

## Structure and function evolution in DNA repair

Zharkov D.O.

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

*Novosibirsk State University, Novosibirsk, Russia*

*dzharkov@niboch.nsc.ru*

**Key words:** DNA damage, DNA repair, base excision repair, nucleotide excision repair, mismatch repair, direct reversal, evolution

*Motivation and Aim:* To perform its function as the genetic material, DNA is subject to constant surveillance and repair by many proteins that fix omnipresent DNA lesions, which otherwise could induce mutations or stall replication. Cellular life evolved several distinct enzymatic mechanisms to repair DNA, including direct damage reversal (DR), base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), recombination repair, and non-homologous end joining, assisted by damage prevention and tolerance pathways such as nucleotide pool sanitization and translesion DNA synthesis. DNA repair enzymes employ many structural folds to recognize DNA damage and catalyze the necessary enzymatic reactions. These folds are often closely related to those present in other proteins that are not involved in DNA repair. The evolution of the repair systems in their present forms apparently occurred very early in the history of life, and very little is known about it.

*Methods and Algorithms:* We have traced the evolutionary relationship of all protein folds in the proteins involved in DR, BER, NER, and MMR, as well as their appearance in the major lineages of life. This allowed us to infer the principal events of appearance and disappearance of certain repair pathways and subpathways in evolution. We have also compared the likely timing of these events with the known chronology of the Earth's chemosphere.

*Results:* BER likely have evolved from an ancient pathway predating the Bacteria/Archaea split that employed endonucleases universally recognizing non-canonical DNA elements, and diversified involving a variety of DNA glycosylases with the increasing complexity of the Earth's chemosphere, with a major impact from the Great Oxidation Event ~2.4–2.0 Ga. NER had likely independently evolved in Bacteria and Eukaryota, with no traces of this pathway in Archaea. MMR as a complete repair pathway also probably evolved twice but was based on an early system of DNA quality control, possibly related to today's bacterial conjugative DNA transport system. DR, although today represented by several structurally unrelated systems (photolyases, suicidal methyltransferases and oxidative demethylases), can be traced to a very ancient group of radical SAM proteins.

*Conclusion:* DNA repair systems likely appeared at almost the same time as DNA replication when the main genotoxic factors were replication errors, UV light and ionizing radiation that produced a limited amount of lesions in oxygen-free aqueous solutions. DNA repair greatly diversified with the expansion of the DNA damage repertoire, including metabolic byproducts and reactive oxygen species.

*Acknowledgements:* The study is supported by RFBR (20-04-00554-a) and Russian Ministry of Science and Higher Education (AAAA-A20-120101690009-7).

# 12

Section “Systems biology of aging:  
experimental and computational  
approaches”



## Multi-stage model of aging and carcinogenesis

Anisimov V.N.

*N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia  
aging@mail.ru*

**Key words:** aging, carcinogenesis, multi-stage model

The current level of knowledge about the nature of aging allows us to approach the creation of a multi-level model of the multi-stage human aging process, taking into account the events that take place during the aging of organisms at different levels of integration: molecular, cellular, tissue, organ, system, organism and development of age associated pathology including cancer. The report formulates the main provisions of the multi-stage model of aging and carcinogenesis, as well as the principles of its construction, approaches to its verification and assessment of its predictive power. There are three levels of systems: homeostatic (whole organism), systemic and cellular-molecular-genetic. Definitions of normal and pathological aging are given, principles of interaction of systems of different levels are postulated. On the examples of manifestations of aging at different levels under the expose to factors, on the one hand, accelerating it (chemical carcinogens, ionizing radiation, some parasites, biological circadian desynchronization, alimentary obesity), and on the other hand, factors preventing accelerated aging and carcinogenesis (administration of melatonin, rapamycin, metformin etc.), possible approaches to using the model to predict the rate of aging and the duration of healthy life of a person and his populations living in different environmental conditions, or exposed to occupational factors accelerating aging, and evaluating the effectiveness of the use of certain means of preventing premature aging and carcinogenesis.

*Acknowledgements:* The study is supported by the Russian Scientific Foundation, grant No. 20-15-00330.

### *References*

1. Anisimov V.N. Aging and cancer biology. In: Extermann M. (Ed.). *Geriatric Oncology*. Cham: Springer Nature Switzerland AG, 2020;91-109. doi: 10.1007/978-3-319-57415-8.
2. Anisimov V.N. Aging delay: of mice and men. *Acta Biomedica*. 2021;92(1):e2021073. doi: 10.233750/abm.v92i1.11273.

## The daily cycle of glutamate in the retina of OXYS rats during aging and the development of retinopathy

Antonenko A.\*, Telegina D.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\**a.k.antonenko98@gmail.com*

**Key words:** glutamate, retina, retinopathy, age-related macular degeneration, aging, OXYS rats

*Motivation and Aim:* Glutamate performs many important functions in the retina, such as visual signal transmission and regulation of the circadian rhythm [1]. Glutamate contributes to the synchronization of the activity of many retinal cells and changes in the expression of their genes. Information on daytime and nighttime fluctuations in glutamate is limited and inconsistent. However, the idea is put forward that biochemical changes and circadian rhythms of glutamate may be risk factors for the development of neurodegenerative processes in the retina, in particular, the development of age-related macular degeneration (AMD) [2], which remains the top cause of blindness and irreversible visual impairment in industrialized countries. Nowadays, there is some information about deviations in neurotransmitter systems in the retina in AMD. Nonetheless, there is little information on retinal changes in the glutamatergic systems during aging and even less at different stages of AMD development. It is also not clear whether these changes are a cause or consequence of the development of the neurodegenerative process. Expectedly, the manifestation of clinical signs of diseases is preceded and accompanied by pathological changes at the molecular level, which are difficult to study in humans.

*Methods and Algorithms:* In this study, we used OXYS and Wistar rats as a control. OXYS rats are a unique model of premature aging whose manifestation is early development of retinopathy similar to AMD. According to our data, even at a young age, the retinopathy that develops in OXYS rats is similar to the dry atrophic AMD type in humans. Male OXYS rats at a preclinical stage (20 days), early stage (3 months), and late stage (18 months) of the pathology and normal controls (age-matched male Wistar rats) were used in this work. To investigate daily cycle of glutamatergic systems in retina the sampling took place during the day (10 a.m.) and at night (2 a.m.). The concentration of unbound glutamate in rat retina samples was measured by colorimetric method. Western blot analysis was used to assess the levels of key enzymes of glutamate metabolism (glutaminase and glutamine synthetase), the ionotropic glutamate receptor (NMDA(R1)) and the glutamate transporter (EAAT1/GLAST). To study the localization of the above proteins, immunohistochemical staining with fluorescent antibodies was used.

*Results:* In Wistar rats with normal aging, there was a characteristic daily difference of glutamate concentrations in all age: increased in daytime and decreased in nighttime. In OXYS rats, the difference in daily concentration disappeared with age. Glutamate is synthesized *de novo* from glutamine in glial cells (astrocytes and Müller cells) by glutaminase. Glutamine synthetase catalyzes the amidation of glutamate to glutamine. Western blotting revealed that the daytime and nighttime levels of glutaminase did not change with age in the retina of OXYS and Wistar rats. But ANOVA showed that the protein level of glutaminase was higher in OXYS rats than in Wistar rats in both daytime

and nighttime at the all ages. The significant differences were found between in 18-months-old OXYS and Wistar rats at the daytime ( $p < 0,05$ ) and between 20-days-old and 3-month-old OXYS and Wistar rats at the nighttime ( $p < 0.05$ ).

At the daytime the protein level of glutamine synthetase increased by the age of 18 months in Wistar rats but did not change with age in the retina of OXYS. By age 18 months, we detected significant upregulation of the glutamine synthetase protein in OXYS rats as compared with Wistar rats ( $p < 0.05$ ). At the nighttime the protein level of the glutamine synthetase increased from the age of 20 days to 3 months in Wistar rats, while no significant differences were found in OXYS rats. No significant interstrain differences were found in the rat retina in night.

In summary, the level of glutaminase was higher and level of glutamine synthetase was lower in old OXYS rats than in old Wistar rats. These data indicated that the glutamine/glutamate ratio was shifting in favor of glutamate in the retina of OXYS rats during development of retinopathy.

GLAST (glutamate aspartate transporter 1, also is known as EAAT1) removes glutamate from the extracellular space and is needed to maintain low nontoxic extracellular glutamate concentrations. At the daytime, Western blotting uncovered a significant increase in GLAST protein levels with age in the retina of OXYS and Wistar rats ( $p < 0.05$ ). Also it was showed that the level of GLAST was lower in OXYS rats than in Wistar rats. By age 3 and 18 months, we detected a significant decline of the GLAST protein level in OXYS rats as compared with Wistar rats ( $p < 0.05$ ).

At night, the level of GLAST protein increased with age in Wistar rats ( $p < 0.05$ ), but did not change in OXYS rats either. During the period of active disease progression (18 months), the level of the GLAST transporter in OXYS rats was significantly lower than in Wistar rats ( $p < 0.05$ ).

Ionotropic NMDA glutamate receptors are responsible for fast neurotransmission. In this work, we assayed glutamate ionotropic receptor N-methyl-D-aspartate (NMDA) receptor subunit NR1 (NMDAR1). The study of the NMDAR1 protein level in the daytime by Western blot did not show significant differences between ages and lines. At the nighttime we detected increasing NMDAR1 protein level in OXYS rats with age and no age-dependent alterations in Wistar rats. The significant upregulated of NMDAR1 protein level in OXYS as compared Wistar rats was detected at the age of 3 and 18 months ( $p < 0.05$ ).

*Conclusion:* It has been shown that similar changes occur in the retina of rats during healthy aging and the development of retinopathy. So the level of the glutamate transporter increases with age in the retina of rats. The development of retinopathy in the retina of OXYS rats is accompanied by a smoothing of the daily rhythm of glutamate, an increase in the level of the glutaminase enzyme and a decrease in glutamine synthetase, as well as a decrease in the level of the glutamate transporter. These alterations are leading to disruption of the glutamatergic system in the OXYS rats retina and an increase in degenerative changes.

*Acknowledgements:* The study is supported by the Russian Science Foundation (grant No. 21-75-00029).

## References

1. McMahon D.G. et al. Circadian organization of the mammalian retina: from gene regulation to physiology and diseases. *Prog Retin Eye Res.* 2014;39:58-76. doi: 10.1016/j.preteyeres.2013.12.001.
2. Jones B.W. et al. Retinal remodeling and metabolic alterations in human AMD. *Front Cell Neurosci.* 2016;10:103. doi: 10.3389/fncel.2016.00103.

## Alterations of glutamatergic/GABAergic systems during healthy aging and Alzheimer's disease-like pathology in rats

Burnyasheva A.O. \*, Rudnitskaya E.A., Telegina D.V.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* burn.alena2505@mail.ru

**Key words:** glutamate, GABA, hippocampus, Alzheimer's disease, OXYS rats

*Motivation and Aim:* Alzheimer's disease (AD) is an incurable neurodegenerative disorder which progresses with age. Disturbance of the balance of the excitatory glutamatergic and inhibitory GABAergic systems accompanies the development of AD. However, the question of whether this dysregulation is the cause or consequence of AD development remains unanswered.

*Methods and Algorithms:* We used senescence-accelerated OXYS rats as a suitable AD model [1] to assess the relationship between changes in glutamatergic and GABAergic systems and the development of AD signs. Wistar rats were used as a control. The content of enzymes regulating glutamate-GABA cycle—glutaminase, glutamate decarboxylase (GAD67), glutamine synthetase (GS) and GABA-transaminase (GABA-T)—was analyzed in the hippocampus of OXYS and Wistar rats at 1.5, 4, 12, and 18 months of age by Western blot and immunohistochemistry. Additionally, the content of NMDA receptor-1 (NMDAR1), AMPA receptor-1 (AMPA1), GABA A receptor alpha-1 (GABAAR1) and glutamate transporter-1 (GLT-1) were evaluated.

*Results:* We didn't observe significant age-related and interstrain changes in the levels of glutaminase and GS. Meanwhile, the level of GABA-T was lower and GAD67 was higher in OXYS rats throughout the lifespan. The level of NMDAR1 decreased with age in the hippocampus of Wistar rats; however we didn't observe this decrease in OXYS rats. Thus, NMDAR1 level was higher in OXYS rats at 18 months of age. The level of GABAAR1 was higher in OXYS rats compared to Wistar rats at 3 months of age. GLT-1 level didn't change throughout the lifespan in Wistar rats, whereas in OXYS rats it increased by the age of 12 months.

*Conclusion:* Development of AD signs in OXYS rats occurs against background of altered balance of the glutamatergic and GABAergic systems with shift towards higher GABA levels which may contribute to the progression of neurodegeneration. Increased NMDAR1 content in OXYS rats at 18 months of age may be considered as compensation aimed to enhance glutamatergic neurotransmission at progressive stage of AD-like pathology.

*Acknowledgements:* This work was supported by grant from the Russian Science Foundation (project No. 21-75-00029).

### References

1. Stefanova N.A., Kozhevnikova O.S., Vitovtov A.O., Maksimova K.Y., Logvinov S.V., Rudnitskaya E.A., Korbolina E.E., Muraleva N.A., Kolosova N.G. Senescenceaccelerated OXYS rats: A model of age-related cognitive decline with relevance to abnormalities in Alzheimer disease. *Cell Cycle*. 2014;13(6):898-909.



## **Hsp70 and donors of H<sub>2</sub>S in aging-associated diseases and inflammation**

Evgen'ev M.B., Garbuz D.G., Zatsepina O.G.

*Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

*Abstract:* Heat shock protein 70, in humans is a key component of the machinery protecting the neuronal cells from various stress conditions whose production significantly declines in the course of aging. Alzheimer's disease (AD) is the most prevalent neurodegenerative pathology in the growing population of elderly humans and leads eventually to dementia and death. Despite tremendous efforts, no effective treatment for AD is currently available. Molecular chaperone Hsp70 plays protective role in various neurodegenerative disorders. Various data suggest that Hsp70 and other molecular chaperones function as a complex neuroprotective system, which fails in the brains of aged people and AD patients. In special experiments we demonstrated that rHsp70 effectively crosses the blood-brain barrier when administered intranasally. We also have shown that chronic administration of recombinant Hsp70 (rHsp70) decreased beta-amyloid level and preserved neuron density in two mouse models of Alzheimer disease. In both cases rHsp70 restored behavior and memory disturbed by Alzheimer disease and aging. We also explored the effect of rHsp70 on neurons morphology and survival in the cortex and the hippocampus of transgenic animals. The proportion of pathologic neurons decreased drastically in Hsp70-treated animals. Therefore, Hsp70 treatment of model mice protects neurons from deterioration and death in brain areas most affected in AD patients. Deep sequencing studies enabled to reveal candidate genes and signal pathways underlying beneficial effects of rHsp70 treatment. In our experiments we also demonstrated that intranasal administration of exogenous recombinant human Hsp70 can promote longevity in male but not female mice. The Hsp70 treatment normalized the synthesis of synaptophysin in aged mice and decreased accumulation of lipofuscin which represents the marker of aging and neurodegeneration processes. The treatment with rHsp70 performed by our group and colleagues from other Institutes demonstrated highly beneficial effects in several models of age-associated diseases including infarction and stroke.

Hydrogen sulfide (H<sub>2</sub>S) represents another compound with strong anti-inflammatory properties able to modulate many biological processes, including ageing. Initially considered a hazardous toxic gas, it is now recognized that H<sub>2</sub>S is produced endogenously and is a key mediator of processes that promote longevity and improve late-life health. In this reports, we will consider the key developments in our understanding of this gaseous signaling molecule in the context of health and disease, discuss potential mechanisms through which H<sub>2</sub>S can influence processes central to ageing and highlight the emergence of novel H<sub>2</sub>S and Hsp70-based therapeutics to treat age-associated diseases in humans.

*Acknowledgements:* This work was supported by grant from the Russian Science Foundation (project 17-74-300-30).

## **Aqueous humor complement factor I concentration is associated with SNP of *CFI* gene**

Fursova A.Zh.<sup>1,2,3</sup>, Nikulich I.F.<sup>1,2,3</sup>, Derbeneva A.S.<sup>1,2,3</sup>, Kolosova N.G.<sup>1</sup>, Kozhevnikova O.S.<sup>1\*</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *State Novosibirsk Regional Clinical Hospital, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State Medical University, Novosibirsk, Russia*

\* *oidopova@bionet.nsc.ru*

Complement activation is a central innate immune mechanism. Aberrant alternative complement activation has been linked to the pathogenesis of age-related macular degeneration (AMD). Complement factor I (FI, encoded by the *CFI* gene) is a negative regulator of complement and plays an important role in limiting complement activation. There is increasing evidence to suggest that local alterations in FI levels may be a priming event in AMD.

Here we evaluated the association of systemic and locally produced FI levels with AMD and a regulatory *CFI* SNP (rs2285714). FI levels were measured by ELISA in plasma (54 AMD cases and 25 age-matched controls) and in aqueous humor (24 AMD cases and 22 age-matched controls). Aqueous humor (AH) samples were obtained from AMD patients before intravitreal injection of anti-VEGF and elderly controls during cataract surgery. rs2285714 was genotyped by real-time PCR. There was no difference in the FI plasma levels in AMD ( $27.8 \pm 7.1$   $\mu\text{g/mL}$ ) and control ( $25.3 \pm 5.3$   $\mu\text{g/mL}$ ) patients. The median FI level in AH was lower in AMD patients (33.1 ng/mL) compared to controls (58.4 ng/mL), although the difference was not significant ( $p = 0.12$ ). Huge concentration gradient between the plasma and AH was found. rs2285714 was significantly correlated with FI concentration in AH, but not in plasma. The median FI level in AH was significantly higher in individuals with the CC genotype (82.5 ng/mL) compared to individuals with the heterozygote (35.9 ng/mL,  $p = 0.002$ ) and TT genotypes (25.9 ng/mL,  $p = 0.0002$ ). These findings support that the importance of local rather than systemic FI production in AMD.

*Acknowledgements:* Supported by RSF 21-15-00014.

## Iron-sulfur nitrosyl complexes as potential therapeutic agents for anti-aging pharmacology and radiation medicine

Gorban' E.N.<sup>1</sup>, Koltover V.K.<sup>2\*</sup>, Sanina N.A.<sup>2</sup>

<sup>1</sup> Institute of Gerontology, NAMS, Kiev, Ukraine

<sup>2</sup> Institute of Problems of Chemical Physics, RAS, Chernogolovka, Russia

\* koltover@icp.ac.ru

**Key words:** aging, nitric oxide (NO), iron-thionitrosyl complexes, radiation, pharmacology

*Motivation and Aim:* It is well known that many of the age-related dysfunctions, including cardiovascular and Alzheimer diseases and the dysfunctions in renal and immune systems, to name just a few, stems from the age-dependent violations in the regulatory functions associated with the nitric-oxide (NO) radical [1–5]. The functional disorders in the NO system are also observed after the organism has been subjected to ionizing radiation. In this regard, iron-thionitrosyl complexes, as the long-living reservoirs and transport remedies of NO, may be of interest for anti-aging therapy and anti-radiation medicine. Here, we present the evidences of geroprotective effects of the iron-sulfur nitrosyl complex,  $\text{Na}_2[\text{Fe}_2(\text{S}_2\text{O}_3)_2(\text{NO})_4] \cdot 4\text{H}_2\text{O}$  (FeSNO), as the novel NO donor for anti-aging therapy.

*Methods and Algorithms:* The thionitrosyl iron complex FeSNO was synthesized in the Institute of Problems of Chemical Physics, RAS [5]. Experiments on the studies of geroprotective and radioprotective effects of FeSNO were performed in the Institute of Gerontology, NAMS. Wistar male rats, the adult animals (7–8 month age) and the old animals (24 month age at the beginning of the experiments), were daily administered intraperitoneally with FeSNO for 14 days. In the experiments on the studies of radioprotective effects of FeSNO, the adult animals were exposed to X-ray irradiation. Experiments on measuring the intensity of free-radical lipid peroxidation, i.e. the levels of malonyldialdehyde (MDA) and thiobarbituric acid, and the levels of activity of the antioxidant enzymes, i.e. superoxide dismutase (SOD), catalase (Cat), glutathione peroxidase (GPO), and glutathione reductase (GR) in liver, heart, and brain of the animals, as well as measurements of the stable NO-metabolites ( $\text{NO}_2^-$ ,  $\text{NO}_3^-$ ) in blood plasma and homogenates of heart and aorta tissues and the measurements of glucose, hemoglobin and glycated hemoglobin in erythrocytes were performed by the standard methods.

*Results:* The prolong service testing, for 605 days, has proved that the administration of rats with FeSNO gives the reliable effect of the lifespan prolongation, 30 percent extension, as compared to the control animals. In addition, the FeSNO administration restores the age-related changes in the levels of MDA as well as the changes in activities of the antioxidant defense enzymes (SOD, Cat, GPO, GR) in liver, heart and brain tissues of the animals. Furthermore, the rats of 7–8 months age were subjected to X-ray irradiation (5 Gy, sublethal dose), after which the irradiated animals were daily administered with FeSNO for 30 days. Again, it was revealed that FeSNO restores the negative impacts of the radiation on activities of the antioxidant protection enzymes, SOD, Cat, GPO and GR, and prevents the activation of peroxidation of lipids in tissues of the irradiated animals. Besides, FeSNO positively impacts on the nitric oxide

metabolism in blood plasma, heart, and aorta of the irradiated animals, as well as on the functional indexes of metabolism of carbohydrates.

*Conclusions:* Thus, FeSNO exhibits the remarkable geroprotective effect as well as the remarkable radioprotective effect. The geroprotective effect evidences that this compound is low toxic. It means that the nitrosyl complexes of iron with the functionalized sulfur-containing ligands, as the NO donors, hold considerable promise as potential agents in creating the novel drugs for anti-aging pharmacology as well as the drugs for radiation medicine, in part, the radiomitigators for minimization of consequences of radiation damages of healthy tissue cells in clinical oncology.

*Acknowledgements:* This work was supported by the Institute of Gerontology, NAMS, Ukraine, and the Ministry of Science and Higher Education of the Russian Federation (theme No. AAAA-19-119071890015-6 (075-15-2019-1662)).

### *References*

1. Murad F. Discovery of some of the biological effects of nitric oxide and its role in cell signaling. *Biosci Rep.* 2004;24(4-5):452-474.
2. Gantner B.N. et al. Nitric oxide in cellular adaptation and disease. *Redox Biol.* 2020;34:101550.
3. Gorban' E.N. et al. Korargin correction of radiation-induced changes in the levels of stable NO metabolites and of lipid peroxidation in blood and tissues of the rats of various ages. *Problems Aging Longevity.* 2012;21(3):266-273.
4. Koltover V.K. Free radical timer of aging: From chemistry of free radicals to systems theory of reliability. *Curr Aging Sci.* 2017;10(1):12-17.
5. Sanina N.A., Aldoshin S.M. Structure and properties of nitrosyl iron complexes with functional sulfur containing ligands. *Russ Chem Bull.* 2011;60(7):1223-1251.

## Regional principle of aging dynamics of visceral lymph nodes of different lymphatic region

Gorchakov V.<sup>1,2\*</sup>, Kolmogorov Yu.<sup>1</sup>, Gorchakova O.<sup>1</sup>, Nikolaychuk K.<sup>2</sup>, Demchenko G.<sup>3</sup>, Bekeneva K.<sup>2</sup>

<sup>1</sup> *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Institute of Genetics and Physiology, Science Committee Ministry of Education and Science of Republic Kazakhstan, Almaty, Kazakhstan*

\* [vgorchak@yandex.ru](mailto:vgorchak@yandex.ru)

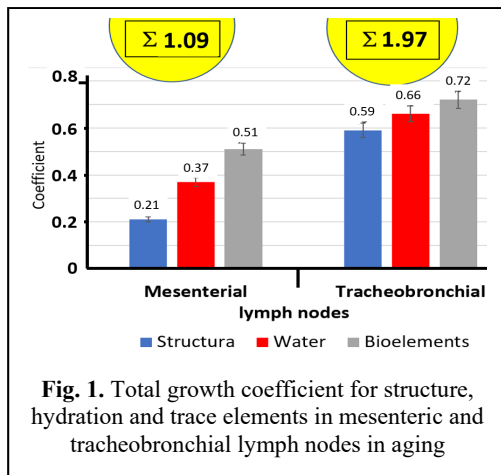
**Key words:** lymph nodes, lymphatic region, hydration, trace elements, gerontology

*Motivation and Aim:* Lymph nodes are grouped into areas belonging to different organs according to the concept of lymph region [1]. Lymphatic regions of organs of respiratory and digestive system change unequally when aging, as regions have features of contact with the external environment. Lymph node is considered as marker of state of drained lymph region [1] and lymph node can be used to estimate rates of changes of morphophysiological parameters of lymph node in aging. The lymph node is a triadic homeostatic system including structural (lymphoid), hydration, and trace elements parts. These components allow to maintain lymph node workability in different periods of life [2]. The purpose of the study is to determine the intensity of age-induced changes of visceral lymph nodes including the structure, hydration and concentration of trace elements.

*Methods and Algorithms:* The experiment was performed on 160 white rats of Wistar with division into an experimental group (rat age is 18–20 months) and a control group (rat age is 3–5 months). Age of rat groups is taken for the final and base period for comparative dynamic estimation of lymph nodes. Mesenteric and tracheobronchial lymph nodes were selected as the subject of the study. We used the following a histological method, thermogravimetric method [3], X-ray fluorescence analysis using synchrotron radiation (RFA SI) [4] and statistical method using the StatPlus Pro 2009 program with determination of basic indicators.

*Results:* The emergence of lymph theory of aging and concept of lymph region [1, 2] poses the need to take into account changes in regional lymph node in aging. Not knowing what happens in organs and regional lymph nodes, we cannot consider age-dependent changes (aging) in isolation. Control at the exit from the lymph region is carried out by the lymph node [1]. Lymphatic regions of the respiratory system and gastrointestinal tract differ from each other in their morphophysiological parameters. The age intensity of changes occurs asynchronously in the mesenteric and tracheobronchial lymph nodes. The mesenteric lymph node undergoes aging to a greater extent. Evidence is the sprawl of connective tissue inside the lymph node and the coarsening of the capsule. There is an intense decrease in cortical plateau, lymphoid follicles and lymphatic sinuses. There is expansion of the medullary substance due to medullary cords with exposure of the reticular stroma of the lymph node. The rate of change indicates the regional specificity of the mesenteric lymph node aging in the intestinal lymphatic region. In aging, the tracheobronchial lymph node is characterized

by less intense changes, demonstrating a significant difference with the mesenteric lymph node. The largest rate of change is noted by lymphoid follicles with a herminative center, medullary cords. Paracortex decreases equally in mesenteric and tracheobronchial lymph nodes. We assume that the difference in structural response to aging depends in large part on the original morphotype of a lymph node. The lymph region with different antigenic load when in contact with the external environment causes accelerated aging of the mesenteric lymph node compared to the tracheobronchial lymph node.



We used to estimate the total growth coefficient, which takes into account the three components of the lymph node – structure (compartments), aqueous saturation and trace element concentration. These three components allow an assessment of the overall operability of a lymph node (Fig. 1). The dynamics of aging of regional lymph nodes are different and are associated with the identity of the drained lymph region. The largest value of the total coefficient is in the tracheobronchial lymph node and a smaller value is typical of the mesenteric lymph node (see Fig. 1). Lymph nodes of

different lymphatic regions are subject to age-dependent changes to varying degrees. The mesenteric lymph node is more susceptible to senile changes, as its parameters regarding structure, hydration, and trace elements differ significantly from the base level. Senile changes are explained by the peculiarity of the lymphatic region, which includes the “organ –lymph node” system, where the organ is dominant, as it directly contacts the external environment [1, 2].

**Conclusion:** Age changes in mesenteric and tracheobronchial lymph nodes occur at different rates. This proves the regional principle of aging depending on the morphophysiological features of the lymphatic region. The mesenteric lymph node is subject to age-dependent changes to a greater extent than the tracheobronchial lymph node, judging by the basic indicators for structure (lymphoid tissue), hydration and trace element concentration. Belonging to different lymph regions determines the morphophysiological response of lymph nodes at different periods of life.

**Acknowledgements:** The study is supported by the state task to research work (FWNR-2022-0012). The work was done on equipment of the G.I. Budker’s Institute of Nuclear Physics, SB RAS (RFMEFI62119X0022).

## References

1. Borodin Yu.I., Gorchakova O.V., Suhovershin A.V., Gorchakov V.N. et al. The concept of lymphatic region in preventive lymphology. Novosibirsk: LAP LAMBERT Academic Publishing, 2018.
2. Gorchakova O.V., Gorchakov V.N., Kolmogorov Yu.P., Nurmakhanova B. et al. Microelement profile and structure of regional lymph nodes during senile involution of lymphoid tissue. *Archiv Euro Medica*. 2021;11(1):48-51. doi: 10.35630/2199-885X/2021/11/1.9.
3. Farashchuk N., Rakhmanin Yu. Water is a structural basis of adaptation. Moscow, 2004.
4. Piminov P. Synchrotron Radiation Research and Application at VEPP-4. *Physics Procedia*. 2016;84:19-26. doi: 10.1016/j.phpro.2016.11.005.

## **Biomarkers of aging and age-related diseases: from linear regression to explainable artificial intelligence**

Ivanchenko M.

*Lobachevsky University, Nizhny Novgorod, Russia  
ivanchenko@unn.ru*

**Key words:** aging, DNA methylation, artificial intelligence

*Motivation and Aim:* Discovery of aging biomarkers and development of biological clocks that enable precise estimation of calendar age and sensitive detection of age acceleration, related to the manifestation of age-related diseases remains central to the field. Despite such recent successes as the Horvath and Hannum epigenetic clocks, PhenoAge clocks and GrimAge clocks, there is still a keen need to (1) encompassing novel and multi-level biomarkers, (2) elaborating the systems description of aging, (3) developing the age estimators based on advanced Artificial Intelligence tools, including the algorithms that allow for explaining AI based predictions both cohort-wide (population-wide) and individual-wise (personalized).

*Methods and Algorithms:* We investigate simultaneous DNA methylation (Illumina EPIC) and immunological data (multiplex analysis) from human peripheral blood samples for the control group and patients with end-stage renal disease (ESRD) receiving hemodialysis. We calculate several epigenetic age estimators, build a linear-regression and elastic-net based immunological age-estimator ipAGE, and implement AI-based age-estimators with individual age-acceleration explained by SHAP values.

*Results:* We demonstrated accelerated aging in terms of the most common epigenetic age estimators in ESRD patients, developed a new clock/predictor of age based on the inflammatory/immunological profile (ipAGE) and identified the differentially expressed inflammatory/immunological biomarkers [1]. IpAGE appeared to be more sensitive than epigenetic clocks in quantifying the accelerated aging phenotype. Moreover, we developed the mixed epigenetic-immunological AI-based age estimators and explain AI models, identifying epigenetic and immunological factors responsible for age acceleration in each single participant.

*Conclusion:* The obtained results uncover and confirm the deep relation between aging and remodeling of the human immune system, previously conceptualized as the inflammaging theory. The developed immunological clocks allow for a direct inference of immune system aging and shed light on the underlying mechanisms. The most to-date advanced methods of Explainable Artificial Intelligence not only rank the most important populational biomarkers of immunological and epigenetic aging, but allow for the first time to differentially identify specific biomarkers that are responsible for age acceleration of each particular individual.

*Acknowledgements:* The study is supported by the Ministry of Science and Higher Education of the Russian Federation, agreement No. 075-15-2020-808. We also acknowledge non-financial provisions of the Ministry of Science and Higher Education of the Russian Federation agreement No. 075-15-2021-639.

### *References*

1. Yusipov I., Kondakova E., Kalyakulina A., Krivonosov M., Lobanova N., Bacalini M.G., Franceschi C., Vedunova M., Ivanchenko M. Accelerated epigenetic aging and inflammatory/immunological profile (ipAGE) in patients with chronic kidney disease. *GeroScience*. 2022;44(2):817-834. doi: 10.1007/s11357-022-00540-4.

## Fighting Alzheimer's disease: is there a chance to conquer?

Khokhlov A.N.

*Evolutionary Cytoogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russia*

*khokhlov@mail.bio.msu.ru*

**Key words:** Alzheimer's disease, senile dementia, aging, model animals, beta-amyloid, free radicals

Being a complete pessimist [1], I always try to avoid overly enthusiastic responses to scientific articles that report the emergence of new promising means of preventing or treating a particular disease. This also applies to research on Alzheimer's disease (AD). Therefore, to the question posed in the title, in my opinion, the following answers are possible, among others: 1) the pathogenesis of AD is still unclear; 2) we do not yet know how to treat AD; 3) the first and second together; 4) AD is not a disease at all; 5) if AD is still a consequence of aging, then we must keep in mind that we also still know almost nothing about the process. In the late 1980s, at a symposium on brain aging in Moscow, D.R. Crapper McLachlan told us about his concept [2] of the occurrence of AD, according to which its pathogenesis is based on aluminum that enters the body with food and drinking water, in particular, when using aluminum utensils. However, in subsequent years, there was no confirmation of this hypothesis. Moreover, in many investigations the inconsistency of this concept was proved. At the same time, numerous works were published in which the leading role in this process of free radicals was postulated (for example, [3]), which were proposed to be combated with the help of antioxidants. However, this idea was later found to be wrong. The Alzheimer's Society website has a dedicated page on the possible role of antioxidants in the prevention and treatment of AD.

It notes that to date there are no reliable clinical data on the advisability of using such drugs against AD. And here is the most important thing: on the same page it is said that data from laboratory studies on model animals, on the contrary, indicate that antioxidants help with AD. However, this is not real AD, but a kind of "surrogate" that is similar to it in some signs and symptoms. After analyzing a fairly large number of articles on this issue, I got the impression that almost all the achievements in the search for means of preventing or treating AD are based on work with model animals. Most often, these are mice with genetic abnormalities in the functioning of the nervous system, which in their manifestations are similar to the symptoms of AD in humans. Unfortunately, experiments on such model animals very often lead researchers to incorrect conclusions. This is well illustrated by the story that I called the "bexarotene saga." Bexarotene (Targretin) is an FDA-approved anti-cancer drug, so there are no problems with its research even in humans. However, attempts to use this drug against AD began precisely with experiments on model animals. In particular, in the sensational work of 2012 [4] published in the prestigious journal *Science*, the authors found that bexarotene removes beta-amyloid from the brain of mice with model AD, and also restores their cognitive capabilities, which is reflected in the restoration of the lost ability to build nests from pieces of paper. Unfortunately, a large number of papers later appeared (for example, [5]), which refuted these results. The same applies to clinical studies of the effectiveness of bexarotene in the



treatment of AD. In my opinion, the correct choice of control objects [6] when conducting such experiments is paramount, and the wrong strategy in this matter almost always leads to incorrect conclusions. At the same time, almost all modern work aimed at finding means to combat AD is carried out on model animals, since this requires much less money and time than clinical experiments. And no one has canceled the ethical requirements for research in humans – as a rule, we cannot administer drugs to healthy people and monitor whether they develop AD. Apparently, this is precisely why we have not made any progress so far in surveys of the prevention and treatment of this scourge of modern humanity, the demographic aging of which (an increase in the proportion of older people in the population) has led to a sharp increase in the number of individuals with AD in almost all developed countries. It is also possible (see the beginning of this text) that, however paradoxical it may sound, AD is not a disease at all, but only a set of signs of non-specific senile brain degradation leading to senile dementia, on the study of which, in this case, efforts should be focused both by gerontologists and neuroscientists.

*Acknowledgements:* This study was performed under the state assignment of Moscow State University, project No. 121032300215-6.

### *References*

1. Khokhlov A.N. Reflections of a pessimistic gerontologist or why we still do not live 1000 years. *Moscow Univ Biol Sci Bull.* 2021;76(4):239-243.
2. Crapper McLachlan D.R. Aluminum and Alzheimer's disease. *Neurobiol Aging.* 1986;7(6):525-532.
3. Retz W. et al. Free radicals in Alzheimer's disease. *J Neural Transm Suppl.* 1998;54:221-236.
4. Cramer P.E. et al. ApoE-directed therapeutics rapidly clear  $\beta$ -amyloid and reverse deficits in AD mouse models. *Science.* 2012;335(6075):1503-1506.
5. Balducci C. et al. The continuing failure of bexarotene in Alzheimer's disease mice. *J Alzheimers Dis.* 2015;46(2):471-482.
6. Khokhlov A.N. et al. On choosing control objects in experimental gerontological research. *Moscow Univ Biol Sci Bull.* 2018;73(2):59-62.

## ROS networks: designs, aging, Parkinson's disease and precision therapies

Kolodkin A.N.

*Luxembourg Institute of Health, Luxembourg, Luxembourg*

*Infrastructure for Systems Biology Europe (ISBE)*

*alexeykolodkin@gmail.com*

**Key words:** systems biology, dynamic modelling, oxidative stress, reactive oxygen species, aging, Parkinson's disease

*Motivation and Aim:* How the network around ROS (Reactive Oxygen Species) protects against oxidative stress and Parkinson's disease (PD), and how processes at the minutes time scale cause disease and aging after decades, remains enigmatic. Challenging whether the ROS network is as complex as it seems, we built a fairly comprehensive version thereof which we disentangled into a hierarchy of simpler subnetworks each delivering one type of robustness.

*Methods and Algorithms:* Our comprehensive model has 63 substances, 61 reactions, and more than 100 parameter values. The basis for substances and for the processes between them is the networks of ROS management known from the scientific literature. 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](http://www.copasi.org)), which is another Systems Biology Markup Language-compliant programme, 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 [1].

*Results:* Our model exhibited a relatively sudden breakdown, after some 80 years of virtually steady performance: it predicted aging. PD related conditions such as lack of DJ-1 protein or increased  $\alpha$ -synuclein accelerated the collapse, whilst antioxidants or caffeine retarded it. Introducing a new concept (aging-time-control coefficient), we found that as many as 25 out of 57 molecular processes controlling aging.

*Conclusion:* We have built a partly validated dynamic model that is able to (i) address aging and PD as results of networked molecular processes going awry, (ii) elucidate the corresponding time warp, and (iii) enable analyses towards individualized medicinal and nutritional therapies [2].

### References

1. Kolodkin A., Westerhoff H.V., Sharma R.P. ROS networks: designs, aging, Parkinson's disease and precision therapies. *FAIRDOMHub*. 2020. doi: 10.15490/fairdomhub.1.investigation.399.3.
2. Kolodkin A.N., Sharma R.P., 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.

## Comprehensive biomarkers of accelerated aging and mortality risk in end-stage renal disease

Kondakova E.<sup>1\*</sup>, Yusipov I.<sup>2</sup>, Lobanova N.<sup>1</sup>, Ivanchenko M.<sup>2</sup>, Vedunova M.<sup>1</sup>

<sup>1</sup> Institute of Biology and Biomedicine, Lobachevsky University, Nizhny Novgorod, Russia

<sup>2</sup> Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky University, Nizhny Novgorod, Russia

\* elen\_kondakova@list.ru

**Key words:** chronic kidney disease, accelerated aging, inflammaging, epigenetics, biochemistry

*Motivation and Aim:* Accelerated aging is a process associated with the accumulation of harmful changes in the body and an increased risk of disease and death. Despite advances in the treatment of chronic kidney disease (CKD) and optimization of the hemodialysis process in end-stage renal disease (ESRD), morbidity and mortality in this group of people remain constantly high [1]. Identification of markers of accelerated aging and mortality risk will allow finding possible interventions to increase the duration and improve the quality of life of patients.

*Methods and Algorithms:* The study participants were men and women aged 24 to 89 years. Of the 420 participants in the study, two groups were selected (patients with ESRD and controls) with the maximum match in age and gender. An enzyme-linked immunosorbent assay (ELISA) and multiplex analysis of cytokines to determine the concentration of FGF21, GDF15 and CXCL9 was performed. DNA methylation quantification were carried out using Illumina Infinium Methylation EPIC BeadChip. Epigenetic age were determined by methylome analysis using four widely models: DNAmAgeHannum, DNAmAge, DNAmPhenoAge and DNAmGrimAge. To develop a model that can predict the chronological age using three biomarkers (FGF21, GDF15, CXCL9), was applied the multiple linear regression model. The Elastic Net regression model (sklearn package in Python), where chronological age was regressed on three biomarkers only in the group of controls were applied. All three biomarkers (FGF21, GDF15, CXCL9) were included in the resulting model with optimal parameters.

*Results:* We demonstrated that ESRD is associated with the significant acceleration of biological age compared with the control group [2]. There was a statistically significant ( $p < 0,001$ ) increase in the content of all biomarkers (FGF21 GDF15, CXCL9) presented in the ESRD group compared to the controls. Moreover, it is accompanied by the dramatic increase in the variance of concentration level of these biomarkers in the ESRD group, irrespective of the etiology of CKD. We analyzed patients in terms of one-year survival and dividing the group of patients according to the disease outcome (survived/died) showed a significant ( $p < 0,05$ ) increase in the FGF21 content in the ESRD group with unfavorable one-year survival prognoses. We developed an estimator of biological age based on the blood concentration of three biomarkers: Age-Estimation Clock. Chronological age was regressed on three biomarkers (FGF21, GDF15, CXCL9) based on the group of controls. When analyzing the one-year survival prognosis, a significant increase in Age-Estimation Clock was found in the group of deceased patients.

*Conclusion:* Our proposed regression model for biological age determination considers the joint contribution of FGF21, GDF15, CXCL9 to the aging processes and allows determining the age of the subjects in the control group with sufficiently high accuracy. The results demonstrating a dramatic increase in the Age-Estimation Clock output in ESRD, up to values significantly exceeding human lifespan that confirmed by finding a statistically significant increase in the Age-Estimation Clock in the group of deceased patients. These findings, together with the evidence on the high prognostic value of FGF21 as a risk of mortality factor in patients with ESRD, opens up the possibility of further improvement of the model's predictive value and the risk group identification.

*Acknowledgements:* The study is supported by a Nizhny Novgorod region grant in the field of science, technology and engineering in 2022.

### *References*

1. Furman D. et al. Chronic inflammation in the etiology of disease across the life span. *Nat Med.* 2019;25:1822-1832.
2. Yusipov I., Kondakova E., Kalyakulina A., Krivonosov M., Lobanova N., Bacalini M.G., Franceschi C., Vedunova M., Ivanchenko M. Accelerated epigenetic aging and inflammatory/immunological profile (ipAGE) in patients with chronic kidney disease. *GeroScience.* 2022;44(2):817-834. doi: 10.1007/s11357-022-00540-4.

# Morphological feature of the postnatal blood-brain barrier formation as one of predictors of AD-like pathology development in rats

Kozlova T.A.\*, Tyumentsev M.A., Rudnitskaya E.A., Stefanova N.A.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\*kozlovatanya21@gmail.com

**Key words:** blood-brain barrier, postnatal brain development, tight junctions, OXYS rats

*Motivation and Aim:* Alzheimer's disease (AD) is destructive neurodegenerative disorder which development is conjugated by dysfunction of the blood-brain barrier (BBB) and decline of the cerebral blood flow. Preconditions of these alterations may originate at early age during BBB formation; nonetheless a probable linkage among altered BBB formation and its dysfunction later in life is studied insufficiently. Difficulties in exploration of this relationship are associated with a deficiency of suitable biological models of the most common sporadic form of AD. Senescence-accelerated OXYS rats are esteemed as an adequate model of the most common sporadic form of AD due to development of all key signs of the disease without mutations in App, Psen1 and Psen2 genes.

*Methods and Algorithms:* Here we investigated the formation of BBB in the hippocampus and frontal cortex of OXYS and Wistar (control) rats during early postnatal development. We studied the number and length of tight junctions (TJ) between endothelial cells together with the distance between TJ using electron microscopy and levels of occludin and claudin 5 as the main TJ proteins by Western blot analysis. Moreover, we evaluated the size and shape of capillary lumens as well as the integrity of a basement membrane.

*Results:* We found that the level of occludin did not significantly change with age in the hippocampus of OXYS and Wistar rats; however the occluding level naturally increased from birth to the end of the second postnatal week (postnatal day 14) in the frontal cortex of rats of both strains. Besides, we found a trend of decreased occludin level in the cortex of OXYS rats ( $p = 0.067$ ). Additionally we found that ultrastructure of capillaries was altered in the hippocampus of OXYS rats at early postnatal age.

*Conclusion:* Alterations of the BBB formation may result in cerebrovascular dysfunction observed previously in OXYS rats even at the early stage of AD-like pathology which therefore may contribute to the development and progression of neurodegeneration late in life.

*Acknowledgements:* This work was supported by Russian Science Foundation (grant No. 19-15-00044).

## Autophagy activation in 3D-spheroid leads to the mesenchymal stem cells rejuvenation

Krasnova O.<sup>1,2\*</sup>, Kulakova K.<sup>1,2</sup>, Bystrova O.<sup>1</sup>, Neganova I.<sup>1</sup>

<sup>1</sup> *Institute of Cytology, RAS, St. Petersburg, Russia*

<sup>2</sup> *Saint-Petersburg State University, St. Petersburg, Russia*

\* *olgakrasnova12@yandex.ru*

**Key words:** mesenchymal stem cells, 3D-spheroids, senescence, autophagy

Mesenchymal stem cells (MSCs) are known as promising candidates in cell and tissue therapy. MSCs transplantation is a highly developing field of the regenerative medicine and to date there are more than 1300 clinical trials concerning chronic and acute diseases treatment using MSCs. However, senescence limits the widespread implementation of this approach, as MSCs drop their potential and therapeutic properties with the passages. 3D-spheroids are considered as strategy to overcome senescence. Indeed, 2D-monolayer cell culture loses the main senescence features such as  $\beta$ -galactosidase-positive staining, cell cycle arrest, abnormally enlarged and flat morphology after cultivation in 3D-conditions. Importantly, 2D cell culture drastically differs from 3D-spheroid – the latter possesses enhanced stemness and differentiation potentials, cell size and cytoplasm-to-nucleus ratio decrease through relaxation of cytoskeleton tension. Among the numerous changes is autophagy gain. Autophagy is considered as ambiguous process during senescence. Pro- and anti-senescence effect of the autophagy depends on cell type and microenvironment. Electron microscopy reveals increased autophagosome formation alongside with high expression of autophagy-related genes such as *TFEB*, *ATG5*, *ATG12*, *ATG16L*, *GABARAPL* and *LC3* in 3D-spheroids after 48hrs in culture. Furthermore, electron and immunofluorescent microscopy demonstrate localization of the mTOR at the nucleus of the cells in 3D-spheroids – it implies on autophagy's activation. Autophagy inhibitor bafilomycin and chloroquine disturb spheroid formation and rejuvenation. Our data suggest autophagy's activation in 3D cell culture as pivotal process to overcome senescence.

*Acknowledgements:* This research was funded by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1075, signed 28.09.2021)

# Longevity and rejuvenation effects of cell reprogramming are decoupled from the loss of somatic identity

Kriukov D.<sup>1\*</sup>, Khrameeva E.<sup>1</sup>, Dmitriev S.<sup>2</sup>, Gladyshev V.<sup>3</sup>, Tyshkovskiy A.<sup>2,3</sup>

<sup>1</sup>Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup>Belozersky Institute of Physico-Chemical Biology, Moscow, Russia

<sup>3</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

\* dmitrii.kriukov@skoltech.ru

**Key words:** cell reprogramming, aging, meta-analysis, rejuvenation, lifespan extension, longevity interventions, gene expression, transcriptomic clocks, aging clocks, time-course data

*Motivation and Aim:* Transient somatic cell reprogramming has been proposed as a perspective rejuvenation intervention [1]. However, its association with the mechanisms of aging and longevity at molecular level remains unclear. Here we conducted a meta-analysis of time-series gene expression data and identified robust transcriptomic signature of reprogramming of mouse and human cells.

*Methods and Algorithms:* We gathered 41 gene expression datasets of time-course cell reprogramming from mouse and human cells. We used meta-analysis technique (mixed effects model) to aggregate multiple datasets into one reprogramming signatures. By comparing them to the signatures of aging and lifespan extending interventions, we revealed genes and functional processes associated with rejuvenation and longevity effects of reprogramming. We also trained multi-tissue transcriptomic aging clocks developed for humans and mice to indicate which particular combination of reprogramming factors provides rejuvenation effects.

*Results:* We observed significant co-regulation of genes perturbed by reprogramming and established lifespan-extending interventions, including those involved in DNA repair and inflammation. On the other hand, age-related gene expression changes were reversed during reprogramming, as confirmed by multi-tissue transcriptomic aging clocks. Importantly, the longevity and rejuvenation effects induced by reprogramming were mainly independent from the changes associated with the gain of pluripotency.

*Conclusion:* Overall, this work reveals specific molecular mechanisms associated with reprogramming induced rejuvenation at gene expression level and provides instruments for discovery of new geroprotective interventions that induce rejuvenation effect of reprogramming without posing the risk of pluripotency and neoplasia.

## References

1. Simpson D.J., Olova N.N., Chandra T. Cellular reprogramming and epigenetic rejuvenation. *Clin Epigenetics*. 2021;13(1):170. doi: 10.1186/s13148-021-01158-7.

## **Aging is a sequential multisystem syndrome, preventable and reversible under certain conditions**

Krut'ko V.<sup>1,2</sup>, Dontsov V.<sup>1</sup>, Khalyavkin A.<sup>1,3\*</sup>

<sup>1</sup> *Federal Research Center "Computer Science and Control", RAS, Moscow, Russia*

<sup>2</sup> *Sechenov First Moscow State Medical University, Moscow, Russia*

<sup>3</sup> *Emanuel Institute of Biochemical Physics, RAS, Moscow, Russia*

\* *antisenesc@mail.ru*

**Key words:** aging plasticity, systems approach, multi-system syndrome

*Motivation and Aim:* It is well known that aging is a natural biological process, as inevitable as sleep is. However, we know now that the "life without sleep" is possible [1] as well as the life without aging [2, 3]. We also know about the partial reversibility of the signs of aging [4] and about the so-called negative aging [5]. In individuals of species subject to negative aging, the risk of death does not increase with age, but decreases [5]. We have our own view on the origin of aging and its control, and especially on the explanation of why we age, despite the fact that we have potentially ageless somatic stem cells that can even reverse aging [6]. Based on this view, we consider aging as a sequential multisystem syndrome.

*Methods and Algorithms:* We used the critical analysis of the array of published experimental findings and our particular interpretation of these findings in order to unite at the one conception many separate and various data obtained from the molecular level up to the level of population.

*Results:* We review the progression of aging as a sequential development of multiple syndromes analogous to other diseases. This generalized approach may allow practicing physicians to consider the signs of aging as manifestations of a poly-syndrome disease and facilitate prevention, diagnosis and treatment of common aging-related dysfunctions.

*Conclusion:* Crucially, considering natural aging as an ordinary poly-syndrome disease makes it possible to apply timely prevention and correction for deteriorative aging using officially authorized drugs that are routinely available to practicing physicians.

*Acknowledgements:* The study is supported by the Ministry of Education and Science of the Russian Federation (RFMEFI60715X0123).

### *References*

1. Koval'zon V.M., Panchin Yu.V. Life without sleep. In: Proceedings VIII International Forum "SLEEP-2021". March 18–20, 2021. Moscow, 2021:10-11. (in Russian).
2. Martinez D.E. Mortality patterns suggest lack of senescence in hydra. *Exp Gerontol.* 1998;33(3):217-225.
3. Ruby J.G. et al. Naked mole-rat mortality rates defy gompertzian laws by not increasing with age. *eLife.* 2018;7:e311157.
4. Rando T.A., Chang H.Y. Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell.* 2012;148(1-2):46-57.
5. Vaupel J.W. et al. The case for negative senescence. *Theor Popul Biol.* 2004;65(4):339-351.
6. Khalyavkin A.V., Krut'ko V.N. How regularities of mortality statistics explain why we age despite having potentially ageless somatic stem cells. *Biogerontology.* 2018;19(1):101-108.



## Geroprotective properties of *Vaccinium uliginosum* L. and *Empetrum nigrum* ssp. *hermaphroditum* L. on the *Drosophila melanogaster* model

Kukuman D.\*, Golubev D., Zemskaya N., Shaposhnikov M., Moskalev A.

*Institute of Biology of Komi Scientific Centre of the Ural Branch of the RAS, Syktyvkar, Komi Republic, Russia*

\* [kukuman@ib.komisc.ru](mailto:kukuman@ib.komisc.ru)

**Key words:** *Vaccinium uliginosum*, *Empetrum nigrum* ssp. *hermaphroditum*, berry extract, geroprotector, lifespan, anthocyanins

**Motivation and Aim:** The fruits of *Vaccinium uliginosum* L. and *Empetrum hermaphroditum* L. are excellent sources of anthocyanins [1, 2], which have a wide range of health benefits [3, 4]. The purpose of this work was to study the potential geroprotective properties and mechanisms of action of plant extracts of local flora of Komi Republic on the *Drosophila melanogaster* model. We have estimated the effect of *Vaccinium uliginosum* L. and *Empetrum hermaphroditum* L. extracts on the lifespan, stress resistance, and locomotor activity of *Drosophila*.

**Methods and Algorithms:** Blueberries and crowberries were collected in the Pechora region near the Kozhva River in 2019. Ethanol solutions of blueberry and crowberry extracts at concentrations of 0.01, 0.1, 1, 2.5, and 5 mg/mL were added to the surface of the nutrient corn medium at the volume of 30  $\mu$ L. As a control, 30  $\mu$ L of 96 % ethanol was applied to the medium. We studied the effect of extracts on lifespan (median and age of 90 % mortality), locomotor activity, and the resistance to adverse environmental factors (oxidative stress, hyperthermia, starvation) according to the method described in this article [5].

**Results:** Blueberry extract at the concentration of 1 mg/mL increased the median lifespan of females by 5 % ( $p < 0.01$ , Gehan-Breslow-Wilcoxon test), and that of males by 3 % ( $p < 0.0001$ ). The maximum lifespan of females was increased when exposed to the extract at concentrations of 1, 2.5, and 5 mg/mL by 1 % ( $p < 0.05$ , Wang-Allisson test), and that of males by 4–6 % ( $p < 0.0001$ ). However, blueberry extract at a concentration of 0.01 mg/mL decreased in maximum lifespan in females by 3 % ( $p < 0.0001$ ), while in males, concentrations of 2.5 and 5 mg/mL decreased median lifespan by 2 % ( $p < 0.0001$ ) and 7 % ( $p < 0.01$ ) respectively.

The crowberry extract at a concentration of 2.5 mg/mL increased the median lifespan of females by 3 % ( $p < 0.05$ ). In males, an increase in median lifespan was observed by 6 % at a concentration of 0.01 mg/mL ( $p < 0.001$ ) and by 4 % at extract at the concentrations of 0.1 and 1 mg/mL ( $p < 0.05$ ). The extract at a concentration of 1 mg/mL led to a decrease in the maximum lifespan of females by 3 % ( $p < 0.0001$ ).

Treatment of *Drosophila* with blueberry extract at a concentration of 1 mg/mL causes a statistically significant increase in the resistance of females to hyperthermia by 21 % ( $p < 0.01$ , Fisher's test) and to starvation by 17 % ( $p > 0.05$ ). The blueberry extract increased the resistance of males to the action of the oxidative stress inducer paraquat by 9 % ( $p < 0.05$ ), but decreased the resistance to starvation by 38 % ( $p < 0.05$ ). Crowberry

extract at a concentration of 2.5 mg/mL increased the resistance of females to hyperthermia by 14 % ( $p < 0.05$ ) and oxidative stress by 33 % ( $p < 0.05$ ). Crowberry extract increased the resistance of males to hyperthermia ( $p < 0.05$ ) and oxidative stress ( $p < 0.01$ ) by 11 %.

There were no statistically significant effects of extracts on the locomotor activity of *Drosophila*, but blueberry extract reduced locomotor activity in females at the age of 8 weeks by 81 % ( $p < 0.01$ , Student's t-test), and in males at the age of 4 weeks by 51 % ( $p < 0.05$ ). Our results indicate that the studied extracts do not have a significant effect on the locomotor activity of *Drosophila*.

**Conclusion:** Ethanolic extracts of *Vaccinium uliginosum* L. and *Empetrum hermaphroditum* L. anthocyanins have positive effects on the lifespan of *Drosophila melanogaster*. The extract of *Vaccinium uliginosum* L. at a concentration of 1 mg/ml increased the resistance of males to oxidative stress, but reduced the resistance to starvation, without affecting the resistance to hyperthermia. In females, dietary blueberry extract increased resistance to starvation and hyperthermia, while the extract had no effect on resistance to the action of the prooxidant paraquat. *Empetrum hermaphroditum* L. extract at a concentration of 2.5 mg/ml increased the resistance of *Drosophila* to oxidative stress and hyperthermia, but had no effect on resistance to starvation. *Vaccinium uliginosum* L. and *Empetrum hermaphroditum* L. extracts have also been shown to have no significant effect on locomotor activity.

**Acknowledgements:** The research was carried out within the framework of the state task on the topic "Genetic and functional studies of the effects of geroprotective interventions on the *Drosophila melanogaster* model" (No. 122040600022-1).

## References

1. Ma L., Sun Z., Zeng Y., Luo M., Yang J. Molecular mechanism and health role of functional ingredients in blueberry for chronic disease in human beings. *Int J Mol Sci.* 2018;19(9):2785.
2. Jurikova T., Mlcek J., Skrovankova S. et al. Black crowberry (*Empetrum nigrum* L.) flavonoids and their health promoting activity. *Molecules.* 2016;21(12):1685.
3. Khoo H.E., Azlan A., Tang S.T., Lim S.M. Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr Res.* 2017;61(1):1361779.
4. Pallauf K., Duckstein N., Rimbach G. A literature review of flavonoids and lifespan in model organisms. *Proc Nutr Soc.* 2017;76(2):145-162.
5. Golubev D., Zemskaia N., Shevchenko O. et al. Honeysuckle extract (*Lonicera pallasii* L.) exerts antioxidant properties and extends the lifespan and healthspan of *Drosophila melanogaster*. *Biogerontology.* 2022;23(2):215-235.

## Role of Hsp70 chaperones in age-related changes in skeletal muscle proteome in *Drosophila melanogaster*

Kukushkina I.V.<sup>1,2\*</sup>, Zgodá V.G.<sup>3</sup>, Makhnovskii P.A.<sup>1</sup>, Popov D.V.<sup>1,2</sup>

<sup>1</sup> *Institute of Biomedical Problems, RAS, Moscow, Russia*

<sup>2</sup> *Lomonosov Moscow State University, Moscow, Russia*

<sup>3</sup> *V.N. Orekhovich Research Institute of Biomedical Chemistry, Moscow, Russia*

\* [vladimirova-bph@yandex.ru](mailto:vladimirova-bph@yandex.ru)

**Key words:** *Drosophila melanogaster*, Hsp70, mitochondria

**Motivation and Aim:** Aging is associated with changes in many cellular functions, including a decrease in mitochondria function in skeletal muscles. Chaperones regulate proteostasis, controlling protein refolding and nonspecific aggregation. We aimed to investigate the role of chaperones of the Hsp70 family in age-related changes in skeletal muscle proteome in *D. melanogaster*.

**Methods:** Males of the *D. melanogaster* Hsp70-(Df(3R)Hsp70A,Df(3R)Hsp70B), strain with knockout of 6 genes from the Hsp70 family, were used in the study, as well as males of the control parent strain *w1118*. General functional status of flies was assessed during the life by vertical climbing speed. Fly's legs were prepared to evaluate genotype- and age-related changes in the proteome of skeletal muscles using shotgun quantitative (with isobaric labeling) proteomic.

**Results:** There were no differences in survival between strains. Young Hsp70-flies show 2 times lower climbing speed compared to *w1118*. During aging only *w1118* showed a decrease in the climbing speed, while the climbing speed of Hsp70-flies remained at a constant (and substantially reduced) level compare to the control. We detected 725 highly abundant proteins in skeletal muscle: the content of 67 and 88 proteins was reduced and increased respectively, in old Hsp70-compared to the age-matched control.

**Conclusion:** Knockout of Hsp70 family genes does not affect lifespan, but reduces locomotor activity and down regulates proteins associates with carbohydrate metabolism and oxidation-reduction processes, in particular, with pentose phosphate pathway and cytochrome c oxidase.

**Acknowledgements:** The study was supported by the Russian Science Foundation (grant 21-15-00405).

## Effect of aging, chronic inflammation and long-term physical disability on human skeletal muscle transcriptome

Kurochkina N.S.<sup>1\*</sup>, Vigovskiy M.A.<sup>2</sup>, Vepkhvadze T.F.<sup>1,2</sup>, Makhnovskii P.A.<sup>1</sup>, Grigorieva O.A.<sup>2</sup>, Lednev E.M.<sup>1</sup>, Efimenko A.Yu.<sup>2</sup>, Popov D.V.<sup>1,2</sup>

<sup>1</sup> Institute of Biomedical Problems, RAS, Moscow, Russia

<sup>2</sup> Lomonosov Moscow State University, Moscow, Russia

\* [nadia\\_sk@mail.ru](mailto:nadia_sk@mail.ru)

**Key words:** skeletal muscle, RNA sequencing, transcriptome, aging

**Motivation and Aim:** At normal physical activity, skeletal muscle tissue plays a key role in the carbohydrate and fat metabolism of the organism and has a positive effect on the functioning of other tissues and organs. Aging is associated with complex changes, including decrease in physical activity, mass and functionality of skeletal muscles (insulin sensitivity, muscular endurance and strength), and with the development of various diseases, often leading to the chronic inflammation. The aim of our study was to investigate effect of aging, chronic inflammation and long-term physical disability on human skeletal muscle transcriptome.

**Methods and Algorithms:** Transcriptomic profile of skeletal muscle of elderly patients with primary knee/hip arthrosis (OP,  $n = 35$ , 72(66–83) years old, body mass index (BMI) 30(21–43) kg/m<sup>2</sup>) was compared with that of young healthy volunteers (YH,  $n = 10$ , 35(25–43) years old, BMI 22.5 (18.9–29.4) kg/m<sup>2</sup>). To identify the role of chronic inflammation and long-term physical disability in gene expression changes, YH was compared with a group of young patients with primary arthrosis (YP,  $n = 7$ , 39(26–45) years old, BMI 25.8(25.4–30.9) kg/m<sup>2</sup>). To evaluate the effect of age, the YP and OP groups were compared. Tissue samples from *m. vastus lateralis* were collected in patients before joint replacement surgery, in healthy people – using needle biopsy. Then single-end RNA sequencing (75 nucleotides, 66 million reads/sample) was performed. Differentially expressed genes were evaluated using DESeq2 with a cutoff criteria: fold change  $> 1.5$  and  $P_{adj} < 0.05$ .

**Results:** Physical score (Short form survey SF-12) did not differ between the groups of patients and was significantly lower than in YH. Inflammatory markers (blood leukocyte and neutrophil levels) were comparable between OP and YP, but significantly greater than in YH. In the OP group, compared with YH, metabolic disorders were observed: an increase in fasting blood glucose and insulin levels, as well as an increase in BMI.

In old people (comparison OP vs. YH) up-regulated genes (2529) were associated with inflammatory response, extracellular matrix reorganization and lipid metabolism; down-regulated genes (1007 genes) mainly encoded mitochondrial proteins and carbohydrate metabolism regulators. In young patients (comparison of YP vs. YH) up-regulated genes (2107) were associated with inflammation and biogenesis of the extracellular matrix, and down-regulated genes (558) mainly encoded mitochondrial proteins, and were associated with G protein signaling. These changes were comparable (but less pronounced) with those in OP vs. YH. Age-related changes in gene expression in patients (OP vs. YP) are less than in previous comparisons: a decrease (536) and an increase

(299) in gene expression are associated with the regulation of the inflammatory response, and down-regulated genes are associated with angiogenesis and G-proteins signaling.

*Conclusion:* In old patients with primary arthrosis, there are marked changes in the transcriptome of skeletal muscles (OP vs. YH). Most of these changes were found in young patients with arthrosis and long-term physical disability compared with young healthy people (YP vs. YH). This suggests that inflammation and decreased physical activity (induced by long-term physical disability) are key factors inducing age-dependent changes in skeletal muscle transcriptome.

*Acknowledgements:* The study was supported by the Russian Science Foundation (grant 21-15-00405, biomaterials from young donors, transcriptomic studies, bioinformatics) the State Assignment of Lomonosov MSU using the equipment provided within the Program of Development of Lomonosov MSU (biomaterials from patients with arthrosis, clinical and laboratory assessment).

# **Innate immunity as an executor of phenoptosis – the programmed death of individual organisms for the sake of the whole population**

Lyamzaev K.G.<sup>1,2\*</sup>, Mulkidjanian A.Y.<sup>1,3</sup>, Chernyak B.V.<sup>1</sup>

<sup>1</sup> *A.N. Belozersky Institute, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *The “Russian Clinical Research Center for Gerontology” of the Ministry of Healthcare of the Russian Federation, Pirogov Russian National Research Medical University, Moscow, Russia*

<sup>3</sup> *Department of Physics, Osnabrueck University, Osnabrueck, Germany*

\* *Lyamzaev@gmail.com*

*Abstract:* In humans, over-activation of innate immunity in response to viral or bacterial infections often causes severe illness and death. Furthermore, similar mechanisms related to innate immunity can cause pathogenesis and death in sepsis, massive trauma (including surgery and burns), ischemia/reperfusion, and some toxic lesions. Based on these observations, we hypothesize that such severe outcomes may be manifestations of a controlled suicidal strategy protecting the entire population from the spread of pathogens and from dangerous pathologies rather than an aberrant hyperstimulation of defense responses. We argue that innate immunity may be involved in the implementation of phenoptosis, an altruistic programmed death of an organism aimed at increasing the well-being of the whole community. We discuss possible ways to suppress this atavistic program by interfering with innate immunity and suggest that combating phenoptosis program should be a major goal of future medicine.

## Vascular wall and molecular markers of age in an ageing population

Malyutina S. \*, Ryabikov A., Maximov V.

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

\* [smalyutina@hotmail.com](mailto:smalyutina@hotmail.com)

**Key words:** Intima-media, atherosclerosis, DNA methylation, epigenetic age, length of telomeres, ageing, population

*Background and Aim:* Atherosclerotic diseases is age-dependent phenotype but the evidence of the relationship between molecular markers of age and atherosclerosis is inconsistent. DNA methylation (DNAm) is known to change throughout the life course. The measure of ‘epigenetic age’ (EA) based on DNAm and leucocyte telomere length (LTL) are considered as ageing biomarker.

We investigated the relationship between EA, LTL and carotid intima-media thickness (CIMT) and plaques in serial measurements in a population sample 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 re-examined in 2015/18 (wave-3, 55–84,  $n = 3898$ ). Carotid ultrasound and LTL measurement were conducted in a subsample at baseline and in wave 3 (773 at wave-1, 995 at wave-3, 526 have repeated measurements). We measured mean-mean CIMT on the far wall of both common carotid arteries and identified carotid plaques (CP) with high-resolution ultrasound. LTL was measured by real-time quantitative PCR-based method. The assessment of DNAm was conducted on Illumina EPIC850k Beadchip in a random subsample (94 subjects) and EA was calculated by Horvath.

*Results:* The mean follow-up period of serial measurements for carotid phenotypes was 7.5 years and 11.9 years for LTL. The mean CIMT was 0.75 (SD 0.18) mm at baseline and 0.91 (0.18) at wave-3 ( $p < 0.001$ ); the average thickening rate was 0.03 (95 %CI 0.029–0.032) mm/year. The mean relative LTL was 1.35 (0.40) at baseline and 0.88 (SD 0.24) at wave-3 ( $p < 0.001$ ), with average shortening rate of  $-0.043$  (95 %CI  $-0.050$ ;  $-0.040$ ) per year. The mean baseline EA was correlated with chronological age across the whole age range ( $r = 0.50$ ,  $p < 0.001$ ) and mean EA was higher compared to chronological age (65.1 years, SD 5.38 and 63.4 years, SD 3.02).

In cross-sectional analysis, at baseline there was no association between CIMT and LTL but in wave-3 LTL negatively correlated with CIMT in men ( $r = -0.130$ ;  $p = 0.016$ ).

In prospective analyses, among men baseline LTL inversely correlated with prospective CIMT, its increment during follow-up and thickening rate per year ( $p = 0.015$ – $0.026$ ). There was no association between baseline EA and prospective CIMT or CIMT thickening rate during follow-up.

For further analyses, a “progression of CIMT” was defined as increment larger than the mean increment during follow-up and contrasted with “no progression”. Baseline LTL associated with CIMT progression ( $p = 0.028$ ). In wave-3, the CIMT progression was more frequent in the 2nd and 3rd vs. 4th quartile of LTL (ordered smallest LTL quartile

to largest) in multivariable models ( $p = 0.032; 0.046$ ). In addition, LTL in wave-3 was shorter in men with CP than in those without CP ( $p = 0.035$ , adjusted for age and smoking). Baseline EA was also associated with CIMT progression: the CIMT progression was more frequent in the 4th quartile vs 1st quartile of EA in multivariable logistic regression model ( $p = 0.041$ ).

*Conclusion:* In a population sample of middle and elderly age, the age-related changes in carotid atherosclerosis and LTL were both of expected direction. At baseline, we did not detect cross-sectional associations between CIMT and LTL or EA. However, baseline LTL inversely correlated with prospective CIMT. Prospectively, among men shorter prospective LTL was independently associated with CIMT progression and with carotid plaques. Similarly, baseline EA was associated with progression of CIMT independently of other factors. The findings allow suggesting that the association between LTL shortening or epigenetic age and subclinical atherosclerosis might be confined to advanced vascular lesion.

*Acknowledgements:* The study is supported by Russian Science Foundation (20-15-00371).



## Relationship between basal metabolic rate, muscle strength and aging

Morgunova G.V. \*, Khokhlov A.N.

*Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russia*

\* [morgunova@mail.bio.msu.ru](mailto:morgunova@mail.bio.msu.ru)

**Key words:** BMR, aging, bioenergetics, metabolism, skeletal muscles

Basal Metabolic Rate (BMR), along with mass, is one of the main characteristics of a species in comparative physiology. Despite the fact that this indicator is periodically criticized, it remains an important measure of the energy expenditure of the body. Of particular interest is the relationship of energy metabolism with aging. Both interspecific and intraspecific comparative studies of the relationship of BMR with life span and other indicators are being conducted. The results so far do not look unambiguous; for example, it has been shown that BMR does not correlate with longevity in eutherians or birds, if a correction for mass is introduced [1]. At the same time, it is clear that passerine birds with high BMR live significantly longer than mammals of the same size.

The body undergoes significant changes during development and aging, so if we look at changes in metabolism in individuals during ontogeny, we will find that BMR probably also changes during life. Previously, it was thought that BMR in humans falls all the time with age, however, in the latest, more detailed and covering a large sample of people study, it was found that the decrease in metabolic rate occurs only after 60 years [2]. Starting from the age of 60, age-related pathologies begin to manifest themselves more clearly, and this probably does not coincide with a decrease in BMR.

Many researchers attribute higher BMR to higher production of reactive oxygen species. In this regard, in long-lived bird species, many researchers are looking for distinctive characteristics that help them in the fight against oxidative stress. However, it has not yet been possible to establish any features – for example, no special composition of fatty acids has been found in cell membranes in long-living birds compared to other ones [3]. In our opinion, it is necessary to look for the reasons for longevity not only in the ability to withstand oxidative stress. The fall in metabolism in people of older age groups is probably due to a loss in the mass and strength of muscle fibers, which is expressed in their atrophy and sarcopenia, as well as in violation of anabolic processes in muscles. It is no coincidence that healthy people after 60 and 70 years of age show on average about 20–40 % lower results in isometric strength tests than young people, and in very old people the decrease is even more noticeable [4]. As is known, skeletal muscles are the main consumer of insulin and, accordingly, glucose; it is actively involved in the use of energy received from food (postprandial blood glucose levels). In this regard, skeletal muscles largely determine the development of insulin resistance. Passerines have a special flight style that requires more energy, so their BMR is higher than in non-passerines [5]. This probably also contributes to the fact that passerine birds have a well-developed glucose sensitivity support system, an extensive network of mitochondria, and many other adaptations for high BMR. In our opinion, it makes sense for researchers dealing with the question of the relationship between metabolism and longevity to take

into account the anatomy, physiology, and biochemistry of muscles in both comparative and longitudinal studies.

*Acknowledgements:* This study was performed under the state assignment of Moscow State University, project No. 121032300215-6.

### *References*

1. Magalhães J.P., Costa J., Church G.M. An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J Gerontol A Biol Sci Med Sci.* 2007;62(2):149-160.
2. Pontzer H. et al. Daily energy expenditure through the human life course. *Science.* 2021;373(6556):808-812.
3. Montgomery M.K., Hulbert A.J., Buttemer W.A. Metabolic rate and membrane fatty acid composition in birds: a comparison between long-living parrots and short-living fowl. *J Comp Physiol B.* 2012;182(1):127-137.
4. Vandervoort A.A. Aging of the human neuromuscular system. *Muscle Nerve.* 2002;25(1):17-25.
5. Gavrilov V.M. Energy expenditures for flight, aerodynamic quality, and colonization of forest habitats by birds. *Biol Bull.* 2011;38(8):779-788.

## **MEK1/2-ERK 1/2–dependent alphaB-crystallin phosphorylation in retina: a focus on aging and age-related macular degeneration**

Muraleva N.A.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

*Myraleva@bionet.nsc.ru*

**Key words:** aging, MEK1/2-ERK1/2 signaling pathway, retina, OXYS rats

*Abstract:* Accumulation of protein aggregates and increase of intracellular damage are universal hallmarks of aging and accompany the development of some age-related diseases including Age-related macular degeneration (AMD). Alpha-B-Crystallin (CryaB) as the molecular chaperone contributes to maintenance of proteostasis by prevention of protein aggregation (e.g. amyloid beta) and enables their correct refolding. CryaB activity is regulated by MEK1/2-ERK 1/2 signaling pathway through its phosphorylation at the Ser45 position. Nevertheless, the link between changes in MEK1/2-ERK1/2–dependent CryaB phosphorylation with age and the development of AMD remains unclear. Here, we examined MEK1/2-ERK1/2-dependent phosphorylation of p45-CryaB in the retina of Wistar rats with normal aging and senescence-accelerated OXYS rats at the different stages of the development of AMD-like pathology, including the presymptomatic stage.

According to transcriptome analysis, the maximum number of differentially expressed genes (DEG) involved in MEK1/2-ERK1/2 pathway in retina of Wistar rats was found at the age from 20 days to 3 months. The activity of the MEK1/2-ERK1/2 pathway at the mRNA level decreased by the age of 18 months. The number of DEG at 18 months was decreased. The western blotting revealed that the level of MEK1/2 and ERK1/2 proteins did not change with age, while the level of MEK1/2, ERK1/2, and CryaB phosphorylation increased by the age of 18 months. Similar but more significant changes accompanied the development of signs of AMD-like pathology in OXYS rats. The activation of MEK1/2-ERK1/2 pathway detected at the manifestation of retinopathy in OXYS rats (3 months old) and it accompanied the elevated of p45CryaB protein. It remained high during the active progression of retinopathy in OXYS rats against the background of the accumulation beta amyloid 1-42 in the retina. Thus, the activation of MEK1/2-ERK1/2-dependent phosphorylation CryaB occurs during normal aging. Manifestation and progression of the AMD-like pathology in OXYS rats occurs against the background of activation of MEK1/2-ERK1/2 pathway and ERK1/2-dependent CryaB phosphorylation.

*Acknowledgements:* This study was supported by Russian Scientific Foundation Grant (No. 22-25-00224).

## Current views on the mechanisms of aging

Panov A.V.<sup>1\*</sup>, Semenova E.V.<sup>2</sup>, Dikalov S.I.<sup>3</sup>

<sup>1</sup> *Federal Scientific Center for Family Health and Human Reproduction Problems, Irkutsk, Russia*

<sup>2</sup> *Ogarev Mordovia State University, Saransk, Russia*

<sup>3</sup> *Division of Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, USA*

\* *Alexander.panov55@gmail.com*

**Key words:** aging, metabolic syndrome, mitochondria, oxidative stress

*Motivation and Aim:* We live at a time when many earlier paradigms are denied and replaced with new ones, which is particularly true for the studies of mitochondrial dysfunctions in the mechanisms of aging and associated diseases in humans. Recently it was discovered that the proteinaceous "shield" of nucleoids protects mtDNA from oxidative damages and that there is no actual proof for the free radical direct effects on mtDNA [reviewed in 1]. Moreover, researchers finally realized that practically all studied oxygen radicals, such as superoxide radical ( $O_2^{\bullet}$ ), nitric oxide ( $^{\bullet}NO$ ), peroxyxynitrite ( $^{\bullet}ONO_2$ ), and even hydroxyl radical ( $^{\bullet}OH$ ), are either chemically not active enough, or live too short, or formed only locally. Thus neither of them can be responsible for the slow and unavoidable process of gradual accumulation of damages and dysfunctions, which we name "aging." Studies of changes in genomics, proteomics, and other "omics" were of little help in understanding mechanisms of aging and pathogenesis of diseases mainly because these approaches have not been accompanied by the simultaneous research on changes in the related functions. Aging is, with few exceptions, a fundamental phenomenon for all living organisms on Earth. In 2013, Barja summarized the numerous studies on life longevities of various species: "Only two parameters currently correlate with species longevity in the right sense (reversely): the mitochondrial rate of reactive oxygen species (ROS) production and the degree of fatty acid unsaturation of tissue membranes" [2]. At first glance, a strange conclusion. Here, we will briefly outline the features of new paradigms related to oxidative stress, mechanisms of aging, and age-related diseases to confirm the correctness of Barja's conclusion.

*Gender variations in the rates of aging.* One of the critical features of aging among mammals is that females age slower and live longer [3]. Studies on human energy metabolism have shown significant gender differences, particularly in the consumption of fatty acids [reviewed in 1, 4]. Olivetty et al. (1995) found that men's hearts undergo a linear loss of approximately 64 million cells from the age of 17 to 95, whereas, in women, the structural properties of the heart remained unchanged [5]. In our experiments, we have shown that long-chain fatty acids oxidize at a high rate only in the presence of supporting substrates, which are amino acids and mitochondrial metabolites pyruvate and succinate [4, 6]. Studies on the gender differences in energy metabolism have shown that women consume more fatty acids (FAs) than men, whereas men, in addition to FAs, consume significant amounts of pyruvate and amino acids [7, 8]. We have suggested that before menopause, women oxidize FAs less efficiently and thus produce less ROS than men [1, 4]. However, after menopause, with the development of

the metabolic syndrome (MetS), women's energy metabolism becomes similar to men's, and the rate of aging accelerates [9].

*Metabolic syndrome is the consequence of the transition to the postreproductive stage of human ontogenesis.* From this point of view [1, 4], most external and metabolic features of MetS reflect the fact that this stage of ontogenesis is governed by the genes of our distant ancestors, at least for the people living in the Northern Hemisphere. At MetS, fatty acids become the primary energy source for all functions, which dramatically increases the production of superoxide radicals and their protonated form perhydroxyl radical ( $\text{HO}_2^\bullet$ ), which is the inducer of the isoprostane pathway of lipid peroxidation (IPLP) in mitochondria [4,10].  $\text{HO}_2^\bullet$  induces IPLP of polyunsaturated fatty acids with the formation of hundreds of products, some of which resemble prostaglandins; others are highly reactive and form stable adducts with cardiolipin, phosphatidylethanolamine, or proteins [Reviewed in 10]. As a result, the rate of aging during MetS becomes accelerated. Most complications observed in patients with MetS originate from the inconsistencies between the metabolic phenotype acquired after the transition to the postreproductive stage and excessive consumption of food rich in carbohydrates and a sedentary lifestyle [1, 4, 10].

### References

1. Panov A., Darenskaya M.A., Dikalov S.I., Kolesnikov S.I. Chapter 3. Metabolic syndrome as the first stage of eldership; the beginning of real aging. In: Amornyotin S. (Ed.). Update in Geriatrics. InTech, 2021:1-31. doi: 10.5772/intechopen.95464.
2. Barja G. Updating the mitochondrial free radical theory of aging: an integrated view, key aspects, and confounding concepts. *Antioxid Redox Signal*. 2013;19(12):1420-1445. doi: 10.1089/ars.2012.5148.
3. Lemaitre J.F. et al. Sex differences in adult lifespan and aging rates of mortality across wild mammals. *PNAS USA*. 2020;117(15):8546-8553. doi: 10.1073/pnas.1911999117.
4. Panov A., Mayorov V.I., Dikalov S. Metabolic syndrome and beta-oxidation of long-chain fatty acids in the brain, heart, and kidney mitochondria. *Int J Mol Sci*. 2022;23(7):4047. doi: 10.3390/ijms23074047.
5. Olivetti G. et al. Gender differences and aging: effects on the human heart. *J Am Coll Cardiol*. 1995;26:1068-1079. doi: 10.1016/0735-1097(95)00282-288.
6. Panov A.V. Synergistic oxidation of fatty acids, glucose, and amino acids metabolites by isolated rat heart mitochondria. *EC Cardiology*. 2018:198-208.
7. Tarnopolsky M.A. Gender differences in substrate metabolism during endurance exercise. *Can J Appl Physiol*. 2000;25(4):312-327. doi: 10.1139/h00-024.
8. Lamont L.S., McCullough A.J., Kalhan S.C. Gender differences in the regulation of amino acid metabolism. *J Appl Physiol*. 2003;95:1259-1265. doi: 10.1152/jappphysiol.01028.2002.
9. Carr M.C. The emergence of the metabolic syndrome with menopause. *J Clin Endocrinol Metabol*. 2003;88(6):2404-2411. doi: 10.1210/jc.2003-030242.
10. Panov A. Perhydroxyl radical ( $\text{HO}_2^\bullet$ ) as inducer of the isoprostane lipid peroxidation in mitochondria. *Mol Biol*. 2018;52:295-305. doi: 10.1134/S0026 893318020097.

## Natural variability in the regulatory regions of longevity genes and ecological adaptation

Pasyukova E.G.<sup>1\*</sup>, Symonenko A.V.<sup>1</sup>, Tsybul'ko E.A.<sup>1</sup>, Rybina O.Y.<sup>1,2</sup>,  
Krementsova A.V.<sup>1,3</sup>, Alatorsev V.E.<sup>1</sup>, Roshina N.V.<sup>1,4</sup>

<sup>1</sup> *Institute of Molecular Genetics, National Research Centre "Kurchatov Institute", Moscow, Russia*

<sup>2</sup> *Moscow Pedagogical State University, Moscow, Russia*

<sup>3</sup> *Emmanuel Institute of Biochemical Physics, RAS, Moscow, Russia*

<sup>4</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

\* [egpas@rambler.ru](mailto:egpas@rambler.ru)

**Key words:** lifespan, transcription, natural populations, polymorphism, SNPs, *Drosophila melanogaster*

*Motivation and Aim:* The elucidation of the molecular mechanisms underlying the relationship between gene expression and phenotype is a fundamental scientific task that is crucial for understanding the basis of development, life, and aging of living organisms. Transcription, which is the most important step that determines the level of gene expression, depends on the interaction between various regulatory elements localized in noncoding DNA and the regulatory proteins that interact with them. In nature, regulatory elements are characterized by significant structural variability, which can affect the level of transcription and, ultimately, complex phenotypes and adaptation. The aim of our work is a comprehensive analysis of the functional role of natural variability in the regulatory regions of genes controlling longevity.

*Methods and Results:* The variability of the 5' and 3' regulatory regions of the *shuttlecraft* and *escargot* genes in flies collected in the Alexandrov and Valдай populations (Russia) was characterized: segregating polymorphisms in populations were identified and their frequencies were compared with those observed in populations living in different parts of the world. The association of the identified segregating polymorphisms with the lifespan of flies and environmental conditions of populations was analyzed. It was shown that lifespan is associated with the variability of at least two polymorphic loci located in the 5' regulatory region of the *stc* gene. The variability of the same polymorphic loci is associated with the air temperature in the habitats of the populations. It was shown that the transcription of the *stc* gene and the expression of the reporter gene in *Drosophila* cell culture depend on which nucleotide is located in each of the two polymorphic loci.

*Conclusion:* This work allowed us to identify functionally important sites of the gene regulatory region and characterize the functional significance of their natural variability for the control of transcription, phenotype, and ecological adaptation.

*Acknowledgements:* The study was supported by the Founder's state assignment (2021–2022), "Investigation of the molecular and genetic basis of variability in lifespan and aging rate in *Drosophila melanogaster*".

# Analysis of changes in microglia and cerebral endothelium in hippocampus of OXYS rats during progression of Alzheimer's disease-like pathology

Peunov D.A.<sup>1, 2\*</sup>, Tyumentsev M.A.<sup>1</sup>, Rudnitskaya E.A.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *d.peunov@g.nsu.ru*

**Key words:** Alzheimer's disease, microglia, cerebral endothelium, OXYS rats

*Motivation and Aim:* Sporadic Alzheimer's disease (AD) is the most spread neurodegenerative disorder that causes dementia in elder people. Among all brain structures the hippocampus is one that suffers most and undergoes the earliest pathological changes. There is growing evidence for the pivotal role of neuroinflammation and dysfunction of cerebral endothelium (CE) in AD. Among brain cells microglia–local phagocytes of the central nervous system (CNS)– are assumed to be the main driver of inflammation. In its turn inflammation staying unresolved for a long time may harm other cells including CE. In the case of AD this damage causes a lack in supplementation of brain with trophic substrates, worsening of amyloid-beta (A $\beta$ ) clearance as well as uncontrollable entrance of factors provoking inflammation. A feedback interaction between proinflammatory microglial activity and cerebral vasculature function appears to exist, and its reveal demands a presence of adequate AD models. Senescence-accelerated OXYS rats are considered as one of such models.

*Methods and Algorithms:* The work was performed using OXYS and Wistar rats (maternal strain used as a control). To reveal AD-associated changes we tested animals at the age of 20, 45 days; 3, 5, 12 and 18 months. The quantity and activation state of microglia were assessed by immunohistochemistry ( $n = 6$ ). Occludin level was measured by western-blot analysis ( $n = 6$ ). Structure of CE was imaged by electron microscopy. For the gene set obtained by intersection of differently expressed in OXYS rats (compared to Wistar rats) genes [1] with microglia– or CE–enriched genes [2] over representation analysis was conducted

*Results:* We showed that OXYS rats in comparison with Wistar rats had higher total and activated microglia density at all ages which for activated microglia was mostly pronounced at 45 days. Upon the results of transcriptome analysis at 20 days OXYS rats had altered expression level of 10 microglia-enriched genes. Among them the biological substrate oxidation pathway was presented (*Cyp4v3*↑, *Paox*↓, *Ptgs1*↓, FDR<0.05). At the 5 months of age expression levels of 4 genes were altered. Among them *Map3k14*↑ – the NF- $\kappa$ B signaling pathway related gene was present. Also no differences in expression of proinflammatory genes downregulated in OXYS rats earlier were detected. At the 18 months of age expression levels of 33 genes were altered. Among these genes the PI3K-AKT↓ (*Pik3cg*, *Spp1*, *Cdk6*), TNF↑ (*Map3k14*, *Cebpb*, *Junb*) signaling pathways and genes related to protein ubiquitination (*Nhlrc3*, *Ubc*), carbohydrates metabolism (*Hk2*, *Slc2a5*, *Mlxipl*) were altered. The occludin level in OXYS rats was lower than in Wistar rats at all ages, being significantly lower at 45 days and trending

downward at 4 months of age. That could be a sign for disruption of tight junctions and was also supported by electron microscopy studies. At 20 days of age we observed 8 CE enriched genes being differently expressed in OXYS rats. Among them genes related to lipid metabolism (*Sgms2*↓, *Hmgcs2*↑), transendothelial transport (*Cav2*↑) and reaction to immune activity (*Sod3*↑, *St3gal6*↑) were presented. At 5 months of age we detected 12 differently expressed genes. Among them genes related to the WNT(*Fzd6*↓) and NOTCH (*Dll4*↑, *Egfl8*↑) were presented. At 18 months of age OXYS rats had 43 differently expressed genes. Among them biological processes namely – blood circulation, negative regulation of cell development, negative regulation of stimuli response, adhesive contacts, regulation of adhesive contacts formation, biosynthesis of prostaglandins and metabolism of sphingolipids – were altered.

*Conclusion:* At the age prior to the manifestation of AD-like pathology (20-45 days) in OXYS rats presumably signs of CE disruption take place (*Cav2*↑). Also transcriptome analysis results indirectly point to the immunoactive milieu enriched with reactive oxygen species (*Sod3*↑, *St3gal6*↑). Concurrently according to the results of immunohistochemistry microglia appear to be slightly more active in OXYS rats at this age. However analysis of microglia-enriched genes showed diminished levels of some inflammation-associated genes (*Paox*↓, *Ptgs1*↓). Together these observations could indicate that activation of microglia is consequent to the disturbance of CE. Suppressed microglial pro-inflammatory response to CE impalement could point to their dysfunction or represent the recovery after preexisting inflammation. At the age when AD signs begin to manifest (3-5 months) we still observe signs of tight junctions disturbance in OXYS rats although signs of direct damage of CE are absent. At the same time the alterations in the growth-related signaling pathways could point to the suppression of angiogenesis. Also the *Fzd6* gene encodes the receptor that mainly activates canonical and planar cell polarity pathways which in turn oppose the WNT/ $Ca^{2+}$  pathway which may enhance inflammation. Hence suppression of the FZD6/WNT branch could be the factor promoting inflammation induction. Microglia phenotype at this age appears to be rather neutral or slightly pro-inflammatory according to the transcriptional analysis. At the age of active AD-like pathology progression and Aβ accumulation (18 months) we observe signs of Aβ metabolism disturbance both in microglia (*Nhlrc3*↓, *Ubc*↓) and CE (*Abcab1*). At the same time, CE appears to undergo endothelial-to-mesenchymal transformation and has weakened connection with the extracellular matrix. Microglia at this age seem to express a wider (compared to the age of 5 months) spectrum of proinflammatory genes at higher level. It's known that inflammation could promote all listed adverse alterations of CE. Ultimately in OXYS rats dysfunction of CE could be an early trigger of pathogenesis development but microglial activation at this time may exert cytoprotective effect while later microglial proinflammatory activity presumably promotes worsening of CE pathology that leads to progression of AD-like pathology.

*Acknowledgements:* The research was supported by RSF (No. 19-15-00044).

## References

1. Stefanova N.A., Ershov N.I., Kolosova N.G. Suppression of Alzheimer's Disease-Like Pathology Progression by Mitochondria-Targeted Antioxidant SkQ1: A Transcriptome Profiling Study. *Oxid Med Cell Longev.* 2019;2019:3984906. doi: 10.1155/2019/3984906.
2. Zhang Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci.* 2014;34(36)11929-11947.



## **Accumulation of damaged mitochondrial DNA in tissues of aged rats and the impact of dietary restriction on this process**

Plotnikov E.Y.<sup>1\*</sup>, Gureev A.P.<sup>2,3</sup>, Andrianova N.V.<sup>1</sup>, Pevzner I.B.<sup>1</sup>, Zorova L.D.<sup>1</sup>, Chernyshova E.V.<sup>2,3</sup>, Sadovnikova I.S.<sup>2,3</sup>, Chistyakov D.V.<sup>1</sup>, Popkov V.A.<sup>1</sup>, Silachev D.N.<sup>1</sup>, Zorov D.B.<sup>1</sup>, Popov V.N.<sup>2</sup>

<sup>1</sup> *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup> *Department of Genetics, Cytology and Bioengineering, Voronezh State University, Voronezh, Russia*

<sup>3</sup> *Laboratory of Metagenomics and Food Biotechnology, Voronezh State University of Engineering Technology, Voronezh, Russia*

\* [plotnikov@belozersky.msu.ru](mailto:plotnikov@belozersky.msu.ru)

The mechanism of accumulation of age-related damage in mtDNA remains unclear. It is unknown whether an increase in the levels of damaged mtDNA correlates with age. In this study, we analyzed the accumulation of mtDNA damage in organs of young and old rats and try to correlate them with the profile of oxylipins. In addition, we tested the possibility of reversing these changes by dietary restriction (DR). We showed that the mtDNA in brain structures (excluding the cerebellum) showed the most pronounced accumulation of age-related damage, similarly shown for the kidney, while the liver, testes and lungs were less susceptible to age-related accumulation of mitochondrial damage. DR prevented age-related accumulation of mtDNA damage in the cerebral cortex and reduced it in the lungs and testes. We also observed changes in the number of mtDNA copies, and these alterations were also tissue-specific. There was a tendency to an age-related decrease in the number of mtDNA copies in the striatum and its increase in the kidneys. DR contributed to an increase in the amount of mtDNA in the cerebellum and hippocampus. Analysis of the profiles of polyunsaturated fatty acids (PUFA) and oxylipins in old and young rats showed that DR reduces the synthesis of arachidonic acid and its metabolites. Thus, we have shown that aging significantly affects the amount of oxidative damage to mtDNA and oxidized fatty acid derivatives. DR significantly affects these processes, returning the studied parameters to those in young rats.

*Acknowledgements:* Study was supported by Russian Science Foundation (No. 21-75-30009).

## **Alteration of hippocampal glial support during early postnatal development as a possible premise of Alzheimer's disease**

Rudnitskaya E.A.<sup>1\*</sup>, Kozlova T.A.<sup>1</sup>, Burnyasheva A.O.<sup>1</sup>, Maksimova K.Yi.<sup>2</sup>, Stefanova N.A.<sup>1</sup>, Kolosova N.G.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Department of Histology, Embryology and Cytology, Siberian State Medical University, Tomsk, Russia*

\* *ekaterina.rudnitskaia@gmail.com*

Recent studies using state of the art diagnostic techniques have confirmed that the preclinical period of sporadic (> 95 % of cases) form of Alzheimer's disease (AD) can last for decades, but the question of when exactly the disease development begins and what contributes to its development remains open. Several epidemiological and experimental studies indicated that the predisposition for accelerated aging, the main risk factor for AD development, can be formed during the early postnatal period of life, at the age of the completion of brain maturation. The results of our work on the senescence-accelerated OXYS rats, a unique model of sporadic AD, have confirmed the validity of this hypothesis. We have identified the features of brain maturation in OXYS rats in the early postnatal period (at the ages P0-P20, P – postnatal day) which can act as prerequisites for the development of initial neurodegenerative changes in the later age. OXYS rats have a lower (comparing with their maternal Wistar strain) body weight at birth, decreased duration of gestation, a delayed physical development and formation of reflexes in the postnatal period against the background of impaired neuronal stem cells differentiation and the formation of mossy fibers in the dentate gyrus of the hippocampus, and a decrease in the efficiency of formation of synaptic contacts in OXYS rats. The completion of brain maturation in OXYS rats occurs against the background of a decrease in astrocytic and microglial support – a key regulator of the neural circuitry function – taking place in the hippocampus and prefrontal cortex. The lack of glial support may be the cause of the decreased efficiency of the formation of synaptic contacts found in OXYS rats, which made it possible to consider it as a key event leading to the development of AD signs in the future.

*Acknowledgements:* This work was supported by Russian Science Foundation (grant No. 19-15-00044).

# Evaluation of the regulatory potential of polymorphic variants of complement system genes associated with age-related macular degeneration

Rumyantseva Yu.V., Kozhevnikova O.S.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [rumyantseva@bionet.nsc.ru](mailto:rumyantseva@bionet.nsc.ru)

**Key words:** age-related macular degeneration, regulatory SNP, complement system

*Motivation and Aim:* Age-related macular degeneration (AMD) is the most common cause of irreversible and severe visual loss in the elderly population. Genome-wide association studies revealed associations between AMD and variants of alternative complement pathway -associated genes, including the locus ARMS2, whose connection with the complement system has been proven recently. Recently we found that rs10490924 (*ARMS2*), rs800292 (*CFH*) and rs6677604 (*CFH*) polymorphisms are strongly associated with exudative AMD in Russian population and the response to antiangiogenic therapy differed according to the patient's specific rs2285714 (*CFI*) and rs2230199 (*C3*) genotype. This study aimed to assess the functional significance of the rs2285714 (*CFI*), rs10490924 (*ARMS2*), rs2230199 (*C3*), rs800292 (*CFH*), and rs6677604 (*CFH*) polymorphisms by *in silico* analysis.

*Methods and Algorithms:* To estimate the potential downstream functional effects of the AMD-associated variants, we used the available data of epigenetic effects (HaploReg, RegulomeDB), non-synonymous functional predictions (SNPinfo), expression and alternative splicing quantitative traits (GTEx consortium atlas).

*Results:* Among the five SNPs studied, 3 loci (rs800292 *CFH*, rs10490924 *ARMS2* and rs2230199 *C3*) cause the replacement of amino acids in the encoded polypeptide. Also, downstream gene variant (rs2285714 *CFI*) and intron variant (rs6677604 *CFH*) were determined. According to the HaploReg database, rs2285714 is located in evolutionarily conserved region, rs2230199—in the promoter histone marks region, and rs6677604 and rs2230199—in the enhancer histone marks region in various tissues. Three SNPs are located in the hypersensitivity region to DNase-1. All SNPs are located in the DNA regulatory motifs regions. For example, the risk allele of rs2230199 may alter affinity to transcription factors SP1, VDR, ZBTB7A and E2F3, the rs2285714 influences on affinity to HSF transcription factor. Also the SNPs in strong linkage disequilibrium with studied polymorphisms were determined: from two SNPs for rs2230199 to more than 20 SNPs for rs10490924.

According to the SNPinfo database, polymorphisms rs2230199 *C3* and rs2285714 *CFI* possess the most significant regulatory potential (0.33 and 0.36 respectively). The same SNPs have the highest the RegulomeDB probability scores (0.86, and 0.6). Notably, both SNPs are located in the regions of exonic splicing enhancer (ESE) and exonic splicing silencer (ESS). SNPs that are located at ESE or ESS may disrupt splicing activity and cause alternative splicing. Recently, we have shown a decrease in the level of factor FI, encoded by the *CFI* gene, in the intraocular fluid, but not in the plasma of patients

carrying the risk rs2285714 genotype. This effect may be partly explained by the effect 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 *PLA2G1A2A*) 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:* Using bioinformatic analysis we showed the regulatory potential of SNPs in AMD-associated genes that possessed epigenetic effects and are involved in the control of gene expression and alternative splicing of genes that contribute to AMD pathophysiology. The previously discovered association of SNPs rs2285714 and rs2230199 with response to antiangiogenic therapy in patients with AMD can be explained by the high regulatory potential of these SNPs.

*Acknowledgements:* The study is supported by State Budget Project No. FWNR-2022-0016.

## Transcription regulation in aging and lifespan control

Rybina O.Y.<sup>1, 2\*</sup>, Pasyukova E.G.<sup>1</sup>

<sup>1</sup> *Institute of Molecular Genetics, National Research Centre “Kurchatov Institute”, Moscow, Russia*

<sup>2</sup> *Federal State Budgetary Educational Institution of Higher Education “Moscow Pedagogical State University”, Moscow, Russia*

\* *flybee@mail.ru*

**Key words:** transcription regulation, lifespan, aging, transcription factor

Spatio-temporal regulation of gene expression governs the functioning of the organism, its development and aging. Tissue-specific changes in the activity of embryonic transcription factors have been shown to cause gene expression reprogramming, resulting in a reduction of age-dependent disorders. Thus, the induced change in transcriptional regulation could be one of the mechanisms for direct control of lifespan. In this regard, the study of tissue-specific transcription factors that are responsible for the organism functioning at a young age is of particular interest. Human *Lhx3/4* is a transcription factor which is required for pituitary development and motor neuron specification. *Lhx3/4* is highly homologous to *Drosophila Lim3*, which encodes an RNA polymerase II transcription factor involved in motor neuron specification.

*Lim3* has a significant impact on *Drosophila* lifespan while being misregulated ubiquitously in the neuronal system and muscles. However, the strongest *Lim3* effect on fly lifespan is observed when it is down-regulated at an early age of *Drosophila* development and is associated with the alteration of mitochondrial and neuromuscular junction activities, an increase in ATP level and *Drosophila* locomotion activity. To determine the mechanisms underlying the impact of *Lim3* on the lifespan of flies, we studied the natural variability of the *Lim3* regulatory region. We found that naturally occurring polymorphism in *Lim3* promoter that is associated with gene expression variations and lifespan control in *Drosophila* are located exclusively in the Polycomb response element (PRE). We revealed this region to be enriched with Polycomb group (PcG) proteins: *Polycomb (Pc)*, Polyhomeotic, Pleiohomeotic *proteins*. Furthermore, mutations in *Pc* affected *Lim3* expression, thus confirming functional interactions between this protein and *Lim3* regulatory region. Thus, PcG proteins that are thought to mediate epigenetic silencing are directly involved in the regulation of *Lim3* transcription, which is associated with the control of *Drosophila* lifespan.

## Aberrant neuroplasticity in aging and neurodegeneration associated with deregulation of neurogenesis and angiogenesis

Salmina A.B.<sup>1,2\*</sup>, Gorina Ya.V.<sup>2</sup>, Komleva Yu.K.<sup>2</sup>, Khilazheva E.D.<sup>2</sup>, Pisareva N.V.<sup>2</sup>, Salmin V.V.<sup>2</sup>, Shuvaev A.N.<sup>2</sup>, Malinovskaya N.A.<sup>2</sup>

<sup>1</sup> *Research Center of Neurology, Moscow, Russia*

<sup>2</sup> *Prof. V.F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia*

\* *allasalmina@mail.ru*

**Key words:** aging, neurodegeneration, brain plasticity, neurogenesis, angiogenesis

**Motivation and Aim:** Neuroplasticity is a common phenomenon that underlies brain's response to the changing environment. Aging is always considered as a gradual loss of brain plasticity caused by impairment of synaptogenesis and neurogenesis, mitochondrial dysfunction, accumulation of misfolded proteins, development of oxidative stress, and excessive cell death [1]. Within neurogenic niches, neurogenesis is supported by local microvessels that provide the establishment of microenvironment needed for stem and progenitor cells maintenance and recruitment. Neoangiogenesis is required for effective neurogenesis and neuroblasts migration toward the loci of neurodegeneration. Moreover, coupling of neurogenesis and angiogenesis in the adult brain contributes to the experience-driven plasticity [2] which is greatly compromised in aging and neurodegeneration, thereby leading to cognitive decline.

**Methods and Algorithms:** We utilized neurogenic niche *in vitro*, *ex vivo* and *in vivo* models of for the assessment of the proliferative activity of neurospheres, extracellular lactate and NAD<sup>+</sup> levels, expression of markers of neurogenesis, angiogenesis and neuroinflammation.

**Results:** *In vitro* neurogenic niche model has demonstrated its functionality in the context of assessing the proliferative and differentiation potential of NSCs/NPCs, as well as *in vitro* modeling of neurodegeneration (exposure to the neurotoxic beta-amyloid A $\beta$ 1-42). We found that neurosphere-forming capacity of hippocampal stem/progenitor cells (NSCs/NPCs) *in vitro* was differently compromised in aging animals and in rodents with the model of Alzheimer's type neurodegeneration. Photoactivated ChR2-expressing astrocytes within the niche increase the proliferative activity of NSCs, change the expression of key differentiation markers. In the presence of neurospheres, there was a decrease in the number of vascular tubes formed *in vitro* from brain microvessel endothelial cells (BMECs). However, glycolytic activity of astrocytes was required for efficient angiogenesis and barriergenesis.

**Conclusion:** Optogenetic activation of astrocytes is effective in controlling the activity of neurogenesis and cerebral angiogenesis and might be considered as a tool for correction of aging- or neurodegeneration-associated changes in brain plasticity.

### References

1. Salmin V.V., Komleva Y.K., Kuvacheva N.V., Morgun A.V., Khilazheva E.D., Lopatina O.L., Pozhilenkova E.A., Shapovalov K.A., Uspenskaya Y.A., Salmina A.B. Differential roles of environmental enrichment in alzheimer's type of neurodegeneration and physiological aging. *Front Aging Neurosci.* 2017;9:245. doi: 10.3389/fnagi.2017.00245.
2. Malinovskaya N.A., Komleva Y.K., Salmin V.V., Morgun A.V., Shuvaev A.N., Panina Y.A., Boitsova E.B., Salmina A.B. Endothelial progenitor cells physiology and metabolic plasticity in brain angiogenesis and blood-brain barrier modeling. *Front Physiol.* 2016;7:599. doi: 10.3389/fphys.2016.00599.

# Investigation of the microRNA expression dynamics in the retina of rats during aging and the development of retinopathy similar to age-related macular degeneration

Shklyar A.<sup>1\*</sup>, Kozhevnikova O.<sup>2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* Shklyaraa@bionet.nsc.ru

**Key words:** AMD, microRNA, OXYS rats

*Motivation and Aim:* Age-related macular degeneration (AMD) is a leading cause of blindness worldwide affecting individuals over the age of 50. MicroRNAs are a class of single-stranded, non-coding RNAs, about 22 nucleotides in length, which negatively regulate gene expression at the post-transcriptional level. MicroRNAs are dysregulated in AMD and may facilitate the early detection of the disease and monitoring disease progression. However there is practically no information about the patterns of microRNA expression in the retina at different stages of AMD development. The confirmed AMD model is OXYS rats, which develop retinopathy, corresponding to AMD in clinical, morphological and ultrastructural manifestations. The aim of the study was to evaluate the microRNA level dynamics (miR-9a, miR-16, miR-27, miR-34a, miR-146, miR-155, miR-320) in the retina of Wistar rats (control) with age and OXYS rats during the development of AMD- like retinopathy at preclinical (20 days), early (6 months), and advanced (12 and 18 months) stages of disease.

*Methods:* The data for analysis were obtained from retina samples using the method of microRNA reverse transcription and Taqman qPCR (Applied Biosystems). Relative fold gene expression changes were calculated by the  $2^{-\Delta\Delta C_t}$  method.

*Results:* We showed that the expression of miR-27a was elevated with age in the OXYS rats. The levels of miR-27 at the age of 12 and 18 months were considerably higher than at the age of 21 days ( $p = 0,0007$  and  $p = 0,007$  accordingly). No significant associations with age or strain were found with miR-9, miR-34a, miR-146a, miR-155a and miR-320a. Given that miR-27a could promote choroidal neovascularization and activate WNT signaling pathway, our data suggest a possible contribution of miR-27 to the development of retinopathy in OXYS rats.

*Acknowledgements:* The study is supported by RSF No. 21-15-00047.

## The emerging role of SIRT6 in the mitochondrial regulation

Smirnov D.<sup>1,2,3\*</sup>, Eremenko E.<sup>2,3</sup>, Stein D.<sup>2,3</sup>, Kaluski S.<sup>2,3</sup>, Jasinska W.<sup>2</sup>,  
Khrameeva E.<sup>1</sup>, Toiber D.<sup>2,3</sup>

<sup>1</sup> Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup> Department of Life Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel

<sup>3</sup> The Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer Sheva, Israel

\* Dmitrii.Smirnov@skoltech.ru

**Key words:** SIRT6, mitochondria, cellular metabolism

**Motivation and Aim:** SIRT6, a member of the SIRTUIN family, is localized in the nucleus and regulates many intracellular processes through two main enzymatic functions – histone deacetylation and mono-ADP-ribosylation. SIRT6 is implicated in genomic stability, DNA repair, telomere maintenance, cellular metabolism and, importantly, has critical roles in the brain protecting against aging-associated diseases [1]. Moreover, SIRT6 deficiency increases mitochondrial defects, including mtDNA damage accumulation and impaired mitochondrial dynamics in the heart and skeletal muscles [2, 3]. However, despite many promising studies in recent years, the molecular mechanisms underlying SIRT6 impact on mitochondrial function in the brain are still poorly understood. In this work, we examined the SIRT6-induced changes in mitochondrial activity by comparing transcriptomic and metabolomic profiles of brain-specific SIRT6 knockout mice (SIRT6-KO) and wild-type mice (WT).

**Methods and Algorithms:** Generation of brain-specific SIRT6 conditional knockout mice, library preparation and RNA sequencing steps were performed in the Dr. Toiber laboratory [4] at the Ben-Gurion University of the Negev in Israel. Raw RNA reads of 22 mouse samples were filtered and trimmed using *fastp* [5] and then processed via version 3.0 of the *nf-core/rnaseq* pipeline [6]. Normalization and differential gene expression (DGE) analysis were performed via the *DESeq2* [7] package in the R programming environment. WT and SIRT6-KO samples from the mouse brain tissue were also collected for the liquid chromatography coupled with mass spectrometry (LC-MS) metabolomics experiment. Obtained metabolomics dataset was processed using an in-house script and analyzed using *MetaboAnalyst* software [8]. Differentially accumulated metabolites (DAMs) were defined according to the following criteria:  $|\text{Fold Change}| > 1.5$  and  $p\text{-value} < 0.05$ .

**Results:** Differential expression analysis of WT vs. SIRT6-KO samples revealed 217 significant genes related to essential mitochondrial processes, including oxidative phosphorylation, mitochondrial import, and tricarboxylic acid cycle. Importantly, down-regulated genes constituted a substantial proportion of all differentially expressed mitochondria-related features, exceeding 91 % of genes. Additionally, we found that the deficiency of SIRT6 leads to decreased mitochondrial content and reduced mitochondrial membrane potential. In line with transcriptomic results, preliminary analysis of mass spectrometry data of SIRT6-KO mouse brains revealed 103 (out of 235) metabolites differentially accumulated in the SIRT6-KO mice compared to the control group. Among them are NAD<sup>+</sup> ( $p\text{-value} = 0.0011$ ), malate ( $p\text{-value} = 0.0012$ ), fumarate ( $p\text{-value} = 0.0015$ ), NADH ( $p\text{-value} = 0.0100$ ), and alpha-ketoglutarate ( $p\text{-value} =$



= 0.0194) metabolites that were previously shown to play a direct role in the oxidative phosphorylation process [9]. Taken together, our results emphasize the crucial role of SIRT6 in the essential mitochondrial processes and provide new insights into its potential targets.

*Acknowledgements:* The research was funded by RFBR and Moscow City Government according to the project No. 21-34-70051.

### *References*

1. Mostoslavsky R. et al. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell*. 2006;124(2):315-29.
2. Pillai V.B., Samant S., Hund S., Gupta M., Gupta M.P. The nuclear sirtuin SIRT6 protects the heart from developing aging-associated myocyte senescence and cardiac hypertrophy. *Aging (Albany NY)*. 2021;13:12334-12358.
3. Cheng M.Y., Cheng Y.W., Yan J., Hu X.Q., Zhang H., Wang Z.R., Yin Q., Cheng W. SIRT6 suppresses mitochondrial defects and cell death via the NF- $\kappa$ B pathway in myocardial hypoxia/reoxygenation induced injury. *Am J Transl Res*. 2016;8(11):5005-5015.
4. [online] [access: 3 November, 2021] <https://toiber.wixsite.com/toiber-lab>
5. Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890.
6. [online] [access: 3 November, 2021] <https://nf-co.re/rnaseq/3.4>
7. Love M.I., Huber W., Anders S. Moderated estimation of foldchange and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
8. Pang Z., Chong J., Zhou G., Morais D., Chang L., Barrette M., Gauthier C., Jacques P.E., Li S., Xia J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(W1):W388-W396.
9. Wilson D.F. Oxidative phosphorylation: regulation and role in cellular and tissue metabolism. *J Physiol*. 2017;595(23):7023-7038.

# The rat brain transcriptome: from infancy to aging and sporadic Alzheimer's disease-like pathology

Stefanova N.A.\*, Kolosova N.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [stefanovan@bionet.nsc.ru](mailto:stefanovan@bionet.nsc.ru)

**Key words:** Alzheimer's disease, RNA sequencing, early postnatal period, OXYS rats

*Motivation and Aim:* It is suggested that anatomical (quantity and connectedness of neurons) and functional (capability to use alternative neural circuits) traits of adult brain – all of which are established in the early life – might impact brain susceptibility to Alzheimer's disease (AD), the most common late-life dementia. The results of our work on the senescence-accelerated OXYS rats, a unique model of sporadic AD (the most prevalent form of the disease; ~95 % of cases), have confirmed the validity of this hypothesis. We have identified the features of brain maturation in OXYS rats in the early postnatal period which can act as prerequisites for the development of initial neurodegenerative changes in the later age. It naturally follows that the events at the molecular-genetic level underlie the phenomenon of the delayed brain maturation in OXYS rats.

*Methods and Algorithms:* To elucidate the nature of the revealed changes in brain maturation we analyzed transcriptomes (RNA-seq data) of the prefrontal cortex (PFC) and hippocampus of OXYS and Wistar rats at the ages of P3 and P10 (P – postnatal day) – which according to our estimates is the critical period of brain maturation in OXYS rats. In addition, we performed a comparative gene expression analysis in the rat PFC and hippocampus from infancy to aging and AD-like pathology using these results and previous RNA sequencing data as comprehensive approach toward understanding the molecular genetic mechanisms for the development, aging and dementia.

*Results:* Analysis of transcriptomes of the hippocampus and prefrontal cortex of OXYS and Wistar rats in the early postnatal period revealed metabolic pathways and processes, changes in which contribute to the delay in brain maturation in OXYS rats. At the age of P3 and P10 in OXYS rats, the number of differentially expressed genes (DEGs) was more than 1.000 in both brain structures ( $P_{adj} < 0.05$ ); changes in the expression of more than 200 of them were unidirectional in the hippocampus and cerebral cortex at both age points. The functional annotation of the DEG indicates a reduced efficiency of the formation of neuronal contacts, insufficiency of astrocytic and microglial support, mitochondrial functions, and angiogenesis delay. Comparison of these results with the results of the transcriptomes of the hippocampus and prefrontal cortex in the period preceding the development of AD signs (age 20 days), during their manifestation (5 mo) and active progression (18 mo) made it possible to evaluate for the first time on a sporadic model forms of AD, molecular genetic prerequisites and mechanisms for the development of the disease at all its stages. Three genes *Thoc3*, *Exosc8*, and *Smpd4* were identified; their expression were increased in both brain regions of OXYS rats throughout life.

*Conclusion:* Thus, we here demonstrated that reduced efficiency of the formation of neural networks in the brain of OXYS rats in infancy obviously contributes to the development of AD-like pathology. The reason for the delay in brain maturation may be

a decrease in the duration of gestation and reduced birth weight that were previously identified in OXYS rats. Our results raise the question that decreased duration of gestation and delayed brain development in infancy may be high risk factors for developing AD in the future.

*Acknowledgements:* This work was supported by a Russian Scientific Foundation grant (19-15-00044e).

## Nerve growth factor in the aging brain: friend or foe?

Stepanichev M.\*, Onufriev M., Lazareva N., Dobryakova Yu., Bolshakov A., Gulyaeva N.

*Institute of Higher Nervous Activity and Neurophysiology, RAS, Moscow, Russia*

\* *m\_stepanichev@ihna.ru*

**Key words:** brain, aging, nerve growth factor, brain-derived neurotrophic factor, basal forebrain cholinergic neurons, TrkA, TrkB, p75NTR

*Motivation and Aim:* It is well-known that nerve growth factor (NGF) plays an important role in the brain development and maturation in early ontogeny. It is considered that in the mature brain NGF controls survival of cholinergic neurons in the basal forebrain. However, possible contribution of NGF to the functioning of the adult and, particularly, aging brain is not completely known. In the present study, we investigated how neurotrophins and their receptors changed with aging and whether the NGF overexpression protected the brain in a model of cholinergic degeneration.

*Methods:* The experiments were performed in 3-, 17–18- or 22–24-month-old male Wistar rats. Animal behavior was studied using water maze. Brain samples were dissected and the levels of NGF or brain-derived neurotrophic factor (BDNF) were measured in hippocampus and frontal cortex using specific ELISA kits (Millipore, USA). The levels of TrkA, TrkB or p75NTR receptors were assayed using western-blot analysis with specific antibodies (SCBT or Millipore, USA).

*Results:* Aged 22–24-month-old rats exhibited slower rate of learning compared to young 3-month-old rats; however, long-term memory indices were similar in these groups. The content of NGF but not BDNF increased in both neocortex and hippocampus during aging. Significant age-dependent decrease in active, maximally glycosylated forms of TrkA (140 kDa) and TrkB (145 kDa) was revealed in the neocortex; while no major changes of partially glycosylated forms (110 and 95 kDa respectively) could be detected. In the hippocampus, similar decrease was revealed in the content of 95 kDa TrkB receptor whereas 110kDa TrkA receptor increased. Age-related changes in neocortical p75 receptor were similar to those of Trk receptors. Taken together, aging induced a complex decrease in neurotrophin receptors in the neocortex indicating significant alterations in neurotrophin signaling in the aged brain and these alterations were more pronounced in the NGF-associated processes.

Overexpression of NGF in the hippocampus significantly improved behavioral impairments induced in rats by administration of immunotoxin 192IgG-saporin into the medial septal nucleus, a model of Alzheimer-like cholinergic deficit. It was associated with a significant improvement of hippocampal neuroplasticity impaired after 192IgG-saporin administration. However, we did not find any effect of this treatment on survival of the basal forebrain cholinergic neurons or their functioning.

*Conclusion:* Thus, aging is accompanied by changes in the neurotrophin system in the neocortex and hippocampus and these changes were unidirectional in the neocortex and the opposite in the hippocampus. NGF overexpression in the hippocampus of rats with cholinergic deficit significantly improved animal behavior and neuroplasticity but did not protect cholinergic neurons from degeneration.

*Acknowledgements:* This study was supported by the Russian Science Foundation (projects No. 22-15-00132 in part of the experiments related to aging and No. 16-15-10403 in part of the experiments with saporin model).

## **Both melatonin and mitochondrial antioxidant SkQ1 had different effect on the retinal glutamate / GABA in healthy Wistar and senescence-accelerated OXYS rats**

Telegina D.V.\*, Fursova An.Zh., Kolosova N.G.

*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *telegina@bionet.nsc.ru*

**Key words:** glutamate, GABA, retina, aging, retinopathy, melatonin, SkQ1, OXYS rats

*Motivation and Aim:* Glutamate and  $\ominus$ -aminobutyric acid (GABA) are the most abundant amino acids in the retina and a proper balance between neural excitation and inhibition is crucial for retinal physiological functions. An imbalance between them may cause a series of pathological changes in the retina involved in the pathogenesis of many neurodegenerative diseases such as age-related macular degeneration (AMD). In the present study, we explored the role of alterations in the glutamatergic and GABAergic systems in the retina during aging and progression of AMD using mitochondria-targeted antioxidant SkQ1 and melatonin as therapeutic agents with previously proven retinoprotective properties.

*Methods:* To conduct such studies, we used senescence-accelerated OXYS rats – a unique model of premature aging whose manifestation is early development of retinopathy similar to AMD. To assess the influence of oral melatonin or SkQ1 administration (from age 12 to 18 months) on aging and progression of the AMD-like pathology, we randomly assigned 12-month-old male Wistar and OXYS rats to one of six groups (8 rats per group). Two groups (Wistar and OXYS rats,  $n = 8$  in each group) consumed a diet supplemented with SkQ1 (250 nmol/[kg body weight]) and two other groups (Wistar and OXYS rats,  $n = 8$  in each group) consumed a diet supplemented with 0.04 mg/[kg body weight] of melatonin (Melaxen; Unipharm, NY, USA). The supplements were provided on dried bread slices between ages 12 and 18 months every day. Each rat in these four groups received either melatonin or SkQ1 with dried bread individually daily at 8 p.m. Twelve-month-old OXYS and Wistar rats ( $n = 8$  each strain) that ingested dried bread without a drug served as controls (two more groups).

All the rats were examined by an ophthalmologist two times: before supplementation at the age of 12 months, and after of the treatment at age 18 months. All the rats underwent funduscopy with a Heine BETA 200 TL Direct Ophthalmoscope (Heine, Herrsching, Germany) after dilatation with 1 % tropicamide. The protein level of glutamate/GABA system were analyzed by Western blotting and immunohistochemistry.

*Results:* We did not detect any signs of retinopathy at the first inspection and during repeated inspections of the retina in control Wistar rats and Wistar rats treated with either SkQ1 or melatonin. There were no differences in the average severity of retinopathy among the groups of OXYS rats during the first examination at the age of 12 months. Pairwise comparisons of the eye condition at the age of 12 and 18 months showed that clinical signs of retinopathy progressed in untreated OXYS rats, whereas melatonin attenuated the rate of progression and SkQ1 completely prevented the progression. As a consequence (one-way ANOVA, Newman–Keuls test), the average severity of

retinopathy in OXYS rats treated with either SkQ1 ( $F_{1,92} = 32.3$ ,  $p < 0.0000$ ) or melatonin ( $F_{1,92} = 6.4$ ,  $p < 0.01$ ) was significantly lower than that in control groups. Supplementation with either melatonin or SkQ1 had no effect on protein amounts of either glutaminase or glutamine synthetase in OXYS rats. In the retina of Wistar rats, supplementation with SkQ1 or melatonin raised the protein level of glutaminase and reduced that of glutamine synthetase ( $p < 0.05$ ). Furthermore, supplementation with melatonin or SkQ1 diminished the GLAST protein level in Wistar rats ( $p < 0.05$ ) and had no significant effect in OXYS rats. Treatment with melatonin had no influence on NMDAR1 and GluA1 protein levels in both Wistar and OXYS rats. There was a marginally significant decrease in the GluA1 level in Wistar rats after melatonin supplementation ( $p = 0.08$ ). Treatment with SkQ1 significantly lowered retinal NMDAR1 expression only in OXYS rats ( $p < 0.05$ ).

Supplementation with melatonin and SkQ1 had no impact on GABA-T and GAT1 protein levels in the retina of the two strains. We observed that SkQ1 raised the level of GAD67 in the retina of Wistar rats but not OXYS rats. By contrast, treatment with melatonin increased the GAD67 protein level in OXYS rats and had no effect in Wistar rats ( $p < 0.05$ ). Supplementation with either melatonin or SkQ1 had no influence on the GABAAR1 protein level in the OXYS retina. Nonetheless, each drug lowered the GABAAR1 protein level in the Wistar retina ( $p < 0.05$ ). Similar data were obtained by immunostaining of the retina from untreated Wistar and OXYS rats at various ages. Both GAD67 and GABA-T were detectable in the ganglion cell layer, inner and outer plexiform layers, and in individual cells of the inner nuclear layer. Nevertheless, fluorescence intensity of GAD67 was higher in the ganglion cell layer and inner plexiform layer, while fluorescence intensity of GABA-T was greater in the outer plexiform layer. GAT1 and GABAAR1 were found to be predominantly expressed in the ganglion cell layer, inner plexiform cell layer, and in individual cells of the inner nuclear layer.

*Conclusion:* Here we confirmed the ability of the supplementation with SkQ1 to completely suppress the progression of retinopathy in OXYS rats already having pronounced clinical signs of AMD and for the first time report the ability of melatonin to suppress the progression of the clinical signs of the disease. In spite of delaying the development of AMD-like retinopathy, supplementation with melatonin or SkQ1 only moderately affected the glutamate/GABA system in OXYS rats. In contrast, supplementation with melatonin and SkQ1 (separately) caused significant alterations of the glutamate/GABA system in Wistar rats. We observed decreasing protein levels of glutamine synthetase and GLAST and an increasing level of glutaminase. These data are suggestive of alterations in the rapid clearance and recycling of glutamate in neurons and glial cells. We found that prolonged use of these drugs can worsen the condition of the retina during normal aging by contributing to gliosis and to an excessive glutamate level in the synaptic gap in the aged retina without a pathology.

*Acknowledgements:* This research was funded by Russian Science Foundation grant No. 21-75-00029.

# Novel multivariate approach identifies LPA and VCAM1 as drug targets for human ageing

Timmers P.<sup>1,2</sup>, Tiys E.<sup>3,4</sup>, Joshi P.<sup>2</sup>, Wilson J.<sup>1,2</sup>, Tsepilov Y.<sup>3,5\*</sup>

<sup>1</sup> MRC Human Genetics Unit, MRC Institute for Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom

<sup>2</sup> Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

<sup>3</sup> Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> Laboratory of Glycogenomics, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>5</sup> Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* tsepilov@bionet.nsc.ru

**Key words:** GWAS, ageing, frailty, healthspan, lifespan, longevity, Mendelian randomisation

*Motivation and Aim:* The length and quality of life is important to us all, yet identification of promising drug targets for human ageing using genetics has had limited success. Quantifying the aging process is not straightforward. A variety of aging-related phenotypes has been studied as proxies, from chronological measurements such as the length of time from birth until the occurrence of a major disease (healthspan) or death (lifespan), to cellular deterioration measurements and holistic measurements. The benefit of combining GWASs of several aging phenotypes, especially in different populations, is the ability to detect biological mechanisms that influence multiple core components of aging. In the present study we characterize the common genetic aging phenotype, identify robust genomic loci and highlight proteins that may be potential drug targets for improving the length and quality of life. Replication was performed using a study of Finnish and Japanese subject survival.

*Methods and Algorithms:* We combine six large European-ancestry genome-wide association studies (GWAS) of human ageing traits – healthspan, father and mother lifespan, exceptional longevity, frailty index, and self-rated health – in a principal component framework that maximises their shared genetic architecture.

*Results:* The first principal component (GIP1) is more heritable than the original studies and shows strong genetic correlations with length of life as well as multiple indices of mental and physical wellbeing. We identify 27 genomic regions associated with GIP1, and provide additional, independent evidence for an effect on human ageing for loci near *HTT* and *MAML3*. Across the genome, GIP1 associations are enriched in genes involved in haem metabolism and pathways related to transcription, neurogenesis, homeostasis, proteolysis, intracellular signalling, immunity, and the muscle system. Finally, using proteome-wide two-sample Mendelian randomisation and colocalisation, we provide robust evidence for a detrimental effect of blood levels of apolipoprotein(a) (LPA) and vascular cell adhesion molecule 1 (VCAM1) on GIP1.

*Conclusion:* Our results demonstrate that combining multiple ageing traits using genetic principal components enhances power to detect biological targets for human ageing.

*Acknowledgements:* This study was supported by the Russian Ministry of Education and Science under the 5-100 Excellence Programme and by the budget project No. FWNR-2022-0020.

## **Analysis of mitochondrial supply of neuronal processes during the postnatal brain maturation in OXYS rats – a model of sporadic Alzheimer’s disease**

Yukhtanov D.<sup>1, 2\*</sup>, Muraleva N.<sup>1</sup>, Tyumentsev M.<sup>1</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *d.yukhtanov@g.nsu.ru*

**Key words:** Alzheimer's disease, neurodegeneration, mitochondria, OXYS rats

*Motivation and Aim:* Age is the most significant risk factor for the sporadic form of Alzheimer's disease (AD, sporadic form accounts for > 95 % of all cases) – the most common form of senile dementia, with one in nine people above 65 and one-third over 85 suffering AD. In recent years, convincing evidence of the validity of the “mitochondrial cascade” hypothesis [1] has been obtained; with said hypothesis positing that age-related dysfunction of mitochondria underlies pathogenesis of the sporadic form of AD. Previously using senescence-accelerated OXYS rats as model of the sporadic form of AD, we have, firstly, confirmed the validity of the mitochondrial hypothesis [2], and, secondly, raised the issue of genetically determined structural and functional changes in mitochondria potentially influencing the process of brain maturation in the early postnatal period and thus determining predisposition to the development of AD. Indeed, recently researchers have turned their attention to the fact that the risk factors of eventual cognitive impairment and accelerated brain aging (a major risk factor of AD development) could be laid down already in the early postnatal period of life, during completion of brain development. Evidently, the basis of successful neuronal differentiation and formation of neural circuits during final stages of brain maturation might be connected to the mitochondrial biogenesis and supply needed to meet brain energy demand. The goal of this study was to compare the parameters of OXYS rats’ brain development with the state of neuronal mitochondria – their distribution across cellular compartments in neurons of the hippocampus and brain cortex of OXYS rats, energy metabolism efficiency, mitochondrial dynamics, and traffic.

*Methods:* We examined prefrontal cortex and CA1 region of the hippocampus (brain regions most vulnerable to AD) of OXYS and Wistar (control) rat strains aged 7, 14 and 20 postnatal days. These ages are associated with the end of migration of neurons (7 days), the peak of synaptogenesis and apoptosis (14 days), and the end of nervous system maturation (20 days). Using electron microscopy, we have assessed such ultrastructural parameters as spatial density of mitochondria in the cortical and hippocampal neuropil and neuronal somata, their distribution to neuronal processes and synaptic terminals, as well as the intensity of mitochondrial dynamics (fission and fusion). Principal proteins governing mitochondrial traffic: TRAK1/2 (motor protein adapters), Miro1/2 (mitochondrial cargo adapters) and syntaphilin (synaptic docking protein) were measured using Western blotting. Functionally we have measured the mitochondrial transmembrane potential (a parameter reflecting mitochondrial ATP-synthesizing capacity), which was accomplished using confocal microscopy of hippocampal acute slices stained with potential-selective dye – this technique owing to the unique hippocampal architecture allows for selective assessment of the transmembrane potential in somata and neuropil.



*Results:* In the OXYS brain cortex even though during the peak synaptogenesis (P14) there is a tendency towards OXYS axonal terminals having increased mitochondrial supply compared to Wistar, by the end of postnatal brain development (P20), less mitochondria were found to reside in axonal terminals. At the same time, total mitochondrial content averaged across neuropil (mitochondria docked in terminals together with mitochondria in transit in neuronal processes between terminals and soma) was increasing in both strains from P7 to P20 but this increase was more pronounced in OXYS rats. In cortical neuropil we have also detected signs of time-dependent intensification of the activity of mitochondrial dynamics from P14 to P20. Neuropil of the CA1 region of the hippocampus mirrors changes found in the cortex in terms of elevated mitochondrial content in the neuropil. However, unlike in cortex, OXYS hippocampal mitochondria were found to localize more to the dendritic terminals. Additionally, while no interstrain differences in the mitochondrial supply to the axonal terminals were revealed, we have found that following the peak of synaptogenesis in Wistar rats the number of axonal mitochondria decreases but in OXYS rats this number remains unchanged. Interestingly we did not find signs of ATP deficiency in hippocampal neurons of OXYS rats – an assessment of the transmembrane potential of mitochondria of the CA1 field of the hippocampus at the ages P7, P14, and P20 using hippocampal acute slices did not reveal significant interstrain differences. Neither have we found marked changes in the ultrastructural parameters in the somata signifying that in the studied ages mitochondria still retain sufficient levels of ATP-synthesizing capacity and mitochondrial biogenesis. Regarding mitochondrial traffic proteins in the studied period, in the cortex of OXYS rats, only the levels of Miro2 at the P7 and P14 ages was higher than in Wistar, and by the age of P20 the interstrain differences were no longer present due to a sharp decrease in the protein level in OXYS rats. In the hippocampus of OXYS rats only the TRAK2 level increased by the end of brain maturation, but its content did not exceed the levels of the age-matched Wistar rats. Syntaphilin is present in both brain structures at a very low level, with no difference between OXYS and Wistar rats.

*Conclusion:* Our data indicate that in OXYS rats the period of brain maturation is marked by alterations of mitochondrial distribution to the neuropil of the brain structures most vulnerable to the AD, these alterations primarily concerning mitochondrial supply of synaptic terminals. Optimal mitochondrial biogenesis and distribution corresponding to the brain energy demands is the basis for the successful differentiation and maturation of neurons, synaptogenesis, formation of neural networks and the completion of the formation of the brain as a whole – processes which were shown to be impaired in OXYS rats [3]. Failure to perform this task in a tightly-regulated, time-sensitive manner might interfere with the synaptogenesis. Following that, the decreased ability of neurons to establish proper contacts in due time might ultimately lead to the formation of aberrant neural networks and exacerbation of energy insufficiency, pointing to the improper mitochondrial distribution as an important underlying factor of neurodegeneration.

*Acknowledgements:* This work was supported by the RSF grant 19-15-00044.

### *References*

1. Swerdlow R.H., Khan S.M. A “mitochondrial cascade hypothesis” for sporadic Alzheimer's disease. *Med Hypotheses*. 2004;63(1):8-20.
2. Tyumentsev M.A. et al. Mitochondrial dysfunction as a predictor and driver of Alzheimer's disease-like pathology in OXYS rats. *J Alzheimer's Dis*. 2018;63(3):1075-1088.
3. Rudnitskaya E.A. et al. Features of postnatal hippocampal development in a rat model of sporadic Alzheimer's disease. *Front Neurosci*. 2020;14:533.

# 13

Symposium “Big genetic  
Data Analysis, deep learning,  
mathematical modeling and  
supercomputing”



# **A novel, computationally tractable algorithm flags in big matrices every column associated in any way with others or a dependent variable, with much higher power when columns are linked like mutations in chromosomes**

Antezana M.A.

*Independent Researcher*

*marcos.antezana@gmail.com*

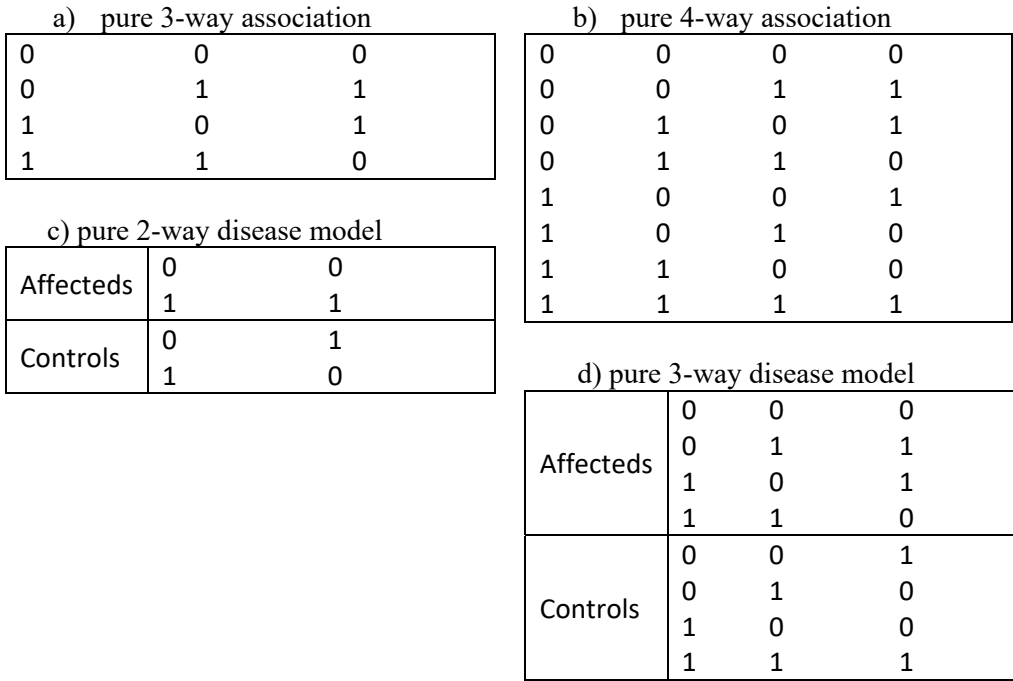
**Key words:** synergism, epistasis, non-additivity, additivity, interaction effect, marginal effect, 2-way, 3-way, higher-order effect, linkage disequilibrium, linkage, fine-mapping, coarse-mapping,  $\chi^2$  partitioning, mutation, combinations, complex trait, complex disease, broad-sense heritability, model heterogeneity, frequentistic, computational tractability, data mining, big data, large matrices, prime numbers

When a data matrix DM has many independent variables IVs, it is not computationally tractable to assess the association of every distinct IV subset with the dependent variable DV of the DM, because the number of subsets explodes combinatorially as IVs increase. But model selection and correcting for multiple tests is complex even with few IVs.

DMs in genomics will soon summarize millions of mutation markers and genomes. Searching exhaustively in such DMs for markers that alone or synergistically with others are associated with a trait is therefore computationally tractable only for 1- and 2-marker effects. Also population geneticists study mainly 2-marker combinations.

I present a computationally tractable, fully parallelizable Participation in Association Score (PAS) that in a DM with markers detects one by one every column that is strongly associated in any way with others. PAS does not examine column subsets and its computational cost grows linearly with the number of columns, remaining reasonable even when DMs have millions of columns.

PAS exploits how associations of markers in the rows of a DM cause associations of matches in the rows' pairwise comparisons. For every such comparison with a match at a tested column, PAS computes the matches at other columns by modifying the comparison's total matches (scored once per DM), yielding a distribution of conditional matches that reacts diagnostically to the associations of the tested column. Equally computationally tractable is dvPAS that flags DV-associated IVs by perturbing the matches at the DV. P values for the scores are readily obtained by permutation and accurately Sidak-corrected for multiple tests, bypassing model selection. A column's PAS P values PASs for different orders of association under the  $H_0$  and false-positive ones are i.i.d. and readily turned into a single P value. Simulations show that i) PAS and dvPAS generate uniform-(0,1)-distributed type I error in null DMs and ii) detect randomly encountered binary and trinary models of significant  $n$ -column association and  $n$ -IV association with a binary DV respectively, with power in the order of magnitude of exhaustive evaluation's and false positives that are uniform-(0,1)-distributed or straightforwardly tuned to be so. Power to detect 2-way DV-associated 100-marker+ runs disjunct and not is non-parametrically ultimate but that to detect pure  $n$ -column associations and pure  $n$ -IV DV associations sinks exponentially as  $n$  increases.



**Fig. 1.** Pure 3- and 4-way associations among binary markers and related pure 2- and 3-way binary disease models. Panels (a) and (b) show schemes of "pure" 3- and 4-way association of binary markers respectively, that characteristically miss half of the eight and sixteen possible 3- and 4-marker same-row combinations, despite showing all possible lower-order same-row marker combinations. If the 1s and 0s in the leftmost columns in (a,b) are relabelled as "affecteds" and "controls," resp., one obtains the models in (c,d) of "perfect" 2- and pure 3-way disease "causation" in which a set of markers in pure association causes disease and the other set protects against it (i.e., both sets have 100 % "penetrance"), without any differences in the lower-order associations of affecteds and controls, i.e., no marker-frequency differences in (c) and neither marker-frequency nor 2-marker-combination differences in (d). Similarly, a pure 5-way pattern becomes a disease model with pure 4-way causation, etc (not shown)

Important for geneticists, dvPAS power increases about twofold in trinary vs. binary DMs and by orders of magnitude with markers linked like mutations in chromosomes, specially in trinary DMs where furthermore dvPAS fine-maps with highest resolution.

# Glume pubescence prediction of spikelets using computer vision techniques

Artemenko N.V.<sup>1\*</sup>, Genaev M.A.<sup>2</sup>, Epifanov R.Yu.<sup>2</sup>, Kurochkina Y.V.<sup>3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

\* n.artemenko@g.nsu.ru

**Key words:** spikelet, pubescence, computer vision, convolutional neural network

*Motivation and Aim:* Glume pubescence of spikelets may be important feature for researching resistance of spikelets to different environmental conditions. In article [1] it was considered that spikelets with glume pubescence are better in storing heat, than those without it. At the same time glume pubescence can be hard feature to detect without special tools. So our aim was to research is it possible to distinguish spikelets with glume pubescence from those without pubescence using modern computer vision techniques.

*Methods and Algorithms:* For solving that task we decided to use Convolutional Neural Networks (CNNs). In our task the following species of wheat were used: '*T. monococcum*', '*T. boeoticum*', '*T. urartu*', '*T. yunnanense*', '*T. aestivum*', '*T. compactum*', '*T. spelta*', '*T. macha*', '*T. vavilovii*', '*T. petropavlovskiyi*', '*T. durum*', '*T. timopheevii*', '*T. dicoccoides*', '*T. turanicum*', '*T. carthlicum*', '*T. turgidum*', '*T. dicoccum*', '*T. aethiopicum*', '*T. polonicum*'. We used pretrained model for segmentation of spikelets to extract mask of spikelets. Then we applied findContours algorithm from OpenCV to distinguish spikelet from background. Also we add some data augmentations to our spikelets (like Crop, Rotation, Scale) using Albumentations library. Finally, we feed this data to CNN. We used following architectures: vgg16, ResNet18, EfficientNet b0-b3. For evaluating models we used accuracy, precision, auc.

*Results:* We tried different crop sizes, augmentations, models and loss function and obtained the following result, which is the best among all trained CNNs.

The best model: EfficientNet b1, trained and evaluated on 512x512 images.

Accuracy: 0.83

Precision: 0.69

AUC: 0.86

*Conclusion:* The obtained results can be used for predicting glume pubescence feature of spikelets.

*Acknowledgements:* The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS.

## References

1. Maes B., Treethowan R.M., Reynold M.P., van Ginkel M., Skovmand B. The influence of glume pubescence on spikelet temperature of wheat under freezing conditions. *Aust J Plant Physiol.* 2001;28(2):141-148. doi: 10.1071/PP00049.

# Deep learning for the development of an OCR for old Tibetan books

Brodt K.<sup>1,2\*</sup>, Rinchinov O.<sup>3</sup>, Bazarov A.<sup>3</sup>, Okunev A.<sup>2,4</sup>

<sup>1</sup> *Université de Montréal, Montréal, Canada*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *Institute for Mongolian, Buddhist and Tibetan Studies, SB RAS, Ulan-Ude, Russia*

<sup>4</sup> *MTS AI LLC, Novosibirsk, Russia*

\* *cyrill.brodt@gmail.com*

**Key words:** deep learning, Tibetan xylographs, optical character recognition (OCR)

*Motivation and Goals:* The funds of the Center of Oriental Manuscripts and Xylographs of the Institute for Mongolian, Buddhist and Tibetan Studies of the Siberian Branch of the Russian Academy of Sciences have the richest Buddhist heritage collections of manuscripts and woodblock printings, numbering about 100 thousand books in Tibetan and Mongolian. The manual processing of these writings and their translation into a machine-readable form is tedious and time-consuming [1]. Automation of optical character recognition or OCR allow us to save cultural treasures for the future generation.

*Methods and Algorithms:* We use computer vision and deep learning advances to apply the OCR for old Tibetan books. Experts in Tibetan studies at the Institute for Mongolian, Buddhist and Tibetan Studies of the Siberian Branch of the Russian Academy of Sciences annotated a dataset of scanned rare woodblock printings, namely Jone (Chone) Kangyur of early 18<sup>th</sup> century. These works have been chosen because they represent the woodcutting and printing quality typical for most Tibetan books. The dataset includes 500 images of book pages. Each image has text annotations.

We split the dataset into 420 training images and 30 testing images. The RGB image width is 2048 pixels and the average height is 650 pixels. Each image contains 8 text lines on average. Each line is framed with bounding box, and the corresponding text annotation is transliterated in Latin characters. The text length per line is 300 characters on average. First, we detect rectangular bounding boxes with text in image. After we crop out the boxes from image and extract the text using OCR model.

*Results:* We evaluate the performance of the baseline model on validation dataset, consisting of 30 high-resolution images each containing 8 text lines. We use standard metrics for OCR. The model reaches: character-by-character recall (char\_recall): 0.9495, character-by-character precision (char\_precision): 0.9539, normalized edit distance (1-N.E.D): 0.9397.

*Conclusion:* We have created a full-featured system for OCR of the Tibetan script. We implement a stream decoding the scanned text in Tibetan. The results of the project were included in the report of the President of the Russian Academy of Sciences Academician A.M. Sergeev, presented to the President of the Russian Federation V.V. Putin.

*Acknowledgements:* We acknowledge the financial support of MTS AI LLC. We also thank the Presidium of the SB RAS for the idea and organizational support.

## References

1. <https://escholarship.org/uc/item/6d5781k5>.

## The system of computer vision for extracting quantitative characteristics of wheat shoots

Busov I.

*Novosibirsk State University, Novosibirsk, Russia*

*i.busov@g.nsu.ru*

**Key words:** wheat shoots, computer vision, image processing, neural networks, CNN, segmentation, classification, statistical analysis

*Motivation and Aim:* Methods of analyzing quantitative characteristics of wheat shoots based on visual or tactile expert assessment have some disadvantages: subjectivity, labor input. An alternative method was implemented in the work, which is based on the analysis of images by computer vision methods.

*Methods and Algorithms:* The object of the research is photographs of wheat shoots. The following methods were used to solve the segmentation problem: a convolutional neural network of the Unet architecture [1], a modified tgi index [3] with further search for the maximum of the target function using a genetic algorithm [4].

Methods used to extract properties from an image: Depth-first search [6], Dijkstra's algorithm [6].

Methods used for statistical analysis of [5] components of the vector of image properties: Student's criterion, Bartlett's criterion, Pearson's criterion.

The methods that were used to solve the classification problem are: statistical regression, Random Forest machine learning algorithm, residual neural network of ResNet architecture [2] with 18, 50 and 101 layers.

The segmentation metric is IoU. The classification metric is accuracy.

*Results:* This computer vision system was tested to determine two quantitative characteristics: ploidy (belonging to the class of haploid or diploid plants), belonging to the genotype L-25 or genotype 1102. The results of segmentation with using the best model on the test sample for determining ploidy are 0.8467 according to the IoU metric, and 0.8938(IoU) on the test sample for determining the L-25/1102 genotype. When analyzing the vector of image properties, no statistically significant differences were found in the distribution of diploid and haploid photos of wheat shoots by components of this vector. The result of the ploidy classification for the best model on the test sample is 0.6000 according to the accuracy metric. The result for classification into the L-25/1102 gene type using this system is 0.9500 according to the accuracy metric on the test sample.

*Conclusion:* In this work, was developed a computer vision system that extracts quantitative characteristics of wheat shoots that are visible in photographs, with high accuracy.

*Acknowledgements:* The images were provided by the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences.

## References

1. Ronneberger O. et al. U-net: Convolutional networks for biomedical image segmentation. *arXiv:1505.04597*. 2015. doi: 10.48550/arXiv.1505.04597.
2. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;770-778. doi: 10.1109/CVPR.2016.90.
3. Alt V.V., Pestunov I.A., Melnikov P.V., Elkin O.V. Automated detection of weeds and evaluation of crop shoots quality based on RGB images. *Siberian Herald Agricultural Sci.* 2018;48(5):52-60. doi: 10.26898/0370-8799-2018-5-7.
4. Price K., Storn R. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J Glob Optim.* 1997;11(4):341-359. doi: 10.1023/A:1008202821328.
5. Gmurman V.E. Probability theory and mathematical statistics. Textbook. Moscow: Yurayt, 2018.
6. Cormen T.H., Leiserson C.E., Rivest R.L. Introduction to Algorithms. MIT Electrical Engineering and Computer Science. Cambridge: The MIT Press, 1990.



# Automatic identification of contaminating sequences in amplicon sequencing data

Fedorov D.<sup>1\*</sup>, Galeeva J.<sup>1</sup>, Manolov A.<sup>1</sup>, Klimina K.<sup>2</sup>, Veselovsky V.<sup>2</sup>, Pavlenko A.<sup>1</sup>, Ilina E.<sup>1</sup>

<sup>1</sup> *Institute of Systems Biology and Medicine, Moscow, Russia*

<sup>2</sup> *Research Institute of Physical-Chemical Medicine, Moscow, Russia*

\**fedorov.de@gmail.com*

**Key words:** microbiome, contamination, amplicon sequencing, metagenomics

*Motivation and Aim:* Amplicon sequencing allows for efficient study of microbiological communities without the need for microorganism cultivation. However, the PCR amplification stage introduces bias into the composition and quantity of the obtained amplicons. This effect can lead to sequencing of sequences from reagents, laboratory plastics, the environment and have a serious impact on the results and their interpretation [1]. This is especially critical when sequencing media with a low bacterial load. In this work, swabs from the nasopharynx of patients with a new coronavirus infection, as well as healthy volunteers, were analyzed. We observed the presence of contamination phenomena of unknown origin and proposed a method for automated search for anomalous structures in data and their subsequent cleaning.

*Methods:* 16s rRNA sequencing data were obtained from the Illumina MiSeq instrument, amplicon sequence variants (ASV) and their abundance in samples were obtained using the DADA2 [2] algorithm, the decontam [3] algorithm was used at the first stage of decontamination, the correlation network was obtained using the SPIEC-EASI [4] method, hierarchical clustering of sequences on the basis of the correlation matrix was performed by the ward.D2 algorithm using the R language.

*Results:* Using a combination of the decontam algorithm and hierarchical clustering based on a correlation matrix, it was possible to achieve automatic search and removal of contaminating sequences. Data cleaning allowed us to reduce the effect of the batch on the estimated composition of the community and to achieve a uniform distribution of samples in a multidimensional space. We observed that the representation of sequences identified as contaminants was specific to individual sequencer runs.

*Conclusion:* A method for the automatic identification and removal of contaminating sequences is proposed, which makes it possible to reduce the bias introduced by individual runs of the sequencer and provides an opportunity for accurate interpretation of the observed compositions of microbiological communities.

*Acknowledgements:* This study was supported by Russian Science Foundation No. 21-15-00431.

## References

1. Salter S.J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014;12:87.
2. Callahan B.J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-583.
3. Davis N.M. et al. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* 2018;6:226.
4. Kurtz Z.D. et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biol.* 2015;11:e1004226.

# Small world effect of the miRNA science field drives its growth

Firsov A.<sup>1\*</sup>, Titov I.<sup>2</sup>

<sup>1</sup> *A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* *artemijfirsov@mail.ru*

**Key words:** k-mer, n-gram, dbscan, identification, mirna, timsort, kofler, digital library, organization co-authorship, small world, power growth

*Motivation and Aims:* Many scientific articles became available in the digital form. One got an opportunity to query articles metadata, which can be used in estimating the statistics of the scientific field, building co-authorship graphs, etc. In our work, we focus on the miRNA science field, building the organization co-authorship network to provide higher-level analysis [1] of scientific community evolution rather than analyzing author-level characteristics. We show that the growth of the field is guided by the small world power-law property [2] on the macro scale by comparing the approximated power growth with models predicted. Additionally, we provide the analysis of the co-authorship graph.

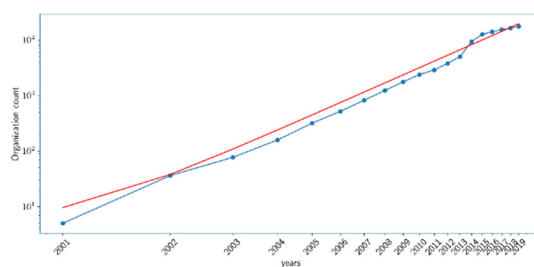
*Methods and Algorithms:* We used PubMed digital library to gather the miRNA science field dataset, which gave us the list of affiliations within the publications, therefore providing us with references to organization names, which published within this field. We disambiguated those affiliations using the lexicographical sorting on the K-Mer Boolean feature vectors (KOFER), normalizing and splitting them beforehand, providing the clustering metrics of such disambiguation. These grouped affiliations contain references to the same organization; therefore, we build the co-authorship graph looking at which organizations published the same paper together. We show that the miRNA science field has the small world property, approximate the power-law growth for the whole miRNA field publication activity and compare it with the models' predictions [2]. We provide the country-level statistics also approximating the power growth for each country. We analyze pioneers, using their publication activity, and show the properties of international interactions.

*Results:* The KOFER algorithm proposed showed these clustering metrics [3]:

1. *homogeneity* = 0.74279
2. *completeness* = 0.95280
3. *v-score* = 0.83

The completeness metric shows that the algorithm is stable to the name variations writings diversity and groups the similar organization names from the affiliation together, however, the homogeneity metric tells us that in some cases it fails to distinguish between very similar names and puts them into one cluster. This corresponds to what we saw while cleaning the disambiguation results.

We show using the «small-world-ness» metric [4] that the miRNA field has the small world property, and with Power law degree distribution parameter equal to 2.01, and the Degree assortativity coefficient equal to -0.03, it follows the models' criteria of the power-law growth [2]. The model states that the initial power growth of the nodes count should have the D-1 exponentiation parameter, where D is the average shortest path length. For our case, the  $D = 3.46$ , and the approximated exponentiation parameter  $(D-1) = 2.64$ , which has 7 % deviation from the model's one. This result shows that the miRNA field follows the models' prediction.



We show that not only organizations count follows the power law, but also the publications count, and we use that to compare the countries publication activity. The leaders of the field are the USA, and China, where the US had the high initial count of publication, but lower exponentiation parameter, than China, leading to China becoming the leader of the field in 2012. However, the US remain the biggest driver for international publications. We also show that countries with high publications count (>500) form almost a full graph, however, some countries did not publish together, such as Iran-Israel, Iran-Brazil, Turkey-Belgium. We also show that the top 10 of the most active organization in this field are primarily Chinese with two of them being American, and the leader organization as of 2019 is Nanjing Medical University.

Another interesting fact about that field, is that it had a leap in international publications count in 2012-2014, however, the overall publications count remained the same. We could not yet find the explanation of this effect, however, it is shown to also be present in other science fields, such as the "material science" research field [5].

Comparing the properties with the authors network properties [6], we see that the organizations graph is denser, leading to the graph diameter, distance, and other characteristics to be higher (or respectively lower, e.g., mean distance). Interestingly, «mean authors per paper» parameter is higher than the «mean organizations per paper» parameter. This might be caused by the fact that authors from same organization tend to publish more, than in co-operation with another institution. The pioneer's analysis show that they tend to publish more papers in collaboration with the organization from other countries compared to the fields' average, and some of them being one of the most active organizations per date, such as the University of California.

*Conclusion:* In this work we have implemented the algorithm for fast institution name clustering based on the K-Mer Boolean feature vector sorting – KOFER. Using that algorithm, we managed to cluster the miRNA science field affiliations data. Using the clustering results, we were able to show that the miRNA's organizations interaction graph satisfies the small-world effect and shows the spreading initial power-law growth, governed by the model [2]. We identified the leaders and pioneers of the field, characterized the country-level interactions, showing that the USA and China are the leaders.

*Acknowledgements:* The study is supported by the TBD.

## 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. doi: 10.48550/arXiv.1301.0801.
2. Vazquez A. Spreading dynamics on small-world networks with connectivity fluctuations and correlations. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2006;74(5 Pt 2):056101.
3. Rosenberg A., Hirschberg J. V-measure: A conditional entropy-based external cluster evaluation measure. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic. Association for Computational Linguistics. 2007;410-749.
4. Humphries M.D., Gurney K. Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One*. 2008;3(4):e0002051.
5. Li Y., Li H., Liu N., Liu X. Important institutions of interinstitutional scientific collaboration networks in materials science. *Scientometrics*. 2018;117(1):85-103.
6. Newman M.E.J., Moore C., Watts D.J. Mean-field solution of the small-world network model. *Phys Rev Lett*. 2000;84:3201-3204.

# Implementation of a spanning trees sampling algorithm to detect 4-motifs

Gam A., Yudina M.

Omsk State Technical University, Omsk, Russia  
 mail@mano.pro, mg-and-all@mail.ru

**Key words:** motifs, computation method, network analysis, graph theory

*Motivation and Aim:* Approximate counting all possible subgraphs is the task connected with some important network analysis methodologies like network motifs detection. Path sampling algorithms [1-4] seem to be the most promising among other statistical algorithms. Path sampling algorithms do not use heuristic assumptions, unlike graph generalization algorithms like TNP algorithm [5]. Path sampling algorithms make it possible to estimate subgraphs frequencies, in contrast to algorithms based on random walk estimate only subgraph concentrations. Finally, path sampling algorithms are implementations of the Monte Carlo method and allow one to obtain consistent, unbiased, effective statistical estimates, in contrast well known Rand-ESU algorithm [6] that does not meet the requirements in the general case [7].

*Methods and Algorithms:* However, the proposed in [1–3] path sampling algorithm does not allow working with directed graphs. As shown in our work [4], statistical experiments using all possible spanning trees are necessary to obtain frequency estimates of all possible subgraphs. The direction of the edges of spanning trees is not taken into account. In a graph, on three vertices there is only one isomorphism class of spanning trees (Fig. 1), on four vertices there are 2 classes, on five there are three, on six there are six. Different values to estimate a subgraph frequency of the same subgraph can be obtained, since the subgraph can be found by the sampling of different spanning trees (Fig. 2).

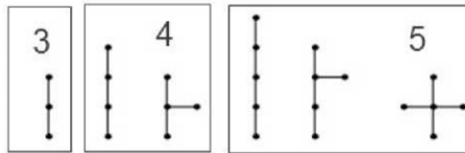


Fig. 1. Spanning trees with three, four and five vertices

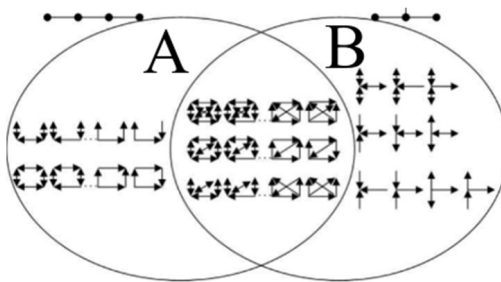
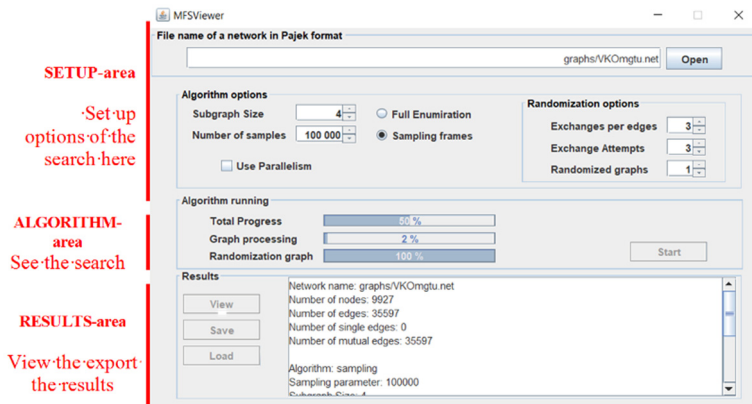


Fig. 2. Venn diagram of the coverage of the spanning trees of subgraphs belonging to different classes of isomorphism; A is the set of subgraphs that are covering by "4-path" spanning tree, B is the set of subgraphs that are covering by "4-branch" spanning tree

*Results:* The spanning trees sampling algorithm is a development of path sampling algorithms. It allows you to analyze not only undirected graphs, but also directed ones. The spanning trees sampling algorithm is implemented in the MFSViewer program. The program (<https://github.com/MNYudina/MFSView>) allows you to detect network motifs. Moreover, the use 68–95–99.7 rule makes it possible to identify network motifs statistically reliably. The program interface is shown in Fig. 3. The graphical interface, the help and the license of MFSViewer are similar to FANMOD ones [6]. However, MFSViewer differs from FANMOD in speed, accuracy [4] and quality indicators of the obtained statistical estimates (such as unbiased and efficiency).



**Fig. 3.** The three areas of the main interface window of MFSViewer. They correspond to the three working steps when using the tool: setting up the algorithm, running the algorithm, and processing the results

*Conclusion:* Despite the fact that compared to the analogues we know, such as FANMOD, Kavosh (Cytoscape), etc., our implementation looks competitive, we continue to work on improvements. At the moment, work is underway on colored motifs.

## References

1. Seshadhri C. et al. Triadic measures on graphs: The power of wedge sampling. In: Proceedings of the SIAM International Conference on Data Mining. 2013:10-18.
2. Jha M., Seshadhri C., Pinar A. Path sampling: A fast and provable method for estimating 4-vertex subgraph counts. In: Proceedings of the 24th International Conference on World Wide Web. 2015:495-505.
3. Wang P. et al. MOSS-5: A fast method of approximating counts of 5-node graphlets in large graphs. In: IEEE Transactions on Knowledge and Data Engineering. 2018:30(1):73-86.
4. Yudina M.N., Zadorozhnyi V.N., Yudin E.B. Mixed Random Sampling of Frames method for counting number of motifs. *J Physics: Conference Series*. 2019:1260(2):022013.
5. Przulj N. Biological network comparison using graphlet degree distribution. *Bioinformatics*. 2007:23(2):177-183.
6. Wernicke S., Rasche F. FANMOD: A tool for fast network motif detection. *Bioinformatics*. 2006:22(9):1152-1153.
7. Yudina M.N. Assessment of accuracy in calculations of network motif concentration by Rand ESU algorithm. *J Physics: Conference Series*. 2019:1260(2):022012.

# Classification of fruit flies by gender in images based on the YOLOv4-tiny neural network using mobile devices

Genaev M.A.<sup>1,2</sup>, Komyshev E.G.<sup>1,2</sup>, Afonnikov D.A.<sup>1,2\*</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [ada@bionet.nsc.ru](mailto:ada@bionet.nsc.ru)

**Key words:** YOLOv4-tiny, fruit flies detection, gender classification, mobile devices, FlyCounter, Android

*Motivation and Aim:* The fruit fly *Drosophila melanogaster* is a classic research object in genetics and systems biology. In genetic analysis of flies, a routine task is to determine the offspring size and gender ratio in their populations. Currently, these estimates are made manually, which is a very time-consuming process. The counting and gender determination of flies can be automated by using image analysis with deep learning neural networks on mobile devices.

*Results:* We proposed an algorithm based on the YOLOv4-tiny network to identify drosophila flies and determine their gender based on the protocol of taking pictures of insects on a white sheet of paper with a cell phone camera. 406 images with labels (total number of flies 35025) were collected for training the network. The increase of the training sample was performed due to generation of synthetic images (superposition of the contours with the artificial background, as well as mosaic transposition of image fragments). which allowed increasing the recall parameter from 0.981 to 0.991, precision from 0.628 to 0.726, and F1 from 0.766 to 0.838. The high value of the recall parameter indicates a high accuracy of fly identification in the image. Gender recognition is less accurate. We compared the accuracy of the gender recognition by neural network algorithm with the results obtained by expert geneticists on the same set of images. On the set of images obtained from mobile devices, the recognition accuracy of machine learning and expert geneticists was close (F1 = 0.726 vs 0.716), while on a sample of images under high-quality illumination and digital camera conditions the machine learning algorithms were noticeably worse (F1 = 0.824 vs 0.900). The results also show that the gender determination for females (F1 = 0.812) is worse than that of males (F1 = 0.873). Among the factors most strongly influencing the accuracy of fly gender recognition, the factor of fly position on the paper was the most important: flies with the lateral position are recognized the worst, but their proportion is the highest. In addition, increased light intensity and higher quality of the device cameras have a positive effect on the recognition accuracy.

*Conclusion:* We implement our method as the Android app FlyCounter for mobile devices, which performs all the image processing steps using device processors only. The time that the YOLOv4-tiny algorithm takes to process one image is less than 4 seconds.

*Acknowledgements:* The data analysis performed using computational resources of the “Bioinformatics” Joint Computational Center supported by the budget project No. FWNR-2022-0020.

# Uniqueness and stability of a cycle in the Elowitz–Leibler type dynamical system

Golubyatnikov V.<sup>1,2\*</sup>, Minushkina L.<sup>1</sup>

<sup>1</sup> Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* Vladimir.golubyatnikov1@fulbrightmail.org

**Key words:** circular gene network, positive feedbacks, negative feedbacks, stable cycle

We consider 6D block-linear dynamical system of Elowitz–Leibler type:

$$\begin{aligned} dx_1/dt &= k_1(L_1(y_3) - x_1); & dy_1/dt &= d_1(\Gamma_1(x_1) - y_1); & dx_2/dt &= k_2(L_2(y_1) - x_2); \\ dy_2/dt &= d_2(\Gamma_2(x_2) - y_2); & dx_3/dt &= k_3(L_3(y_2) - x_3); & dy_3/dt &= d_3(\Gamma_3(x_3) - y_3). \end{aligned} \quad (1)$$

This dynamical system simulates functioning of circular gene network described in [1]. The decreasing step-functions  $L_i$  describe negative feedbacks, increasing step-functions  $\Gamma_j$  correspond to positive feedbacks in the gene network;  $j = 1, 2, 3$ . We have constructed in [2] an invariant toric domain  $W_1$  in the phase portrait of the system (1) and we have shown there that this domain contains at least one cycle of this system. Now, our main result is:

**Theorem.** If  $L_j(0) = \max L_j > 1$ , and  $\max \Gamma_j > 1$ , then the system (1) has exactly one cycle in the domain  $W_1$ ; this cycle is stable.

*Acknowledgements:* Supported by grant FWNF-2022-009 of SB RAS.

## References

1. Elowitz M.B., Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403:335-338.
2. Golubyatnikov V.P., Minushkina L.S. Combinatorics and geometry of circular gene networks models. *Lett Vavilov J Genet Breed*. 2020;6(4):188-192.

# Mathematical and numerical modelling of the circadian oscillator

Golubyatnikov V.<sup>1\*</sup>, Akinshin A.<sup>2</sup>, Ayupova N.<sup>1</sup>, Kirillova N.<sup>1</sup>, Podkolodnyy N.<sup>3,4</sup>,  
Podkolodnaya O.<sup>3</sup>

<sup>1</sup> Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia

<sup>2</sup> JetBrains, St. Petersburg, Russia

<sup>3</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>4</sup> Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia

\* Vladimir.golubyatnikov1@fulbrightmail.org

**Key words:** circadian oscillator, positive feedbacks, negative feedbacks, cloud application

We construct smooth 6-dimensional dynamical system of the kinetic type

$$dx_1/dt = k_1(\Gamma_1(x_2) \cdot \gamma_1(x_3) - x_1); \quad dx_j/dt = k_j(\Gamma_j(x_5) \cdot L_j(x_1) - x_j); \quad j = 2,3,4;$$

$$dx_5/dt = k_5(\Gamma_5(x_6) - x_5); \quad dx_6/dt = k_6(C \cdot L_6(x_4) - x_6); \quad C = \text{const}; \quad (1)$$

as a model of circadian oscillator, see [1]. Here, all the variables denote concentrations of the components and are non-negative, all parameters and functions are positive. The functions  $L_i$  decrease monotonically and describe negative feedbacks,  $i = 2,3,4,6$ . The functions  $\gamma_1$  and  $\Gamma_s$  increase monotonically and describe positive feedbacks in this oscillator,  $s = 1, \dots, 5$ .

We find conditions of uniqueness and non-stability of an equilibrium point of the system [1], and in this case, we show existence of a cycle of this system near this equilibrium point.

For numerical experiments with trajectories of the system (1), and for visualizations of these results, a specialized cloud application is elaborated on the basis of the R language: <https://andreyakinshin.shinyapps.io/clock-bmall/>

Limit cycles and their bifurcations were observed in these experiments for appropriate values of parameters of the system (1).

*Acknowledgements:* The authors were supported by the State Contracts of Sobolev Institute of Mathematics, SB RAS (project No. 0314\_2019\_0011), Institute of Cytology and Genetics, SB RAS (FWNR-2022-0020), Institute of Computational Mathematics and Mathematical Geophysics, SB RAS (No. 0251-2021-0004) and partially by the Russian Foundation for Basic Research (project No. 20\_31\_90011).

## References

1. Golubyatnikov V.P., Podkolodnaya O.A., Podkolodnyy N.P., Ayupova N.B., Kirillova N.E., Yunosheva E.V. Conditions of existence of cycles in two basic models of circadian oscillator of Mammalians. *J Appl Ind Math.* 2021;15(4):1-14.



# Acoustic tomography: inverse problem of source recovery

Guber A.<sup>1\*</sup>, Shishlenin M.<sup>1,2</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia

\* a.guber@g.nsu.ru

**Key words:** inverse problem, numerical analysis, acoustic tomography

*Motivation and Aim:* Acoustic tomography of the lungs is a method of detecting various pathologies of the respiratory system. This is safer than CT, because during the examination a person is not exposed to radiation. Moreover, acoustic tomographs are much cheaper than classical computed tomographs, which makes it possible to massively use acoustic tomography methods for the diagnosis of, for example, malignant tumors or complications of COVID-19 – pneumonia, which is considered to be the most common cause of death of patients.

The report investigates the inverse problem of determining the source of acoustic waves in the three-dimensional case, the solution of which is necessary for a tomographic reconstruction of an acoustic image of the lungs. Similar problems also often arise in orthopedics when determining the degree of joint wear [1, 2], microseismic monitoring of hydraulic fracturing [3], etc.

*Methods and Algorithms:* Let us consider a direct problem for the acoustics equation in the domain  $\Omega = \{(x, y, z): x \in (0, L_x), y \in (0, L_y), z \in (0, L_z)\}$ :

$$u_{tt} = c^2(x, y, z)(\Delta u - (\nabla \rho(x, y, z), \nabla u)) + q(x, y, z, t), (x, y, z) \in \Omega, t \in (0, T), \quad (1)$$

$$u|_{t=0} = 0, u_t|_{t=0} = 0, \quad (2)$$

$$u|_{\partial\Omega} = 0 \quad (3)$$

by finding of the function  $u(x, y, z, t)$  given  $q(x, y, z, t), c(x, y, z), \rho(x, y, z)$ .

The inverse problem is to determine the source of vibrations  $q(x, y, z, t)$  according to the indications of the acoustic sensors located on the chest of the subject:

$$u(x_m, y_m, z_m, t) = f_m(t), m = \overline{1, M}. \quad (4)$$

Here  $(x_m, y_m, z_m)$  is the position of  $m$ -th receiver,  $M$  is the number of receivers. We assume that the functions  $c(x, y, z), \rho(x, y, z)$  are known.

The inverse problem of determining the function  $q(x, y, z, t)$  after finite-difference discretization is reduced to solving a large-dimensional SLAE.

$$Aq = f. \quad (5)$$

Here  $A$  is a  $M \cdot (N_t - 1)$  by  $(N_x + 1) \cdot (N_y + 1) \cdot (N_z + 1) \cdot (N_t - 1)$  matrix. The system of equations (5) is strongly underdetermined.

The singular value decomposition method and the  $r$ -solution are used to solve this SLAE.

*Results:* The decrease in the condition number of the matrix  $\mathbf{A}$  of the discrete operator of the inverse problem is investigated, and the behavior of the degree of sparsity of the matrix of the direct problem with an increase in the density of the grid partition over time is studied. Numerical methods for solving direct and inverse problems using OpenMP, MKL and CUDA technologies are implemented. The results of numerical calculations are given.

### *References*

1. Hirasawa Y., Takai S., Kim W.C., Takenaka N., Yoshino N., Watanabe Y. Biomechanical monitoring of healing bone based on acoustic emission technology. *Clin Orthop Relat Res.* 2002;402:236-244.
2. Wierzcholski K. Acoustic emission diagnosis for human joint cartilage diseases. *Acta Bioeng Biomech.* 2015;17(4):139-148.
3. Shishlenin M.A. Microseismic monitoring of hydraulic fracturing. In: Proceedings of the International Conference "Computational Mathematics and Mathematical Geophysics": Novosibirsk, October 08–12, 2018. Novosibirsk: ICM&MG SB RAS, 2018;407-413.
4. Godunov S.K., Antonov A.G., Kirilyuk O.I., Kostin V.I. Guaranteed accuracy of solving systems of linear equations in Euclidean spaces. Novosibirsk: Nauka, 1992.

## **Multidimensional analysis, dimension reduction, categorization with statistical approach – stability and reproducibility**

Karpenko D.V.

*National Research Center for Hematology, Moscow, Russia*

*d\_@list.ru*

**Key words:** multidimensional data analysis, binary clustering, secondary parameters, prognosis self check, statistical weight, statweight

*Motivation:* The analysis of modern medical and biological data arrays formulates various tasks in the multidimensional space of experimental data. Categorization tasks are associated with the formation of forecasts; dimensionality reduction can also solve the problem of extracting key parameters that describe groups. Modern powerful tools using neural networks involve the use of large training datasets. However, medical data are often characterized by their paucity and, in some cases, even blurring of the boundaries of the original groups and the presence of transitional elements. In most cases, we even don't know if groups differ by the given set of parameters. The complexity of the data obtained at different times in different sources adds more difficulties.

*Aims:* Create a unified, automated clustering and group description tool capable of processing mosaic data, which is resistant to single outliers. Additional tasks are formulated as taking into account individual fluctuations of measurements, taking into account the uncertainty of the group, and taking into account the pairwise relationships of parameters.

*Methods and Algorithms:* The boundaries of the dynamic range of data are determined, data outside this range are determined as fixed values corresponding to the maximum or minimum limit, and overestimated error values are signed to it. Values approaching the boundary of the dynamic range are retained, but are associated with a larger error. The data are converted to a logarithmic scale, and the group average and the scatter of values are calculated for each parameter separately. The mean and spread are calculated recursively using an exponential statistical weight (statweight) determined from the deviation from the mean and measurement error divided by the group scatter. After calculating the average for the group, the parameters of paired bonds are formed: in a two-dimensional space, a pair of parameters is determined – at what angle and with what spread the point of the group is observed relative to the center of the group. The set of obtained angles is interpreted as a set of independent parameters, thus the primary and secondary parameters become equal. Next, the problem of binary clustering is solved for each parameter separately through the question of which group the point differs more from. The difference is calculated as the difference between the value of the point and the group divided by the total error of the related objects, and the ratio of differences from the first and second groups is compressed in the range from 0 to 1. The calculated average for all parameters answers the question of binary clustering or forecasting. At this stage, a set of parameters that are optimal for group separation is selected. The stability of the prognostic model is checked by alternately excluding one element of the group and evaluating its prognostic value, and so on for all elements. Thus, the influence

of individual elements on small groups is checked. Early versions of algorithm was applied previously [1, 2].

*Results:* An algorithm was developed and tested in Python v.3.7 in Microsoft Windows. The program allows to process incoming data as separate blocks in Excel list, scripts perform the primary sorting of incoming data. The output is a list of parameters arranged according to their significance for the difference between the target groups and a graph in the form of dual ratios of points between the target groups. The mode of checking the stability of categorization through the exclusion of elements from groups results in a graph, where each point is marked separately and is located in the range from 0 to 1, depending on the differences from which group are more significant.

*Conclusion:* Dual clustering and space reduction are the development of the idea of scalar projection onto the axis connecting the centers of groups. A significant rethink is the comparison of distances from a point to a group in terms of the errors in determining the point and the group. Creating new parameters from the original ones is an attempt to define primitive relations between parameters and consider them as independent elements. It can be seen that the least squares method with a linear interpretation of dependence or the principal components analysis have similar foundations, namely, the selection of linear dependences. However, the schemes for introducing the used approaches with statweight, exclusion testing and binary prediction into these tools are not obvious. The scheme used made it possible to compare directly and equally the primary and secondary parameters. The creation of new parameters in the original multidimensional parameter space for small groups should be done with caution, since the data features, which are found that way, may be sporadic. The algorithm is able to find differences and reliably distinguish elements of groups, but when trying to carry out the proposed check, there is a strong inconsistency and the absence of a reliable predictive conclusion as a result of comparing insufficiently distinguishable groups. However, the algorithm and its checks perform well, when comparing sufficiently distinct groups. It is important to emphasize the significance of the proposed verification approach not only for the proposed algorithm, but also for other automated calculation schemes.

### *References*

1. Petinati N., Kapranov N., Davydova Y., Bigildeev A., Pshenichnikova O., Karpenko D., Drize N., Kuzmina L., Parovichnikova E., Savchenko V. Immunophenotypic characteristics of multipotent mesenchymal stromal cells that affect the efficacy of their use in the prevention of acute graft vs host disease. *World J Stem Cells*. 2020;12:1377-1395. doi: 10.4252/wjsc.v12.i11.1377.
2. Petinati N.A., Bigildeev A.E., Karpenko D.S., Sats N. V., Kapranov N.M., Davydova Y.O. et al. Humoral Effect of a B-Cell Tumor on the Bone Marrow Multipotent Mesenchymal Stromal Cells. *Biochemistry. (Mosc)*. 2021;(86):207-216. doi: 10.1134/S0006297921020097.

## Recursive matrix algorithm for calculating differential expressions

Karpenko D.V.

National Research Center for Hematology, Moscow, Russia

[d\\_@list.ru](mailto:d_@list.ru)

**Key words:** differential expression, statweight

*Motivation:* Differential gene expression calculations involve comparing the expression of a gene of interest and one or more housekeeping genes whose expression level is considered stable. Stability is defined as the maintenance of the pattern of gene expression relative to each other in the studied samples or groups of samples. Difficulties arise when it is necessary to compare expression in tissue cells with significantly different physiologies, since the expression of some genes including housekeeping genes can alter prominently among tissues. There will be a separate set of stable genes for separate sets of tissues. This gives rise to mosaicism in multiple comparisons; a similar mosaicism appears when comparing results from different experiments, when the set of studied genes is different. This creates natural difficulties in data processing, which are often solved individually in each analysis.

*Aims:* Create a unified, automated differential expression calculation tool capable of processing mosaic data and resistant to single outliers.

*Methods and Algorithms:* Expression data are calculated in logarithmic scale. First of all, the relative pairwise expression of all genes against all for each sample is calculated. For each pair separately, the mean and scatter are calculated between samples. Exponentially decaying statistical weight (statweight) is assigned to the values of the pair in the sample in accordance with the difference from the pair mean and the scatter for the pair. Taking into account the calculated statistical weight, we recursively recalculate the mean and the spread until the result stabilizes between iterations. A similar procedure is performed for the differences of a sample pair from the average pair across samples, but now the enumeration occurs for the sample gene over all its pairs – in fact, how stable is the deviation of the gene relative to the general pattern for each paired gene. The statweight is also used, which now takes into account the total instabilities of the gene relative to pairs in the sample, the instabilities of these pairs relative to each other in the samples, and the instability of the sample itself. For a gene, the differences of each pair from its average for samples are summed up, taking into account the statweight of this pair, so we get a matrix of corrections for gene expression by samples. We correct the initial differential expressions with the obtained corrections and recalculate recursively until stabilization in form of the corrections becomes less than a predetermined value. We sum up the final corrections accumulated in iterations, and this is the resulting differential expression data.

*Results:* An algorithm was developed and tested in Python v.3.7 for Microsoft Windows. The program allows processing incoming data as separate blocks in Excel list, scripts perform the primary sorting of incoming data. The output is in the form of an Excel table with matrices of values of relative expressions and estimated errors. The program allows assigning a deliberate statweight to the elements under consideration: a low statweight

can be set for obviously unstable objects or for a whole group of related genes that are stable relative to each other, but not relative to a reference gene. This allows introducing known assumptions about biological objects into calculations.

*Conclusion:* The program allows accelerating and automating the calculations of differential gene expression. Moreover, the algorithm as well as the program can also be applied in other areas requiring differential analysis and assuming the presence of characteristic profiles of the ratio of objects, such as, for example, in tasks on microbiota populations. The introduced ability to determine the preset statweight enables preferences in the calculations according to third-party information about the object. The use of the statweight makes the algorithm resistant to outliers and dynamically handles them. The algorithm assumes the presence of a distribution similar to a normal one and cases of multimodal distributions will probably be handled incorrectly.

## Artificial intelligence implementation in pharmaceutical products lifecycle

Koshechkin K.A.<sup>1</sup>, Lebedev G.S.<sup>1</sup>, Fartushny E.N.<sup>1</sup>, Orlov Y.L.<sup>1,2\*</sup>

<sup>1</sup> *The Digital Health Institute, I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** Machine learning, Deep learning, neural network, pharmaceutical products lifecycle, pharmaceutical industry, e-Health, review

*Motivation and Aim:* Artificial Intelligence (AI) meets more and more applications in biology, medicine, pharmacy challenging fundamental computational problems [3–5]. To examine systematically the overall approach for artificial intelligence implementation in pharmaceutical products lifecycle.

*Methods and Algorithms:* Published studies in PubMed and IEEEXplore were searched from inception to March, 2021. Studies were screened for relevant outcomes, publication types, and data sufficiency, and a total of 73 (1,2 %) out of 6131 studies were retrieved after study selection. Extraction of data was performed by authors according to PRISMA (<http://www.prisma-statement.org>), and the quality assessment was performed according to QUADUS-2 [1]. We reviewed pharmaceutical solutions (<https://www.cyclicarx.com/>, <https://intervenn.com/>, <https://www.ibm.com/watson-health>) and publications.

*Results:* The first two stages of literature search yielded 2617 hits from PubMed and 1466 hits from IEEEXplore. The volume of search results from Sechenov-AI Information System was 2048 publications in 446 journals. After screening titles and abstracts, we excluded 2837 studies for duplication and 2745 studies for non-relevant abstracts or publication types. A total of 549 studies went through full-text review, of which 189 studies were excluded for result that was not of interest and 215 were excluded for insufficient data for analysis. After further exclusion, only 154 studies were retrieved after the second stage of literature selection. The third stage of literature review, and 23 studies were added, which result in 73 studies retrieved after literature selection, resulting in a total number of included 177 AI implementations for final analysis. All of Artificial Intelligence systems could be divided in multiple overlapping categories. For the total of 177 projects found the most popular area of AI implementation is Clinical trial and preclinical tests (34 %). On the second place stood the Novel small molecules design systems with 33 % from total. The third most popular scope for AI implementation is Target identification for novel medicines. More than 25 % of the systems provide this functionality. It is interesting that most of the systems specialize in only one area (102 systems – 57 % are single purpose). None of the systems provide functionality for full coverage of the lifecycle and function in all categories of the tasks. We reviewed also COVID-19-related applications of AI (in predicting respiratory failure, prognosis of patients) [2]. COVID-19 pandemic has evaluated novel chemical and biologic entities as potential therapeutics for SARS-CoV-2 infection, the repositioning and off-label use of existing agents approved for unrelated conditions is

widely advocated as a therapeutic approach against COVID-19 as it offers more rapid, actionable interventions against the virus.

*Conclusion:* This meta-analysis demonstrated that Artificial Intelligence solutions in Pharmaceutical Products Lifecycle could find numerous implementations, and none of the market available solutions cover them all. Natural Language Processing could be used for real world evidence-based trials, reports and summaries generation. Artificial Intelligence methods could be used for large amount of unstructured information analysis for new drug targets identification as well as novel molecules *generation in silico* and their activities prediction. Drug repositioning and repurposing could be done with text mining. Clinical trials utilize real world evidence with Artificial Intelligence support. Image classification can automatically discover, generate, and learn features of images which is useful in preclinical and clinical trials results processing. Artificial Intelligence aided Personalized therapy is used for risk prediction, and multiple factors analysis including genetics. Drug dispense control helps to discover counter indications and drug combinations interactions.

### *References*

1. Chan H.C.S. et al. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol Sci.* 2019;40(8):592-604.
2. Ferrari D. et al. Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia—Challenges, strengths, and opportunities in a global health emergency. *PLoS One.* 2020;15(11):e0239172.
3. Koshechkin K., Lebedev G., Radzievsky G., Seepold R., Martinez N. Review of Blockchain Technology Projects to Provide Telemedical Services. *J Med Internet Res.* 2021;23(8):e17475. doi: 10.2196/17475.
4. Lebedev G. et al. Systematization of the principles and methods of applying for digital medicine in oncology. *Procedia Computer Sci.* 2021;192:3214-3224. doi: 10.1016/j.procs.2021.09.094.
5. Shaderkin I.A., Shaderkina V.A. Remote medical consultations for patients: what has changed in Russia in 20 years. *Russ J Telemed E-health.* 2021;7(2):7-17. doi: 10.29188/2712-9217-2021-7-2-7-17. (In Russian)



# Button press detection from EEG signals using deep learning

Kuzucu E.<sup>1\*</sup>, Savostyanov A.<sup>1,2</sup>

<sup>1</sup> *Department of Mechanics and Mathematics, Novosibirsk State University, Novosibirsk, Russia*

<sup>2</sup> *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk State University, Novosibirsk, Russia*

\* *e.kuzucu@g.nsu.ru*

**Key words:** deep learning, EEG, button press detection

*Motivation and Aim:* Electroencephalogram (EEG) signals are mainly used for neuroscience studies and for BCI applications with a rising trend. Majority of the work includes conducting experiments with different paradigms, and studying the results of experiments for scientific discoveries or for various practical applications. It is likely that in the near future, all EEG related scientific discoveries will combine into one comprehensive automated EEG analysis framework where different paradigms can be analyzed in terms of detection of events, markers and signals of interest. In this study we created an application which estimates time of button press actions from stop-signal paradigm EEG recordings, using deep learning methods. And verified the proposed deep learning model's inter subject compatibility by using one-left-out cross validation method for 11 subjects.

*Methods and Algorithms:* We down sampled EEG data by 20 and added extra dimension to the data in order to allow compatibility with 2D convolutional layers. And for the model we utilized 3 convolutional layers with filter sizes of 64, 32 and 16 respectively. And as a loss function as well as accuracy metric, we used Root Mean Square Error (RMSE). To have more interpretable results we also used L1 distance as secondary accuracy metric.

*Results:* For intra subject experiment, we split the dataset into train and test dataset and trained it for 300 epochs. During the training, minimum RMSE were detected as 6.04 and minimum average L1 distance of 40.4 ms also observed. For inter subject experiments, we calculated RMSE and L1 for all test subjects during left-one-out iteration and take average of these results. Average min RMSE was calculated as 5 and average of minimum average L1 distance of 14.2 ms also calculated.

*Conclusion:* The results obtained during these experiments shows that proposed model successfully decodes spectral features related to button press task in EEG data. And it can be used without further calibration for various Brain Computer Interface (BCI) applications as well as part of experimental marker detection framework for Stop Signal Paradigm experiments.

*Acknowledgements:* The EEG data collection was supported by grant No. 22-15-00142 of Russian Science Foundation.

## **Bioinformatics and nanobioelectronics. Informatics based on biocomputer technologies**

Lakhno V.D.

*Institute of Mathematical Problems of Biology – the Branch of Keldysh Institute of Applied Mathematic,  
RAS, Pushchino, Moscow Region, Russia  
lak@impb.ru*

**Key words:** DNA, data bases, charge transfer, biocomputer

Bioinformatics is the field that develops methods and software tools for understanding of biological data, which includes sequence analysis, gene and protein expression, analysis of cellular organization, structural bioinformatics, data bases etc. A new and more general direction is to consider bioinformatics as informatics on the bases of biocomputer technologies.

DNA molecular is a clear example of data storage and biocomputing. Performing millions of operations simultaneously DNA – biocomputer allows the performance rate to increase exponentially. The limitation problem is that each stage of paralleled operations requires time measured hours or days. To overcome this problem can nanobioelectronics [1–3].

The central problem of nanobioelectronics is the realization of effective charge transfer in biomacromolecules. The most promising molecule for this goal is DNA. Computer simulation of charge transfer can make up natural experiment in such complex object as DNA. Such processes of charge transport as Bloch oscillations, soliton evolution, polaron dynamics, breather creation and breather inspired charge transfer are modeled. The supercomputer simulation of charge dynamics at finite temperatures is presented. Different molecular devices based on DNA are considered.

### *References*

1. Lakhno V.D. DNA Nanobioelectronics. *Int J Quantum Chem.* 2008;108:1970-1981.
2. Lakhno V.D. Theoretical basis of Nanobioelectronics. *EPJ Web Conferences.* 2020;226:01008.
3. Lakhno V.D., Vinnikov A.V. Molecular devices based on DNA. *MBB.* 2021;16:115-135.

# Beta negative binomial mixture model facilitates identification of allele-specific gene regulation in high-throughput sequencing data

Meshcheryakov G.<sup>1\*</sup>, Boytsov A.<sup>2</sup>, Abramov S.<sup>2</sup>, Kulakovskiy I.V.<sup>1,2</sup>

<sup>1</sup> *Institute of Protein Research, RAS, Puschino, Russia*

<sup>2</sup> *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

\* *georgy@vega.protres.ru*

**Key words:** allelic imbalance, beta negative binomial, mixture model

*Motivation and Aim:* High-throughput sequencing allows identifying allele-specific regulatory sites, where alternative alleles of homologous chromosomes have different chromatin state (e.g. allele-specific chromatin accessibility, ASC) or different transcription factor binding affinity (allele-specific binding, ASB). The ASC and ASB events are highlighted by single-nucleotide substitutions in read alignments and provide valuable information on functional consequences of single-nucleotide variants in non-coding regions. The key approach to identify AS events is to assess the allelic imbalance of read counts supporting alternative alleles in ChIP-Seq (for ASBs) or DNase-Seq/ATAC-Seq (ASC) read alignments.

There are two challenges in statistical scoring of the observed allelic imbalance at candidate sites. First, the relative allelic copy numbers (allelic dosage) of the particular genomic segments is generally unknown. This challenge is solved by BABACHI[1], that performs unsupervised estimation of the relative background allelic dosage (BAD) directly from variant calls. Knowing  $BAD$ , the expected frequency of two allelic variants is  $f_1 = \frac{BAD}{BAD+1}$  and  $f_2 = 1 - f_1 = \frac{1}{BAD+1}$ .

The second challenge is to estimate whether the observed imbalance significantly deviates from the BAD estimates. In Abramov et al. [1], we used the negative binomial mixture model with  $f_1$  and  $f_2$ . This approach works reasonably well for ASBs, but fails to model the observed distributions of allelic read counts for ASCs. ATAC-Seq and DNase-Seq data yield higher variance and the negative binomial model fails to adequately approximate the empirical distribution. Here, we propose an improvement by assuming frequency to be a beta random variable.

*Methods and Algorithms:* Similarly to the previous work, we model the number of reads  $x_c$  supporting one of the alleles given a certain fixed number of reads  $c$  at the other allele. We assume  $x_c$  to be distributed as a mixture of two negative-binomial random variables with success rates  $f_1$  and  $f_2 = 1 - f_1$  respectively. In this study, we also let  $f_1$  to be a beta random variable centered at  $\mu = f_1 = \frac{BAD}{BAD+1}$ . The latter can be conveniently achieved with reparametrization of  $Beta(\alpha, \beta)$ :  $\alpha = \mu\kappa$ ,  $\beta = (1 - \mu)\kappa$  with mean  $\mu$  and concentration  $\kappa$ . Thus,

$$x_c \sim \omega NB(r, p) + (1 - \omega) NB(r, 1 - p), p \sim Beta(\mu\kappa, (1 - \mu)\kappa) (1)$$

where  $r, \omega$  are active parameters and “+” is a mixing operator of two distributions, whereas  $\omega, (1 - \omega)$  are mixture weights (note that in the special case of  $BAD = 1$  it

holds that  $p = 1 - p$  and  $\omega$  parameter is redundant). The proposed generative model at Equation 1 is simplified by integrating out the frequency parameter  $p$  via marginalization. The result is known as beta negative binomial distribution, and the final model attains the form of

$$x_c \sim \omega \text{BetaNB}(r, \mu, (1 - \mu)\kappa) + (1 - \omega) \text{BetaNB}(r, 1 - \mu, \mu\kappa). \quad (2)$$

Here  $r, \kappa, \omega$  are obtained as maximum likelihood (ML) estimates. To fit the model parameters to the observed distribution of read counts  $x_c$  at various fixed  $c$ , we used JAX automatic differentiation library to obtain analytical gradients of ML objective function and scipy's SLSQP solver to obtain local minima.

*Conclusion:* The proposed method is integrated into the MixALime python software (<https://github.com/autosome-ru/MixALime>) and will facilitate reliable identification of allele-specific regulatory events in a wide repertoire of sequencing assays, including DNase-Seq and ATAC-Seq.

### *References*

1. Abramov S. et al. Landscape of allele-specific transcription factor binding in the human genome. *Nat Commun.* 2021;12:2751.

## Towards artificial minds through computer simulation of natural ones, from simple to more complex

Palyanov A.

*A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*  
*palyanov@iis.nsk.su*

**Key words:** virtual organism, computational modelling, *C. elegans*, *Xenopus*, tadpole, computational neuroscience, biomechanics, hydrodynamics, Sibernetic software

*Motivation and Aim:* The problem of building an artificial mind (at least a simplest one capable of subjective experience) based on a computational system still remains a challenge. Even a fundamental possibility or impossibility of this remains unknown. There are approaches aimed at building artificial cognitive architectures and those which try to reproduce the natural ones at neuronal level in the form of a computer simulation. The advantage of the latter is the existence of biological prototypes which provide clues and guides for building their dynamical digital copies.

*Methods and Algorithms:* Modeling nervous systems requires the data about their neuronal connectivity and electrophysiological properties. However, it is highly preferable to provide them with meaningful sensory signals and ability to observe what actions correspond to the activity of their motor neurons. Being a part of two international projects working on investigation and computational modeling of nervous systems of simple multicellular creatures, we developed a software system named Sibernetic [1], which is capable of simulation and visualization of organism's bodies placed into virtual physical environment. Each virtual organism is provided with sensory and muscular systems connected with its simulated nervous system, so it can both move itself through contracting its muscles and receive the information about what is going on nearby as a result of this movement or for any other reason. In Sibernetic all simulated objects are represented by particles of different types (which are linked to each other in case of elastic objects) used to build organisms and environment. The latter includes incompressible liquid with density and viscosity of water at 20 °C, because both organisms use swimming for locomotion. The liquid is modeled via prediction-correction incompressible smoothed particle hydrodynamics (PCISPH) algorithm which we implemented in parallel form optimized for GPU computations using C++ and OpenCL.

*Results:* The organisms which we were attempting to simulate are invertebrate *Caenorhabditis elegans* nematode with 302 neurons (OpenWorm project) [2] and more complex one, vertebrate *Xenopus* tadpole, with thousands of neurons (2308 neurons of 12 types were included into the model) [3]. Nervous system of *C. elegans* appeared to be quite complicated to simulate, because its neurons are so specific that even lack action potentials and use passive electrotonic propagation of signals instead, so it is still far from being complete. However, virtual *Xenopus* tadpole driven by its nervous system, represented by a network of Hodgkin-Huxley type neurons, appeared to behave pretty natural – when touched in the middle of the body, it starts to swim (in nature it is

necessary to escape from predators) until the tadpole's head bumps into an obstacle and attaches to it with mucus from the cement gland. Even the swimming velocity of virtual tadpole turned out to be very close to the real one – 19.3 and 21 mm/s respectively.

*Conclusion:* Positive experience in modeling relatively simple organisms and their nervous systems together with the knowledge that human brain (86 billion neurons) and rat brain (0.2 billion neurons) are too complex to digitize them entirely at cellular level and simulate on a supercomputer, leads to the idea that the next object for modeling should be chosen thoughtfully so that the goal is achievable and useful. An adult frog has about 16 million neurons, but its cognitive abilities are not too impressive. *Drosophila*'s brain with its 200 thousands of neurons is already investigated quite well[4], and its neural circuits responsible for flying and navigation are of great interest. And, finally, ants with just 250 thousand of neurons are known to pass mirror test, i.e. distinguish between own reflection in a mirror and another ant seen through a glass [5], which is highly related to the origins of consciousness [6].

*Acknowledgements:* The study was performed according to the Russian Federation Government research assignment for A.P. Ershov Institute of Informatics Systems, SB RAS, project FWNU-2021-0005.

### *References*

1. Palyanov A. et al. Three-dimensional simulation of the *Caenorhabditis elegans* body and muscle cells in liquid and gel environments for behavioural analysis. *Phil Trans R Soc B*. 2018;373:20170376.
2. Sarma G.P. et al. OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Phil Trans R Soc B*. 2018;373:20170382.
3. Ferrario A., Palyanov A. et al. From decision to action: Detailed modelling of frog tadpoles reveals neuronal mechanisms of decision-making and reproduces unpredictable swimming movements in response to sensory signals. *PLoS Comput Biol*. 2021;17(12):e1009654.
4. Hulse B.K. et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife*. 2021;10:e66039.
5. Cammaerts M.-C., Cammaerts R. Are Ants (Hymenoptera, Formicidae) Capable of Self Recognition? *J Sci*. 2015;5(7):521-532.
6. Barron B.B., Klein C. What insects can tell us about the origins of consciousness. *PNAS*. 2016;113(18):4900-4908.

# Inverse modeling of biological processes in sensitivity operator framework

Penenko A.<sup>1\*</sup>, Skorik V.<sup>1</sup>, Zubairova U.<sup>1,2</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

\* [a.penenko@yandex.ru](mailto:a.penenko@yandex.ru)

**Key words:** inverse modeling, sensitivity operator, adjoint equations

*Motivation and Aim:* Due to the diversity of the inverse problem statements arising in the course of sophisticated natural systems modeling, it is essential to develop a general framework that allows both solving and estimating the inverse problem solution result before its actual solution. The examples of widely used inverse problems are the coefficient identification problems of finding the unknown kinetic parameters or locating synthesis regions in bio-chemical reaction systems based on the concentration measurements for the selected set of species.

*Methods and Algorithms:* We work in a setting when there is a biological system model defined by a system of differential equations and measurement data on the system state provided with a measurement operator connecting measurement result with the model state function. We use an approach which consists of reducing various inverse problems to a unified quasi-linear matrix equation family with sensitivity operators constructed from an ensemble of independent solutions of adjoint problems [1]. Ensemble structure of the algorithms allows for efficient implementation on modern HPC systems. The sensitivity operator properties allow predicting the efficiency of the inverse problem solution. The projector to the orthogonal complement of the sensitivity operator matrix kernel can be used as a predictor. Parameter identification problems for biological models are often solved by various derivative-free and gradient-based algorithms. We compare the developed sensitivity operator-based algorithm results to those obtained by standard implementations of the parameter identification algorithms [2].

*Results:* A general sensitivity operator-based framework can be efficiently applied for solving and analyzing various inverse modeling problems arising in biological processes studies. The developed kinetic parameter identification algorithm shows high efficiency compared to the standard implementations of gradient-based and derivative-free algorithms both in accuracy and computation time.

*Conclusion:* The sensitivity-operator-based approach can be successfully applied to the inverse modeling of the biological processes described by the differential equation models.

*Acknowledgements:* The study is supported by RFBR grant 20-01-00560.

## References

1. Penenko A.V., Mukatova Z.S., Salimova A.B. Numerical study of the coefficient identification algorithm based on ensembles of adjoint problem solutions for a production-destruction model. *Int J Nonlinear Sci Numer Simul.* 2020;22:581-592. doi: 10.1515/ijnsns-2019-0088.
2. Penenko A., Konopleva V., Bobrovskikh A. Numerical Comparison of the Adjoint Problem-based and Derivative-free Algorithms on the Coefficient Identification Problem for a Production-Loss Model 2021. *IEEE.* 2021. doi: 10.1109/opcs53376.2021.9588680.

## Non-Euclidean machine learning methods in systems biology

Poleksic A.

*University of Northern Iowa, Cedar Falls, Iowa, USA*

*poleksic@cs.uni.edu*

Accurate modeling of the geometry of a biological system enables the discovery of novel relationships from sparse, noisy and biased data. Machine learning methods in systems biology rely on representing biological objects as points in a low dimensional space, known as the *feature* or *latent* space. Traditionally, the native (latent) biological space is assumed to be the classical Euclidean space. While this assumption has been convenient (due to the availability of efficient statistical and numerical optimization tools in the flat space), recent theoretical studies suggest that a better modeling of the proximity of biological entities in a biological system can be achieved using the hyperbolic distance metric. The authors of these studies argue that, in contrast to the Euclidean space, the negatively curved (hyperbolic) space can accommodate the exponential growth of a number of relevant biological features, since the circumference of a hyperbolic circle and the disk area are exponential functions of the radius. Nevertheless, these theoretical claims have never been validated in practice nor they have been built upon to create improved methods for biological relationship inference.

To test the abovementioned hypotheses, we have developed a variant of the classical matrix factorization method, termed *hyperbolic matrix factorization*, that uses the hyperbolic distance metric to predict interactions and associations between drugs, genes, diseases, and other objects in a complex biological network. To assess the practical benefits of our technique, we carried out a head-to-head comparison of the Euclidean and hyperbolic matrix factorization on the task of predicting drug-target interactions. The results of our benchmarking experiments demonstrate a substantial increase in the accuracy of hyperbolic representation compared to Euclidean representation. In addition, we show that the latent dimension of an optimal hyperbolic embedding is, by an order of magnitude, smaller than the latent dimension of an optimal Euclidean embedding. We see this as direct evidence that hyperbolic geometry (rather than Euclidean) underlines the biological system.



# Functional domain annotation of protein sequences with deep metric learning

Russkikh N.<sup>1,2,3\*</sup>, Ryazantsev G.<sup>2</sup>, Antonets D.<sup>1,2,3</sup>, Shaburova E.<sup>1,2</sup>, Vyatkin Yu.<sup>2,4</sup>

<sup>1</sup> State Research Center of Virology and Biotechnology “Vector”, Koltsovo, Novosibirsk, Russia

<sup>2</sup> AcademGene LLC, Novosibirsk, Russia

<sup>3</sup> A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

\* russkikh@nprog.ru

**Key words:** representation learning, protein functional annotation, domain, deep metric learning

*Motivation and Aim:* Functional domain annotation of amino acid sequences is a very important and long-standing task. Identification of functional domains within the protein can provide us insights not only into its functions but also into its origin and evolution. Current methods of choice, such as PHMMER [1], HHSEARCH [2] etc. are based on homology detection using hidden Markov models (HMMs). Despite the achieved progress, the task remains challenging in certain cases, especially when dealing with sequences with very little or no homology towards the annotated proteins. We propose a deep metric learning approach to address this issue. Benchmark on remote homology task constructed from Pfam-seed dataset demonstrated that the proposed method makes fewer errors than the conventional methods and performs on par with deep learning-based methods while being able to predict multiple domains in the protein at once.

*Methods and Algorithms:* We treat protein functional annotation problem as a tokenwise classification deep learning task. Our model consists of a multilayer transformer encoder with a small two-layer perceptron on top of its output. As a transformer backbone, we used ProtBert [3]. Output features from four ProtBert intermediate layers, from 10 to 13, are concatenated and passed to the feedforward head, consisting of two layers with 1024 and 64 neurons respectively, with ReLU activation in between them. The model was trained on Pfam-seed dataset using soft triplet loss [4]. In order to classify the positional tokens, we used the NED (Normalized sum of Exponential of the Distances) algorithm [5].

*Results:* In this section we provide some benchmarking on the Pfam-seed dataset on both token-wise and domain-wise level. We use Pfam-seed dataset splits published at [6], both the random splits and the clustered ones. The only difference is regarded to multi-domain proteins. Since we consider samples on the whole protein level rather than individual domain level, some multi-domain proteins could be mixed (contain both train and validation domains, or train and test domains etc.) thus couldn't be correctly assigned to any set. In order to keep the test set exactly the same and to keep the property of remote homology between the train, validation and test sets, such “mixed” multi-domain proteins were discarded from the train and validation sets. On the aminoacid-level (tokenwise level), our model achieved 72.6 and 95.0 % macro averaged recall on clustered and random splits respectively. In order to obtain the domain-level predictions (and to compare with the existing domain-level methods), we averaged tokenwise predictions inside the domain borders assuming they're known.

## Clustered split of pfam-seed

Model	Error rate, %	Number of errors
Top Pick HMM	18.1	3844
phmmer	32.6*	6942
BLASTp	35.9*	7639
ProtCNN	27.7*	5882
ProtENN	12.2*	2590
Protomenal (ours)	20.9	4035

\* the results were taken from [6].

*Conclusion:* The results suggest that the deep metric learning is a very promising approach for annotating multi-domain proteins, especially with few or no known homologs. The model is deployed as a web service at <https://protomenal.com>. We are always open for collaboration, and especially with the experimental groups interested in proteomic research.

*Acknowledgements:* This work was supported by the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-15-2019-1665).

## References

1. Eddy S. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:e1002195.
2. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21:951-960. doi: 10.1093/bioinformatics/bti125.
3. Elnaggar A. et al. ProfTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing. *arXiv:2007.06225.* 2020. doi: 10.1101/2020.07.12.199554.
4. Qian Q., Shang L., Sun B., Hu J., et al. SoftTriple Loss: Deep Metric Learning Without Triplet Sampling. 2019;6449-6457. doi: 10.1109/ICCV.2019.00655.
5. Karpusha M. et al. Calibrated neighborhood aware confidence measure for deep metric learning, *arXiv:2006.04935.* 2020.
6. Bileschi M.L. et al. Using Deep Learning to Annotate the Protein Universe. *bioRxiv.* 2019;626507. doi: 10.1101/626507.

## Improved MHC–peptide class I interaction prediction with pretrained transformer

Shaburova E.V.<sup>1, 2\*</sup>, Antonets D.V.<sup>1, 2</sup>

<sup>1</sup> State Research Center of Virology and Biotechnology “Vector”, Koltsovo, Novosibirsk, Russia

<sup>2</sup> Novel Software Systems LLC, Novosibirsk, Russia

\* shaburova.liz@yandex.ru

**Key words:** MHCI-peptide presentation prediction, transformers, vaccines development

*Motivation and Aim:* Increase in amount of promising neoantigens and viral antigens has paved the way for development of highly specific vaccines. Precise prediction of the binding and presentation of peptides to major histocompatibility complex (MHC) alleles is an important step toward such therapies. The ideal tool for such a task should be panspecific, be able to learn from both binding affinity and presentation data sources, and be interpretable. Here we present a transformer neural network model which leverages self-supervised pretraining from a large corpus of protein sequences with the supervised training for MHCI-peptide interaction prediction, and benchmark it with current SOTA models.

*Methods and Algorithms:* For training and validation we used publicly available datasets on binding affinity data (BA) and mass-spectrometry data (EL), which reflects both events of binding and presentation respectively [1]. The model takes peptide and pseudosequence of MHC I allele, process it with their own pretrained BERT (we took pretrained protBERT model [2] and additionally pretrained it with our own data) to get embedding presentation of both amino acid sequences. After embeddings being concatenated, they undergo through 2 dense layers to get a prediction of binding affinity and presentation. As a benchmark models we selected MHCNuggets [3] and MHCFlurry 2.0 [4]. In order to compare performance on the test dataset, we measured accuracy, recall, precision, F-score, MCC (Matthews correlation coefficient) and RocAUC.

*Results:* Our model show comparable performance with both MHCNuggets and MHCFlurry models (MHCFlurry-BA, MHCFlurry-AP and MHCFlurry-PS). Also, being trained on EL dataset only, our model slightly outperformed the other models trained on the data of the same modality (MHCNuggets and MHCFlurry-AP), which suggest the capability of transformer architecture to accumulate more relevant information and generalize better.

*Conclusion:* The results obtained surmise transformer neural network model is being helpful in peptides binding and presentation prediction, which means they could be presented as a part of vaccine development pipeline. The further research should be focused on the ways of immunogenicity data inclusion to the training dataset and experiments with model architecture. Moreover, implying the fact that presentation of both class I and II epitopes are critical for the induction of a sustained effective immune response, the same model could be trained for both MHCI and MHCII peptide presentation prediction.

## References

1. Reynisson B. et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 2020;48(1):449-454.
2. Elnaggar A. et al. ProfTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Learning. *bioRxiv.* 2020.
3. Shao X.M. et al. High-throughput prediction of MHC class I and class II neoantigens with MHCnuggets. *Cancer Immunol Res.* 2020;8(3):396.
4. O'Donnell T.J., Rubinsteyn A., Laserson U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst.* 2020;11(1):42-48.

# Computational identification of Anti-CRISPR proteins by using information extracted from the primary structures

Sidra L.<sup>1</sup>, Saiqa A.<sup>2</sup>, Maryum B.<sup>1</sup>, Wajid A.A.<sup>1\*</sup>

<sup>1</sup> Computational Biology and Data Analysis Lab., Department of Computer Sciences & Information Technology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan

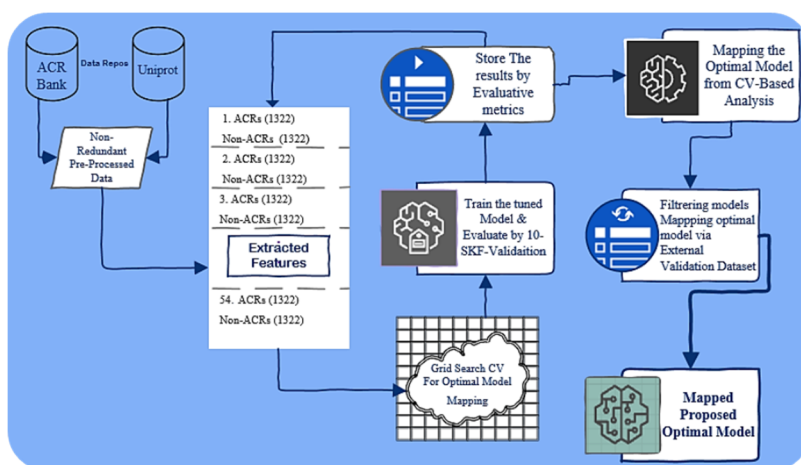
<sup>2</sup> Biotechnology Lab., Department of Zoology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan

\* wajidarshad@ajku.edu.pk; wajidarshad@gmail.com

**Key words:** Anti-CRIPR, proteins, machine learning, sequence information, predictive modeling

**Motivation and Aim:** The CRISPR-CAS technique has revolutionized genome editing, resulting in intrinsically engineered disease-resistant crops, livestock & other products. It has immense application in the personalized treatment of human diseases such as cancer, cystic fibrosis, and muscular dystrophy, and can be effective against infectious diseases such as malaria and HIV. The problem with the CRISPR-CAS system is that it lacks an off-switch to turn off Cas9's activity, which leads to off-target edits and mutations, resulting in adversity [1]. Anti-CRISPR (ACR) proteins proved to be natural off-switching-inhibitors for CAS's activity. Due to the immutable significance of ACR's a more consistent, quicker, and accurate approach is needed for discovering the Anti-CRISPR proteins to increase the rate and volume of discovered ACR's.

**Methods and Algorithms:** We have employed various machine learning algorithms such as large margin (Support Vector Machines), bagging based ensembles (Random Forest), and boosting based ensembles (Gradient boosting Machine). We used sequence derived structural descriptors of protein sequences to train and test different models. Various biological significant evaluation coupled with machine learning based matrices have been employed to gauge the generalization performance of our proposed solution. A generic framework of this study is shown in Fig. 1.



**Fig. 1.** Generic Framework of how we have mapped out the optimal ML-model for ACR-Predictor

*Results:* Using 10-fold cross validation we have achieved an area under the precision-recall curve of 0.96 using SVM and sequence derived structural descriptors. Similarly, using external validation dataset we have also achieved an area under the precision-recall curve of 0.97. These results show that our proposed predictive model predicts Anti-CRISPR (ACR) proteins with high precision (all the predicted ACRs are actually ACRs) and high recall (all the actual ACRs predicted correctly as ACRs). We have also checked the generalization performance through GO term enrichment analysis where we found a significant overlap between the terms of predicted and actual ACR proteins.

*Conclusion:* Thus, on the basis of deduced results, a reliable and appropriate SVM-based approach is carved out, based on selected optimal features to identify the candidate ACR's precisely and to deal with experimentally restrictive time and cost trade-offs.

### *References*

1. Carlaw T.M., Zhang L.H., Ross C.J.D. CRISPR/Cas9 Editing: Sparking Discussion on Safety in Light of the Need for New Therapeutics. *Hum Gene Ther.* 2020;31(15-16):794-807. doi: 10.1089/hum.2020.111.

## GeneCut – a software tool for oligonucleotide design, assembly and cloning of gene constructs

Sukhomlinov D.<sup>1\*</sup>, Pereverzev I.<sup>2</sup>, Bondiuk S.<sup>1</sup>, Seryakov A.<sup>1</sup>

<sup>1</sup> NCIT UNIPRO, LLC, Novosibirsk, Russia

<sup>2</sup> Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

\* [dsukhomlinov@unipro.ru](mailto:dsukhomlinov@unipro.ru)

**Key words:** oligonucleotides, codon optimization, gene assembly, cloning

*Motivation and Aim:* The synthesis of long nucleic acids is of great importance for modern molecular biology and genetic engineering, because it allows one to artificially create fragments with a given nucleotide sequence without using a DNA template. However, at present, it is technologically possible to synthesize only short DNA fragments, called oligonucleotides, which are subsequently assembled into a long DNA chain. But, when performing assembly *in vitro*, oligonucleotides in a mixture can form not only the necessary target, but also numerous non-target complexes – hairpin loops, self- and heterodimers. Computer modeling of oligonucleotide complexes makes it possible to provide the absence of stable non-target complexes and choose the optimal set of input data without carrying out experiments on real molecules.

The aim was to develop a tool for *in silico* experiments on gene splitting and molecular cloning. The tool should allow one to split the input target sequence into a set of oligonucleotides, that could be used to the further assembly the original sequence using the Polymerase Chain Reaction technology.

*Methods and Algorithms:* Unipro UGENE [1] is a desktop multiplatform software package that integrates dozens of widely used bioinformatics tools. Its core has been used as the basis for the technical implementation of a new product called GeneCut.

To split into fragments, you need to specify one long sequence and from one to four steps of the experiment, which are executed sequentially.

In the first step, "Optimize codon context", the sequence is modified in such a way that the fraction of occurrence of each codon in this sequence is as close as possible to the codon fraction values of the target organism from the corresponding codon frequency table.

At the second step "Exclude restriction and splicing sites", the chosen sites are replaced by sequences of the same length that do not contain these sites. The important point of the first two steps is to preserve the resulting amino acid sequence.

The third step "Long fragment assembly" is intended for situations when the input nucleotide sequence is long enough (about several thousand bases) sequences. In this case, first, splitting into long blocks is carried out, and then each block is split into short oligonucleotides. This step models the splitting for subsequent assembly by *Gibson assembly (GA)* [2] or *Overlap extension polymerase chain reaction (OE-PCR)*[3] methods. The difference between them is that the splitting by the *GA* method is carried out into blocks of a given length, and by the *OE-PCR* method - into blocks with specified temperatures of the sticky ends.

At the fourth step, "Oligonucleotides assembly", the splitting of the input sequence into a set of short oligonucleotides is modeled for subsequent assembly by *Polymerase*

*cycling assembly (PCA)* [4] or *Thermodynamically balanced inside-out (TBIO)* [5] methods. Their difference lies in the fact that when splitting by the *PCA* method, oligonucleotides are arranged in turns on opposite DNA strands, while when splitting by the *TBIO* method, they are located from the center on a direct strand towards the 5' end and at the reverse complementary towards the 3' end.

During cloning, the algorithm selects oligonucleotide primers corresponding to the specified melting temperature parameters, which allow completing and combining the specified vectors and/or fragments into a single construct. The DNA cloning methods Gibson Assembly and Gateway Technology [6] have been implemented.

*Results:* GeneCut has been used for *in silico* splitting and cloning of target genes. The “Codon optimization” feature has been used for: IL2 gene: wild type (403 nt), eukaryotic optimized (403 nt), prokaryotic optimized (426 nt) EPO gene: eukaryotic optimized (582 nt), prokaryotic optimized (603 nt).

The “Long fragments assembly” *OE-PCE* method has been used for splitting of bacteriophage MS2 RNA-dependent RNA polymerase gene (1323 nt). This gene has been split into 2 fragments of 668 nt and 699 nt.

The “Oligonucleotides assembly” Gibson assembly method has been used for splitting of the gene 5.5 kilobyte size (private data) into 8 fragments of 730 nt each of them.

The resulting oligonucleotides have been synthesized, gene constructs have been assembled, cloned into a plasmid, and sequenced.

*Conclusion:* A software tool GeneCut has been developed. This tool contains a set of algorithms that model the splitting of a DNA sequence into short oligonucleotides for their subsequent assembly. The efficiency of assembly from the obtained oligonucleotides was confirmed experimentally.

*Acknowledgements:* This project has been founded by the Candidate of chemistry Shevelev Georgy Yurevich, who tragically died in July 2021. The scientific component of this work is the result of his long work.

## References

1. Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28:1166-1167. doi: 10.1093/bioinformatics/bts091.
2. Gibson D., Young L., Chuang R.Y. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6:343-345. doi: 10.1038/nmeth.1318.
3. Horton R.M., Ho N.S., Pullen J.K., Hunt H.D., Cai Z., Pease L.R. Gene splicing by overlap extension. *Methods Enzymol*. 1993;217:270-279. doi: 10.1016/0076-6879(93)17067-f.
4. Stemmer W.P.C., Cramer A., Ha K.D., Brennan T.M., Heyneker H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*. 1995;164:49-53. doi: 10.1016/0378-1119(95)00511-4.
5. Gao X., Yo P., Keith A., Ragan T.J., Harris T.K. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res*. 2003;31(22):e143. doi: 10.1093/nar/gng143.
6. Katzen F. Gateway(®) recombinational cloning: a biological operating system. *Expert Opin Drug Discov*. 2007;2(4):571-589.



# Pygenomics: Python package for processing genomic intervals and bioinformatic data formats

Tamazian G. \*, Cherkasov N., Kanapin A., Samsonova A.

*Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia*

\*[g.tamazian@spbu.ru](mailto:g.tamazian@spbu.ru)

**Key words:** computational analysis, data formats, genomic intervals

*Motivation and Aim:* Computational analysis of genome sequencing data and its derivatives such as assembled genome sequences, annotated genes, repeats, and genomic variants plays an important role in modern bioinformatic studies. Such studies are usually implemented in the form of computational pipelines which combine invoking bioinformatic programs (e.g., genome assemblers or gene prediction programs) with extra routines that convert input or output files of the programs between various bioinformatic data formats and query the produced files. Many bioinformatic data formats are based on genomic intervals [1], and thus querying files in such formats requires operating with the intervals.

*Methods and Algorithms:* We present pygenomics – an open-source Python package that provides routines for reading and writing bioinformatic data in various formats and operating with genomic intervals. Pygenomics is implemented in pure Python and does not require any other libraries except for the Python standard library. The package is developed according to the functional programming paradigm that ensures immutability of the package entities, absence of side effects in the package functions except for ones related to input-output (I/O), and extendable stream-based I/O. The package source code is provided with type hints [2] and a property-based testing framework that is based on the Hypothesis package [3]. Interval operations in the package are implemented using implicit intervals trees [4].

*Results:* Pygenomics implements reading and writing from a number of bioinformatic data formats, including BED, GTF, GFF3, SAM, and VCF. The package provides the application programming interface (API) and the command-line interface (CLI) for calling its routines from a source code or as stand-alone programs respectively. Implementation of pygenomics in pure Python allows to seamlessly incorporate the package routines into Snakemake [5] pipelines and to run them using CPython and PyPy interpreters. Absence of external dependencies, implementation in pure Python, and the property-based testing framework facilitate deployment of pygenomics to various computational platforms. Examples of Python scripts that use the pygenomics package routines are available in the public GitLab repository: [https://gitlab.com/gtamazian/pygenomics\\_examples](https://gitlab.com/gtamazian/pygenomics_examples).

*Conclusion:* Pygenomics provides routines that facilitate development of computational tools and pipelines for working with bioinformatic data. The routines are implemented in a transparent and robust way due to following the functional programming paradigm and utilizing features provided by the Python programming language and its standard library. The package with its source code is publicly available on GitLab: <https://gitlab.com/gtamazian/pygenomics>

*Acknowledgements:* This research was funded by RSF, grant 20-14-00072.

## References

1. Lawrence M., Huber W., Pagès H., et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9(8):e1003118. doi: 10.1371/journal.pcbi.1003118.
2. Van Rossum G., Jukka L., Langa Ł. PEP 484 – Type hints. Python Enhancement Proposals. 2014.
3. MacIver D.R., Hatfield-Dodds Z. Hypothesis: A new approach to property-based testing. *J Open Source Softw.* 2019;4(43):1891. doi: 10.21105/joss.01891.
4. Li H., Rong J. Bedtk: finding interval overlap with implicit interval tree. *Bioinformatics.* 2021;37(9):1315-1316. doi: 10.1093/bioinformatics/btaa827.
5. Köster J., Rahmann S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics.* 2012;28(19):2502-2502. doi: 10.1093/bioinformatics/bts480.

## Space-time correlation analysis in bio-macro-molecules, based on molecular dynamic trajectory data

Titov A.I.<sup>1,3\*</sup>, Shvetsov A.V.<sup>1,2,3</sup>, Isaev-Ivanov V.V.<sup>1</sup>

<sup>1</sup> Saint Petersburg Nuclear Physics Institute named by B.P. Konstantinov, NRC “Kurchatov Institute”, Gatchina, Russia

<sup>2</sup> Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

<sup>3</sup> NRC “Kurchatov Institute”, Moscow, Russia

\* [titov\\_ai@pnpi.nrcki.ru](mailto:titov_ai@pnpi.nrcki.ru)

**Key words:** molecular dynamics, molecular modelling, GROMACS, algorithms, phytase, escherichia coli

*Motivation and Aim:* Since the end of the last century, biologists state a question: is it possible to model processes that happen in living beings at the molecular level? With rapid development of computing technologies and appearance of super-computer clusters, this question becomes even more actual. It is a pity that current computing levels are not enough for real-time modelling, even a cell-sized system is most likely not possible to recreate. On the other hand, calculation of inter-molecular interactions between biological molecules in real solvent for reasonable time durations is possible for the current-gen computing systems. However, modelling is not be-all and end-all, data analysis is even more important than getting it, but methods are lacking behind in development – such as dynamic properties analysis and inner-molecular interaction analysis.

*Methods and Algorithms:* There are a few programming packages for molecular modelling. Few should be mentioned among many: GROMACS[1] and AmberTools[2]. We have selected the first package, as it is open-source and allows to use its code and libraries for creation of personal programs to analyze computed trajectories of molecular dynamics (MD). We have created a few algorithms, in particular “Time and space correlation analysis in bio-macro-molecules from MD trajectory data”[3]. This work is about its application.

*Results:* We have selected *phytase* protein from *Escherichia* for inner-molecular analysis, since it is widely analysed in literature. All possible mutations were computed for 12 potential positions. During correlation analysis, we figured out that in more thermally stable variants the connectivity net is more spread, with relatively less connection nodes.

*Conclusion:* The obtained results allow us to consider that the created method of bio-macro-molecule analysis helps finding a list of potential thermo-stabilising variants. Based on the current data we plan to run a few *in vitro* experiments to prove this hypothesis.

# Multi-class abdominal aortic aneurysm segmentation via 3D neural networks

Tuchinov B.<sup>1\*</sup>, Amelina E.<sup>1</sup>, Pnev S.<sup>1</sup>, Groza V.<sup>1</sup>, Nikitin N.<sup>2</sup>, Karpenko A.<sup>3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Medical Center "Avicenna", Novosibirsk, Russia

<sup>3</sup> Meshalkin National Medical Research Center, Novosibirsk, Russia

\* [bairt@nsu.ru](mailto:bairt@nsu.ru)

**Key words:** aorta, aneurysm, thromboembolism, medical imaging, machine learning

*Motivation and Aim:* In recent years, there has been an increase in the number of patients with atherosclerotic lesions of the aorta and its branches. The most common and clinically significant complications of the course of atherosclerosis are cholesterol and thrombotic emboli. Moreover, the presence of calcinates (calcium deposits in the aortic wall and thrombotic masses) significantly effect on the treatment of another serious disease – aortic aneurysm. Aortic aneurysm is a serious disease, the course of which is accompanied by many adverse consequences for the patient's health and often ends in rupture, leading in many patients to death. The detection rate of aortic aneurysm in different study cohorts ranges from 3.2 to 6 % of the population. Moreover, 98.1 % of the aneurysms were small. The multi-class abdominal aortic aneurysm segmentation task is one of the most challenging segmentation tasks. The aim of the current project is to develop diagnosis approach via deep learning techniques for assessment to aneurysmal lesions of the abdominal aorta.

*Methods and Algorithms:* The source of clinical cases (clinical base) is the Meshalkin National Medical Research Center (Novosibirsk, RF), which provides high-tech medical care (up to 20000 surgical operations per year). The dataset consists of 30 consecutive study participants who underwent clinically indicated computed tomography angiography (CTA) of the abdominal aorta and iliac arteries. Annotations comprise labels of the aortic lumen, thrombotic masses and calcifications. Segmentation labels were independently approved by three board-certified radiologists experienced in cardiovascular imaging and any label misclassification was manually corrected by an additional expert. All CTA examinations were performed by using a 320-row CT scanner (Aquilion ONE; Toshiba). All images were reconstructed with 1-mm section thickness and 0.8 mm increment by using a CTA-dedicated kernel and adaptive iterative dose reduction algorithm. We perform three different experiments according to the Ground Truth (GT) modification. Within the dataset, not all the presented samples have the equal available annotations. This led to the three following approaches: a) consider only GT from expert #1; b) consider the calcification GT only when 2 experts are agreed; c) consider the calcification GT only when all available experts are agreed. The aorta and thrombotic masses GT remain the same among all experiments.

Firstly, we use nnU-net [2] adapted to our datasets, as a basic architecture for comparison. It has proven to be a strong basic solution for segmenting medical images. In our case, we used the 3D version with a base number of 32 channels and a maximum number of channels in a bottleneck of 320. The model was trained in a 5-fold cross-validation mode for 50 epochs, with patch size of 96x160x160 and a default batch size

equal 2. As an augmentation to the input images were applied: rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma and mirroring. Another method that we used is the Modified\_3DU-net. This is less adaptive though a more classic and robust approach. In the basic principles it relies on the 3D U-net like network architecture. For this particular task we used the modification consisting of 4 descending encoder context pathway blocks, 1 intermediate block, and 4 decoder localization pathway blocks relatively. In this research we used the modification with a base number of 16 channels in the first encoder block and multiplier of 2 on each descending step. The model was trained in a 5-fold cross-validation mode for 500 epochs with the patches of 320x320x64 and batch size of 4. The patch size was selected of 64 in the Z direction due to the variability of given scans and for correct processing of all volumes. This model was also trained with the use of Cross-entropy loss function. As for the augmentations only Horizontal and Vertical flips were used.

*Results:*

Method	Dice_Aorta	Dice_Thromb	Dice_Calc
nnUnet_1Exp	0,955	0,765	0,710
nnUnet_2Exp	0,950	0,758	0,672
nnUnet_3Exp	0,951	0,755	0,665
Modified_3DU-net_1Exp	0,928	0,785	0,664
Modified_3DU-net_2Exp	0,873	0,764	0,545
Modified_3DU-net_3Exp	0,909	0,775	0,626

The table shows the results on validation set. We can observe, both methods have good results to segment the aortic lumen and thrombotic masses. One can observe that more strict conditions on the calcification such as agreement of all experts lead to slight reduction in the performance, while the minimum agreement of 2 experts leads to a strong drop in the accuracy. This might be explained the challenging task of the structure’s annotations and that the complete agreement within the experts lead to significant reduction in the calcification GT segmentation zones. It becomes very small and so difficult for the processing. The complexity of the task is also due to the fact that usually calcinates are analyzed using a non-contrast CT. But both studies require separate image acquisitions and thus carry radiation burden. We sought to develop and validate a method to segment calcinates using only CT angiography.

*Conclusion:* The obtained results will make it possible to develop a comprehensive methodology diagnosis of rupture and the occurrence of thromboembolic complications in abdominal aortic aneurysm. One of the most important ways to improve the quality of the proposed methods is to significantly increase and create the reliable training dataset, such that the obtained result could be improved for robustness and stability.

*Acknowledgements:* The study is supported by the Russian Science Foundation grant No. 21-15-00091.

*References*

1. Isensee F. et al. nnU-net for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. Springer, Cham, 2020;118-132.

## Argo\_CEL: GPU based composite elements discovery

Vishnevsky O.<sup>1,2\*</sup>, Bocharnikov A.<sup>2</sup>, Ignatieva E.<sup>1,2</sup>, Kolchanov N.<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* *oleg@bionet.nsc.ru*

**Key words:** motif discovery, composite elements, transcription regulation, ChIP-Seq

*Motivation and Aim:* Composite elements play an important role in the regulation of transcription and provide a way for crosstalk between different regulatory pathways. Existing methods for the revealing of potential composite elements (PCE) are usually based on assessment of the significance of the mutual presence of the predicted transcription factor binding sites (TFBS) or ChIP-Seq pikes. Thus, such methods essentially depend on the completeness of training samples and experimental information about localization of TFBS. Methods for *de novo* discovery of significant oligonucleotide motifs are rather fast and require neither multiple alignments of query sequences nor the information about localization of transcription factor binding sites. However, existing heuristic approaches do not guarantee discovery of the most represented and significant motifs in huge experimental data sets. We developed new GPU based exhaustive approach ARGO\_CEL for potential composite elements discovery in large DNA datasets.

*Methods and Algorithms:* Total scheme of ARGO\_CEL approach is:

1. *De novo* discovery of the significant oligonucleotide IUPAC motifs specific for the set of regulatory sequences. The degenerate oligonucleotide motif, which was obtained as a result of this procedure, is considered significant if it meets the following criteria:

$$\begin{cases} F > f_0 \\ Q > q_0 \\ P(n, N) > p_0 \end{cases}$$

F is the proportion of query sequences containing the motif; Q denotes the enrichment of the motif expected by chance; P(n,N) is the binomial probability of observing by chance the motif in n out of N analyzed sequences;  $f_0$ ,  $q_0$ ,  $p_0$  are the threshold levels.

2. Clustering of the revealed motifs. Clusters include the motifs significantly mutually appearing by chance in a set of random sequences. Estimation of correlation of all revealed motifs to each other on the set of random sequences was performed using phi-coefficient of Pearson for binary variables.

3. Revealing in the analyzed set of regulatory sequences significantly correlated clusters of motifs not correlated by chance.

4. Classification of the clusters with Tomtom. The assessment of similarity with the known TFBS was carried out in the JASPAR and HOMOCOMO databases. When annotating, a significant similarity with the weight matrices of several TF binding sites was normally observed for each oligonucleotide motif revealed. In this case, if JASPAR or HOMOCOMO annotations showed similarity with the weight matrix of the major TF, for which the experiment was carried out, then we considered that this oligonucleotide motif corresponds to the binding site of this TF. Otherwise, we identified PWMs that

had a significant similarity to the annotated motif in both JASPAR and HOMOCOMO databases.

**Results:** In order to carry out a comparative assessment of the efficiency of the developed motif discovery system, test sets of randomly generated sequences were created. In these sets, random IUPAC motifs of a given level of degeneracy were placed at random positions and in a given part of the sequences. Average information content (AIC) was used to measure the level of degeneracy of the motifs. Well-known motif discovery programs MEME-ChIP and CisFinder have been used to reveal embedded motifs. It was demonstrated that as soon as the abundance,  $F$ , of motifs falls below 30 %, the degeneracy of motifs increases, and AIC decreases below 1.5, the number of sequences  $N$  decreases and  $pr$  (probability, of observing by chance motifs embedded in the set) becomes higher than 0.05, the efficiency of all programs decreases abruptly. Nevertheless, as  $pr$  falls below 0.05, Argo\_CEL still accurately identifies motifs, while all the other methods mentioned do not, especially at  $AIC = 1.25$ , – their outputs are significantly different from the original motifs embedded in the sets. It is most clearly observed at  $F < 30$  %. Thus, it becomes necessary to use Argo\_CEL to identify poorly abundant and highly degenerate context signals. In [1], we showed that  $\sim 70$  % of weight matrices from TRANSFAC and Jaspas have  $AIC \leq 1.25$ . Thus, we have proven that Argo\_CEL can be more efficient than some other approaches in *de novo* discovery of about half of the TFBS, especially if they are present in less than 30 % of the sequences of the DNA set. We applied Argo\_CEL system, to detect *de novo* conserved motifs in the sets of DNA sequences obtained during ChIP-Seq experiments with three transcription factors: FoxA2, Sp1, and GATA1 downloaded from the Cistrome DB [2]. For each of the ChIP-Seq experiments, 50 most significant motifs were annotated using Tomtom. In each of the considered ChIP-Seq experiments, for the annotations of the 50 most significant motifs obtained in this way, a search was made in scientific publications. For the vast majority of the considered motifs, experimental confirmation of the synergistic interaction of predicted TFs with target TFs was found.

**Conclusion:** We developed a new method ARGO\_CEL for revealing of correlated binding sites in the ChIP-seq data, based on *de novo* motif discovery approach. This is an exhaustive method based on the use of high-performance GPU technologies. In comparison with existing methods of *de novo* motif discovery, it demonstrates higher efficiency, especially for strongly degenerate and underrepresented motifs. We have proven that Argo\_CEL can be more efficient than some other approaches for *de novo* discovery of about half of the TFBS, especially if they are present in less than 30 % of the sequences of the DNA set. These correlated binding sites may correspond to sites – partners of transcription factors that take part in coordinated regulation of expression. The use of ARGO\_CEL for the analysis of ChIP-Seq data made it possible both to identify potential binding sites for previously experimentally confirmed transcription factors – partners, and to propose motifs – targets for further experimental analysis.

**Acknowledgements:** The study is supported by the budget project of Institute of Cytology and Genetics, SB RAS (FWNR-2022-0020).

## References

1. Vishnevsky O.V., Bocharnikov A.V., Kolchanov N.A. ARGO\_CUDA: Exhaustive GPU based approach for motif discovery in large DNA datasets. *J Bioinformatics Computation Biol.* 2018;16(1):1740012.
2. Mei S., Qin Q., Wu Q. et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.* 2017;45(D1):D658-D662.

# Mathematical modeling of information diffusion in Twitter

Zvonareva T.<sup>1\*</sup>, Krivorotko O.<sup>1,2</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia

\* t.zvonareva@g.nsu.ru

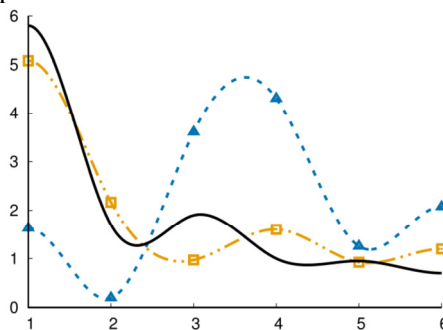
**Key words:** diffusive logistic model, source problem, inverse problem, data analysis, optimization, regularization, tensor train, multilevel gradient method, high performance computing

*Motivation and Aim:* The social media allow for the rapid diffusion of various information around the world and this information is not always neutral. Such processes could be described with the mean-field games that consists in Kolmogorov-Fokker-Planck (KFP) and Hamilton-Jacobi-Bellman (HJB) equations [1]. KFP characterizes the information diffusion in social media as well as HJB describes the control of such diffusion. In this work we investigate the information diffusion in Twitter represented by the simplified KFP equation which is characterized by its coefficients and the initial condition [2]. These parameters describe the main features of information diffusion in a social network. The aim of the work is to clarify the source of information propagation for formulation the control problem of such diffusion [3].

*Methods and Algorithms:* The problem of determining the initial condition is reduced to the problem of minimization of the misfit function (inverse problem) which was solved by the global, local and hybrid optimization methods:

- particle swarm optimization (PSO) is a stochastic method for finding the global minimum region;
- multilevel gradient method (MGM) [4] is a modification of the local gradient descent method which has a higher rate of convergence;
- tensor train (TT) [5] is a global optimization method based on tensor decomposition;
- combination of TT and MGM as a hybrid optimization approach.

*Results:* For the benchmark problem the source (solid black line) of information propagation was specified from additional information about the process at 21 points in time for each hour. The figure shows the solutions of the inverse problem using combinations of PSO + MGM (blue dashed line with triangles) and TT + MGM (orange dashed line with squares). The x-axis is the distance from the source measured in the minimum number of links between the source and the user. The y-axis shows the density of users involved in the process.





*Conclusion:* For additional information at different points in time without spatial separation the solution of the source problem is nonunique and unstable. The TT+MGM combination could be found the solution closer to the exact one in sense of minimization of the relative accuracy of model and benchmark solutions of inverse problem than the PSO + MGM. However, TT + MGM requires a large computational cost, i.e., to obtain an accuracy of about  $10^{-2}$  it requires 17 hours of computation on the SSCC SB RAS node with 2 Intel Xeon Gold 6248R processors (3 GHz, 24 cores) and 384 GB of main memory. This algorithm will be tested for popular news on Twitter in order to analyze social influence.

*Acknowledgements:* The study is supported by Russian Science Foundation (project No. 18-71-10044-II).

### *References*

1. Bauso D., Tembine H., Başar T. Opinion Dynamics in Social Networks through Mean-Field Games. *SIAM J Control Optim.* 2016;54(6):3225-3257.
2. Wang F., Wang H., Xu K., Wu J., Jia X. Characterizing Information Diffusion in Online Social Networks with Linear Diffusive Model. *Proc ICDCS.* 2013;307-316.
3. Krivorotko O., Zvonareva T., Zyatkov N. Numerical solution of the inverse problem for diffusion-logistic model arising in online social networks. *Commun Comput Info Sci.* 2021;1476:444-459.
4. Zvonareva T.A., Krivorot'ko O.I. Comparative analysis of gradient methods for source identification in a diffusion-logistic model. *Zh Vychisl Ma. Mat Fiz.* 2022;62(4):694-704.
5. Oseledets I.V. Tensor-train decomposition. *SIAM J Sci Comput.* 2011;33(5):2295-2317.

# Author index

- Abakumov A. 238  
Abasov R. 385  
Abdikarim G.M. 755  
Abdulkhakov S. 499  
Abramov S. 1107  
Abramova M.Yu. 386  
Abramovich A.V. 549  
Abushakhmina A. 932  
Adameyko K. 108  
Adamovskaya A.V. 191, 206  
Adelshin R. 677  
Adonina S. 926  
Afanasenko O. 640  
Afonnikov D.A. 50, 57, 154, 182,  
593, 609, 615, 617, 640, 1094  
Afonnikova S.D. 510  
Afonyushkin V.N. 883  
Aftanas L.I. 961  
Agapov A. 983  
Agapova Y. 335  
Aglyamova A. 22  
Ahmetov I.I. 392  
Aiden E. 113  
Akberdin I.R. 182, 211, 214,  
234, 573, 584, 883, 903  
Akhatov A.M. 392  
Akhatov I. 579  
Akhberova N. 858  
Akhmetzianova L.U. 24  
Akimkin V. 518, 526, 845  
Akimniyazova A.N. 755, 880  
Akimov S.A. 288  
Akinshin A. 1096  
Akmacic I.T. 474  
Akopyan A.A. 727, 731, 732, 736, 744, 751,  
798  
Akulov A. 194, 686  
Alatortsev V.E. 1062  
Alborova I.E. 482  
Aleksandrovich V. 594  
Alekseenko L.L. 217  
Alekseeva M.G. 507  
Alemasova E. 969, 1008  
Aleoshin V.V. 697  
Alexeevski A. 158, 162, 184, 313, 576  
Alharbi M. 362  
Ali Ashfaq U. 362  
Aliev A.M. 378  
Aliev T. 757  
Aliper E.T. 267  
Alsakaji H.J. 895  
Alsalloum M. 928  
Alshammari A. 362  
Altukhov D. 335  
Amelin M.E. 804  
Amelina E.V. 804, 1124  
Amstislavskaya T.G. 729, 731,  
732, 733, 735, 736, 747, 751  
Ananeva A. 379  
Anarbaev R. 1008  
Anashkina A.A. 160, 263, 269, 271, 375  
Andreenkova O.V. 146  
Andreeva E.A. 427  
Andreeva I.N. 494  
Andrianova N.V. 1065  
Anikin D. 227  
Anisimov V.N. 1029  
Anisimova M.V. 668  
Anisimova N. 26  
Anokhin K. 956  
Antezana M.A. 1083  
Antonenko A. 1030  
Antonets D.V. 40, 62, 1113, 1115  
Antonets K.S. 172, 543, 626  
Antonov E.V. 491, 513, 557, 570  
Antonov I. 87  
Antonova O. 640  
Antontseva E. 757, 777  
Antropov D. 841  
Antropova E.A. 206, 759  
Anufriev K.E. 560  
Anufrieva E. 694  
Arapidi G. 911  
Aravin A.A. 983  
Arbatsky M. 860  
Arefieva N. 866

Aripova A. 761  
Arrasate P. 274  
Arseniev A.S. 275, 286, 287, 291, 293,  
300, 303, 311  
Artemenko N.V. 1085  
Artemieva L.E. 287  
Artemyeva A. 26  
Artomov M. 469  
Artykova A. 520  
Artyomov M. 99, 469  
Artyushin I.V. 892  
Askari V. 787  
Aslam S. 362  
Atabiev B. 481  
Ateiah M. 355  
Aulchenko Y.S. 464, 467, 470, 474, 476  
Axenovich T.I. 464, 473, 478, 710  
Ayginin A.A. 348, 428, 867  
Ayriyants K. 715, 929  
Ayupova N. 1096  
Azhar I. 789

**B**  
Babak O. 26  
Babenko R. 28, 678  
Babenko V.N. 28, 54  
Babenko V.V. 216  
Babich T.L. 511, 537  
Babochkina T.I. 668  
Baboshin M.A. 544, 551  
Babovskaya A. 453  
Babusenko E.S. 566  
Babushkina N.P. 381, 383  
Bagautdinova Z.Z. 652, 658  
Bagirov V.A. 710  
Bagrov D.V. 327  
Baiandina Iu. 29  
Bakman A. 970  
Bakulina A. 331  
Bakunova A. 273  
Balabaev N.K. 298  
Baltin S. 665  
Baneh H. 688  
Bankin M. 595, 623, 642  
Bannikova S. 491, 546, 555, 557, 559  
Baranov V.S. 417  
Baranovskiy A.G. 1004  
Barberis M. 192  
Barbitoff Y.A. 85, 385, 404, 407  
Baricheva E. 712  
Bartsev S. 239, 252  
Barysheva E. 415  
Bashirzade A.A. 731, 732  
Bashkirov P. 274  
Basu A. 869  
Batlutskaya I.V. 386  
Battulin N. 115, 120

Bauer T.V. 62  
Baulin E.F. 313  
Baulin E. 343  
Baydin G.V. 921  
Bazarevich M. 111  
Bazarov A. 1086  
Bazhenov S.A. 705  
Bazhenova E. 926  
Bazovkina D. 926, 944  
Beilin A.K. 768  
Beissert S. 1007  
Bekeneva K. 1037  
Beklemisheva V.R. 113, 137  
Beldiia E. 46  
Beletsky A. 520  
Belichenko V.M. 731, 732, 744  
Belikov S. 163  
Beljanski M. 888  
Belokopytova P.S. 117, 129, 134, 440  
Belonogova N.M. 464, 473, 478  
Belousova E.A. 998, 1023  
Belova I.N. 244  
Belova M.G. 921  
Belykh A. 398  
Berdyugina D.A. 747  
Berestovskaya J.J. 530  
Berezhnaya A. 31  
Berezina O. 103  
Berman D.I. 703  
Bergardt V. 49  
Bershatsky Y. 275, 277  
Bersimbaev R. 761  
Beskrovnaya M. 983  
Besova E.V. 921  
Bessonova D. 407  
Bezdvornyykh I. 762  
Bezsudnova E. 273, 321  
Bezuglov V. 68  
Bgatova N.P. 764, 798, 807, 821  
Bhaumik P. 1009  
Bhavya B. 873  
Bichevaya N. 390  
Bidzhieva S. 537  
Bikhurina T.I. 670  
Bilyalov A. 48  
Birnbaum R. 124  
Birykov M. 992  
Biziukova N. 341  
Blinov A. 557, 597  
Bobyleva L. 415  
Bocharnikov A. 101, 1126  
Bocharov A. 931  
Bocharov E. 275, 277, 279, 311, 333, 335  
Bocharov G. 212, 589, 765, 871  
Bocharova A. 388  
Bochkov D. 491, 557

Bogachev S.S. 979, 1014  
Bogacheva A.I. 348  
Bogacheva N. 491, 557  
Bogdanov M. 858  
Bogomolov A.G. 655, 689, 825, 863  
Bogomyakova O. 194  
Bograya M. 193  
Bohdan D. 343  
Boichenko S. 970  
Boiko A. 785  
Bokov R. 84  
Boldinova E.O. 1004, 1026  
Boldyreva L.V. 721  
Bolshakov A. 1076  
Bolshakova E. 390  
Bolsheva N.L. 33  
Bondar N. 42, 715, 929  
Bondarenko N.A. 802  
Bondaryuk A.N. 139  
Bondiuk S. 1119  
Borisova M.A. 946  
Borisova N.V. 936  
Börner A. 615  
Borodin P.M. 710  
Borshchevskiy V.I. 262, 265, 294, 300,  
302, 303, 309, 337  
Bose K. 1009  
Bosov D. 127  
Boulygina E.A. 392  
Boyarchenkov A.S. 221, 223  
Boyko K. 273, 321  
Boytsov A. 1107  
Bragina E.Yu. 381, 383, 394  
Bratus A. 765  
Britikov V. 277  
Britikova E. 277, 333  
Brodt K. 1086  
Brovkina O. 396  
Bryanskaya A.V. 491, 513, 546, 555,  
557, 559, 568, 570  
Bryzgalov L. 466  
Budkina A. 87  
Bukharina T.A. 201  
Bukin S. 514, 540  
Bukin Y.S. 139, 677, 697, 866  
Bulaev A. 520, 523  
Bulakhova N.A. 703  
Bulavin I. 35, 644  
Bulgakov N. 1026  
Bulgakova O. 761  
Bulusheva I.A. 867  
Bulyakova T. 956  
Bulygin A.A. 971  
Buneva V. 785  
Burkatovskii D. 337  
Burnyasheva A.O. 1032, 1066  
Bushueva O. 398, 400, 415, 421  
Buslaev V. 402  
Busov I. 1087  
Butina T.V. 139  
Bychkov I. 442  
Bychkov M.L. 319  
Bystrova O. 1046  
Cai G. 425  
Cao X. 648  
Chabina A.S. 217  
Chadaeva I. 432, 689, 825, 863  
Chagarov O. 481  
Chaley M. 896  
Chamo M. 765  
Changalidi A.I. 404  
Chechetkina S.A. 438, 440  
Chechyotkina O.Yu. 482  
Chekanov N. 407  
Chelebieva E. 64, 644  
Chemeris A.V. 24, 612  
Chen M. 766  
Chen W.-L. 425  
Chepurnov G. 597  
Cheraghi S.F. 323  
Cherdantsev S.V. 486  
Cherenko V. 841  
Cherevko A. 194  
Cherezov V. 265, 294, 302  
Cherkashina O.L. 768  
Cherkasov A. 71, 92, 111  
Cherkasov N. 1121  
Chermnykh E.S. 768  
Chernitsyna S. 540  
Chernogor L. 163  
Chernushkin D. 562  
Chernyak B.V. 1054  
Chernykh V.V. 802  
Chernyshkov A. 518, 526  
Chernyshova E.V. 1065  
Chernyshova I.A. 972  
Chesnokov D. 557  
Chesnokov Y. 319  
Chicherina G.S. 180  
Chistyakov D.V. 1065  
Chivileva A.S. 440  
Chizhova N.D. 729, 733  
Chugunov A.O. 319  
Chupakhin A.P. 229, 236  
Churkina M. 103  
Churkina T.V. 419, 452  
Chuvakova L.S. 965  
Chvileva A. 438  
Cornette R. 92

Daly M. 469  
 Damarov I. 466  
 Danilchenko V.Yu. 770, 771  
 Danilenko A. 331  
 Danilenko K.V. 961  
 Danilenko N. 594  
 Danilenko V.N. 490, 493, 496, 500, 507, 564  
 Danilov L.G. 176, 448  
 Danilova A.L. 405, 436, 450, 459  
 Danilova S.S. 918  
 Das B.B. 974  
 Das R. 773, 873  
 Dashkevich G.E. 939  
 Daugavet M. 97, 140  
 Davletgildeeva A.T. 975, 976  
 Davletkulov T.M. 24  
 Davletshina G.I. 167, 179  
 Davydenko K. 442  
 Davydenko O. 594  
 Davydova A. 775  
 Davydova Yu. 932, 937  
 Dedkov V.G. 892  
 Dedysh S.N. 516  
 Deeva A.A. 281, 329  
 Degermendzhi A. 239  
 Degtyareva A. 757, 777  
 Deineko E. 657  
 Demakova V. 407  
 Demchenko G. 1037  
 Demchenko K.N. 651  
 Demeneva V. 457  
 Demenkov P.S. 191, 196, 206, 220,  
 225, 231, 575, 759, 790, 877  
 Dencher N.A. 337  
 Denikina N.N. 139  
 Denisov E. 437  
 Derbeneva A.S. 801, 1034  
 Derevianchenko S. 196  
 Dergilev A.I. 44, 197  
 Deryusheva E. 156, 779, 818  
 Desforges B. 1019  
 Deviatiiarov R.M. 71, 92, 573, 780  
 Devyatkin V.A. 801  
 Dey P. 75  
 Deykin A. 398  
 Diatlova E. 977  
 Digris A. 106  
 Dikalov S.I. 1060  
 Dikaya V. 70, 140  
 Djordjevic Magdalena 897, 899, 916  
 Djordjevic Marko 897, 899, 916  
 Dmitrenko O. 796  
 Dmitriev A. 360  
 Dmitriev A.A. 33, 60, 631, 633  
 Dmitriev S. 1047  
 Dmitrieva D. 283  
 Dmitrieva E.M. 958  
 Dobrovolskaya M.V. 480  
 Dobryakova Yu. 1076  
 Dolgikh V. 599  
 Dolgova A.S. 892  
 Dolgova E.V. 979  
 Dolinnaya N. 1011  
 Dome A. 841  
 Dontsov V. 1048  
 Dorofeev A.G. 530  
 Dorogova N. 712  
 Drachkova I.A. 638  
 Drenichev M.S. 972  
 Drozd E. 26  
 Drozdova P. 169, 188  
 Dubenko T. 665  
 Dubey A. 367, 370  
 Dubrovina N.I. 727, 731, 732, 735,  
 736, 744, 751  
 Duk M.A. 600  
 Dutta S. 1009  
 Duvalov E. 106  
 Dvoinishnikov S.V. 548  
 Dvorianinova E.M. 33, 60, 631  
 Dvornikova A. 1026  
 Dyachenko A.I. 696  
 Dyachkova M.S. 505  
 Dymova M. 819  
 Dyrkheeva N.S. 972, 981  
 Dyuzhikova N.A. 58, 722  
 Dzhaubermezov M. 481  
 Dzhioev Y.P. 866  
 Efimenko A.Yu. 1052  
 Efimov K.V. 142, 874  
 Efimov V.M. 142, 615, 874  
 Efimova I. 856  
 Efimova O. 111  
 Efimova O.A. 417  
 Efremov A. 174  
 Efremov R.G. 267, 275, 285, 311, 333  
 Efremova A.D. 781  
 Efremova E. 665  
 Egorov V.V. 878  
 Egorova A. 518, 526  
 Egorova A.A. 606  
 Egozova E.S. 966  
 Ekomasova N. 481  
 Elatkin N. 688  
 El-Deeb A. 355  
 Elgaeva E.E. 467, 476  
 Elisafenko E.A. 144, 602  
 Elistratova M.G. 199, 232, 773  
 Elkina Yu. 520

- Emirsaliiev A. 35  
 Endutkin A.V. 982  
 Engovatova A.V. 482  
 Enikeeva R. 932, 937  
 Epifanov R.Yu. 1085  
 Epimakhova E.V. 783  
 Eremenko E. 1072  
 Eremin D.V. 928, 934, 964  
 Ermakov E.A. 783, 785  
 Ermakov M.S. 839  
 Eroshenko D. 977  
 Ershov A. 521, 537  
 Ershov N.I. 52, 434, 678  
 Ershova A. 576  
 Esipova N.G. 263  
 Eslami G. 787  
 Esyunina D. 983  
 Evgen'ev M.B. 965, 1033  
 Evsiukova V.S. 683  
 Evtushenko E.V. 144, 602  
 Ezhikeeva S.D. 749
- Fadeev S.I. 211  
 Farhi E. 124  
 Fartushny E.N. 1103  
 Fateev D. 26  
 Fedorov A.I. 450, 459  
 Fedorov D. 1089  
 Fedorova O.S. 970, 971, 975, 976,  
 988, 1001, 1018, 1022  
 Fedorova S. 712  
 Fedoseeva L.A. 713, 732  
 Fedotova A. 49  
 Feng C. 766  
 Feofanov A. 1005  
 Feofanova N.A. 668  
 Ferguson-Smith M.A. 167, 179  
 Filatova A. 407, 442  
 Filatova S. 457  
 Filimonov A.S. 990  
 Filimonov D. 341, 345  
 Filipenko M.L. 452  
 Firsov A. 1090  
 Fishman V.S. 117, 118, 122, 123, 129,  
 131, 134, 148, 407, 411  
 Fofanov M.V. 113  
 Fomin E. 604  
 Fomin I.N. 606  
 Fomina T. 604  
 Franco M.L. 291  
 Freidin M.B. 394, 464, 467  
 Frisman E.Ya. 243, 246, 255  
 Frolov P.V. 253, 259  
 Frolov V. 274  
 Frykin R.I. 631  
 Fu X.G. 423
- Furman D.P. 201  
 Fursova A.Zh. 801, 1034, 1077
- Gabidullina L. 481  
 Gabidullina T. 453  
 Gadzhi A. 694  
 Gagarinskaya D.I. 1004  
 Gaifullina R. 379  
 Gainova I. 876  
 Galachyants Y. 528  
 Galeeva J. 1089  
 Galyamina A.G. 738  
 Galytsyna A. 135  
 Galzitskaya O.V. 156, 298  
 Gam A. 1092  
 Gangapshev A. 539  
 Gaponov Y. 335  
 Garafutdinov R.R. 24  
 Garbuz D.G. 1033  
 Gatskaya S. 602  
 Gatupov M. 442  
 Gavrilenko M. 453  
 Gavrilenko T.A. 609, 640  
 Gaysler E. 877  
 Gazizova G.R. 48, 71, 84, 92, 829  
 Gelfand M. 133  
 Genaev M.A. 50, 593, 609, 615, 617,  
 1085, 1094  
 Generozov E.V. 392  
 Gentzbittel L. 607, 688  
 Gerashchenko T. 437  
 Gerasimov A. 302  
 Gerasimova S.V. 606, 616  
 Gerasimova T.V. 560, 590  
 Geraskina O. 1005  
 Gerlinskaya L.A. 668, 674  
 Gibadullina S. 327  
 Gierke P. 615  
 Gild E.-Y.V. 735  
 Ginolhac A. 1007  
 Giovannotti M. 179  
 Gladyshev V. 1047  
 Glagoleva A. 640  
 Glotov A.S. 385, 404  
 Glukhov G.S. 327  
 Golikova P. 409, 459  
 Golubev D. 1049  
 Golubeva L.I. 544, 551  
 Golubeva T. 38  
 Golubyatnikov V. 1095, 1096  
 Golushko S.K. 804  
 Golyshev S. 127  
 Goncharov N.P. 597  
 Goncharov V.I. 354  
 Goncharova I.A. 381, 383, 394  
 Goncharuk M.V. 286, 300, 303

- Goncharuk S.A. 286, 287, 291, 293, 300, 303  
 Gorban' E.N. 1035  
 Gorbenko D. 355  
 Gorbenko I.V. 56, 608  
 Gorchakov V. 1037  
 Gorchakova O. 1037  
 Gordeliy V. 302  
 Gorelov S.V. 878  
 Gorina Ya.V. 1070  
 Gorshkova T. 22, 37, 654, 660  
 Goryachkovskaya T.N. 491, 513, 546, 554,  
 555, 557, 559, 568, 570  
 Govorun V. 911  
 Grafaskaia E. 359  
 Graphodatsky A.S. 113, 137  
 Grebennikov D. 212, 871  
 Grebennikova O. 35, 644  
 Gretsova M. 140  
 Gridina M.M. 118, 120, 123, 131, 411  
 Grigoreva E.A. 630, 688  
 Grigoreva T.P. 450  
 Grigorieva O.A. 1052  
 Grigoryeva T. 499, 534  
 Gromyko D. 73  
 Grouzdev D. 521, 537  
 Groza V.V. 804, 1124  
 Gruntenko N.E. 146  
 Gruzdev E. 520  
 Gubanova N.V. 197  
 Gubaydullin I.M. 24, 612  
 Guber A. 1097  
 Gugin P. 992  
 Gul R. 789  
 Gulevich R.G. 690  
 Gulyaeva N. 1076  
 Gunbin K.V. 182  
 Guo X.Y. 423  
 Gurazheva A. 103  
 Gureev A.P. 1065  
 Gurevich A. 357  
 Gurinova E. 409  
 Gursky V.V. 217, 685  
 Gurtovoy D. 398  
 Gusach A.Yu. 262, 265, 294, 302  
 Gusev O.A. 48, 71, 84, 92, 379, 573, 780, 829  
 Guseva E.D. 651  
 Gushchin I. 296, 302  
 Gutnik D. 163  
 Guzenko E.V. 493  
  
**Halder R.** 1007  
 Hantemirova M. 419  
 Hapaev R.S. 849  
 Harina O.K. 427  
 Harshan S. 75  
  
 Heidari M.M. 305, 323  
 Herbeck Yu.E. 690  
 Hillert-Richter L. 989  
 Hofestädt R. 202  
 Hong S.-J. 826  
 Hussain K. 789  
  
 Ibadullaeva E. 644  
 Ibragimova S. 645  
 Ibragimova S.M. 609, 640  
 Ignatieva E.V. 101, 204, 1126  
 Ignatov A. 177  
 Igonin V. 535  
 Igoshin A.V. 692  
 Ilchibaeva T. 928, 935, 964  
 Ilic B. 899, 916  
 Il'icheva I.A. 160  
 Ilina E.L. 651  
 Ilina E.N. 216, 851, 1089  
 Ilina E.S. 984  
 Ilinskaya A. 484  
 Ilinsky V. 484  
 Illarioshkin S.N. 564  
 Ilyasov R.A. 490, 493  
 Ilyinsky V.V. 55  
 Imanmalik Y. 407  
 Iolchiev B.S. 710  
 Isaev-Ivanov V.V. 878, 1123  
 Ishchenko A.A. 975  
 Ishchenko I.Yu. 810, 811  
 Ishtiaq A. 789  
 Itskovich V. 186  
 Ivanchenko M. 742, 1039, 1043  
 Ivanisenko N.V. 206, 290, 986, 989  
 Ivanisenko T.V. 206, 790, 877  
 Ivanisenko V.A. 191, 206, 220, 231, 575,  
 759, 766, 790, 812, 877, 986  
 Ivankov O. 339  
 Ivanoshchuk D. 430  
 Ivanov A. 38, 952  
 Ivanov G.A. 972  
 Ivanov N.V. 39  
 Ivanov R. 792  
 Ivanov S.M. 341, 360, 364  
 Ivanov V. 540  
 Ivanov V.P. 400  
 Ivanova N. 35  
 Ivanova S.A. 783, 785, 958, 962  
 Ivanova V. 125, 502  
 Ivashchenko A.T. 755, 880  
 Ivashkina O. 956  
  
**Jasinska W.** 1072  
 Jiao Y. 648  
 Joshi P. 1079

Kabanikhin S. 909  
Kabardin I.K. 548  
Kabieva S.S. 79  
Kabilov M. 775  
Kabirowa E. 115  
Kacher Yu. 327  
Kalabusheva E.P. 768  
Kalinin A.V. 549, 550  
Kalinin V. 208  
Kalinina T.I. 560, 590  
Kallistova A.Yu. 532  
Kalmykov S.N. 511  
Kalueff A.V. 729  
Kaluski S. 1072  
Kalyakulina A. 742  
Kalyuzhnaya M. 584  
Kamanova E.P. 40, 62  
Kaminskaya Y. 935  
Kaminskiy G. 901  
Kamynina A. 311, 337  
Kan V.A. 248, 257  
Kanapin A.A. 595, 600, 623, 642, 762, 1121  
Kantún N. 344  
Kaplan V.S. 513  
Kapunac S. 888  
Kapushchak Ya. 799, 862  
Karagodin D. 712  
Karagyaur M. 956, 1016  
Karamysheva T.V. 672  
Karasev D. 345  
Kardymon O.L. 290, 407  
Karetnikov D.I. 609, 621  
Karim S. 789  
Karlova M.G. 325, 327  
Karogodina T.Yu. 793, 815  
Karpeeva E. 193  
Karpenko A. 1124  
Karpenko D.V. 1099, 1101  
Karpova A. 936, 952, 967  
Karpova N. 794, 796  
Karsen L. 470  
Kartavtseva I.V. 670  
Karyagina A. 174, 184, 313, 576  
Karzhaev D.S. 630  
Katraeva I. 524  
Kavai-Ool U. 419, 952  
Kawakami Y. 455  
Kazakov A. 818  
Kazantsev F.V. 211, 218, 220, 225,  
227, 578, 952  
Kazantseva A. 932, 937  
Kazantseva D.V. 783, 962  
Kazantseva E.S. 244  
Kazhevarova A. 457  
Kechko O.I. 375  
Kelbin V.N. 610  
Khabarova A.A. 115, 148, 438, 440  
Khabudaev K.V. 892  
Khalid H.R. 362  
Khalyavkin A. 1048  
Khan M. 789  
Khan R. 113  
Khandaev B. 432, 825, 863  
Khantakova J.N. 715  
Kharchenko V.E. 649  
Kharkov V. 413  
Kharlampieva D. 911  
Khatami M. 305, 323  
Khe A. 194  
Khilazheva E.D. 1070  
Khisametdinov M. 537  
Khiutti A. 640  
Khlebodarova T.M. 211, 546, 555, 560  
Khlestkina E.K. 616  
Khlystov O. 514  
Khobta A. 987  
Khodyreva S.N. 984, 998  
Khokhlov A.N. 1040, 1057  
Khomenko T. 733  
Khorn P.A. 262, 265, 302  
Khotskin N.V. 934  
Khotskina A. 686  
Khozyainova A. 437  
Khrameeva E. 111, 124, 133, 135, 1047, 1072  
Khristichenko M. 212  
Khrunin A.V. 378  
Khurmuzakiy A.A. 263  
Khurshid M. 362  
Khusainov G. 302  
Khusnutdinova E.K. 427, 481, 932, 937  
Khvorykh G.V. 378  
Kikawada T. 92  
Kilchevsky A.V. 26, 461, 493, 494  
Kim A. 484  
Kim D.V. 987, 1026  
Kim I.I. 802  
Kireev I. 127  
Kirichenko A.V. 464, 473, 478  
Kirillov B. 579  
Kirillova N. 1096  
Kirpichnikov M.P. 317, 319  
Kiryanova O. 612  
Kiryushkin A.S. 651  
Kisaretova P. 42  
Kiselev I.N. 214, 234, 903  
Kiseleva A.A. 31, 614, 635  
Kiseleva E.V. 349, 438, 555, 681, 854  
Kisselevskaya-Babinina V. 905  
Kiyamova R. 858  
Kladova O.A. 988



Klemeshova D.I. 939  
Klimenko A.I. 146, 225, 241, 498, 578, 580,  
582, 952  
Klimenkov I.V. 749, 941  
Klimina K.M. 500, 505, 1089  
Klimontov V.V. 766, 837, 890  
Klimova N. 689, 825  
Klimovich P.S. 860, 956, 1016  
Klyuchereva A.A. 690  
Klyushnik T.P. 744  
Knyazev G. 931  
Knyazheva M.A. 740  
Kobalo N. 882  
Kobzeva K. 398, 415, 421  
Kocharovskaya M.V. 317, 319  
Kochetkova A.S. 984  
Kochetov A.V. 609, 621, 640  
Kochurova A. 46  
Kohchi T. 645  
Koksharova G. 148  
Kolchanov N.A. 101, 182, 206, 220, 241, 432,  
560, 590, 638, 689, 790, 825, 863, 1126  
Kolesnikova M. 929  
Kolmogorov Yu. 1037  
Kolmykov S.K. 77, 94, 573, 584  
Kolodkin A.N. 1042  
Kolokoltsov V. 907  
Kolomiets E.I. 490  
Koloshina K.A. 606  
Kolosov A. 520, 523  
Kolossova N. 331  
Kolossova N.G. 717, 801, 825, 1034, 1066,  
1074, 1077  
Kolossovskaya E.V. 616  
Kolpakov F.A. 77, 84, 94, 214, 234, 573,  
883, 903  
Kolpakov K. 843  
Kolpashchikov D.M. 355  
Koltover V.K. 1035  
Koltsova A.S. 417  
Komar A. 193  
Komissarov A.S. 70, 71, 448  
Komleva Yu.K. 1070  
Komyshev E.G. 615, 1094  
Konanov D.N. 216  
Kondakova E. 1043  
Kondaurova E.M. 717  
Kondrakhin Y. 77, 94  
Kondrashov O.V. 288  
Konenkov V. 830, 832  
König C. 989  
Kononkova A. 111  
Kononov V.A. 44  
Konstantinov Yu.M. 56, 608, 749  
Konstantinova A.V. 290  
Kontermann R. 1007  
Kontsevaya G.V. 668  
Kopylova G. 46  
Korenskaia A.E. 146, 580, 582  
Korneenko E.V. 885, 892  
Kornienko T.E. 990  
Kornilov F. 303  
Korobeynikova A.V. 79, 197, 232, 766  
Korobkov A.A. 263  
Korobov D. 481  
Korol A. 635  
Korolenko E.T. 798  
Korolenko T.A. 736, 798  
Korolev M. 843  
Korosteleva A.V. 427  
Korotkova A.M. 616  
Korovina A. 1005  
Korzhuik A.V. 491, 513, 557  
Kosareva A. 937  
Koshechkin K.A. 1103  
Koshenko T.A. 500  
Koshkina D. 1005  
Kosman E.S. 62  
Kostina M.V. 633  
Kosykh A.V. 768  
Kot E.F. 287, 291, 293  
Koteneva E.A. 549, 550  
Kotova T. 294  
Koval O.A. 839, 992, 1025  
Koval V. 615, 617  
Kovalenko I.L. 738  
Kovalev A. 524  
Kovalev K. 302  
Kovalev S. 419  
Kovaleva E.S. 551  
Kovaleva M. 302  
Kovaleva V.Y. 142, 874  
Kovner A. 799  
Kovrizhnykh V.V. 652  
Kovtun A.S. 490, 496  
Kozhekin M. 617  
Kozhemyakina R.V. 689, 690, 825  
Kozhevnikova E.N. 721, 946  
Kozhevnikova O.S. 801, 1034, 1067, 1071  
Kozlov A.P. 149  
Kozlov K.N. 211  
Kozlov N. 665  
Kozlova A. 858  
Kozlova A.P. 152  
Kozlova L. 654, 660  
Kozlova O. 48, 71, 92  
Kozlova T.A. 1045, 1066  
Kozlovsky Yu.E. 490  
Krasikova A. 49, 118  
Krasikova Y.S. 994  
Krasner K.U. 802  
Krasnoperova E. 665

Krasnova O.A. 217, 1046  
Kratasyuk V.A. 329  
Kravchenko E. 339, 539  
Kravchuk E. 327  
Kravchuk O. 108  
Krementsova A.V. 1062  
Krintsina A. 619  
Kriukov D. 1047  
Krivenko O. 29, 35  
Krivonosov M. 742  
Krivorotko O. 909, 914, 1128  
Kropochev A. 498, 578  
Kropocheva E. 983  
Krotenko N.M. 962  
Krut'ko V. 1048  
Krylov A. 514  
Krylov N.A. 267  
Krysko D. 856  
Kubareva E. 1011  
Kucher A.N. 381, 383  
Kudryavtseva L.P. 60  
Kudryavtseva N.N. 28, 738  
Kudryavtseva O. 421  
Kuianova Iu. 229  
Kukuman D. 1049  
Kukushkina I.V. 829, 1051  
Kulaeva E. 803  
Kulakov M.P. 243  
Kulakova K.A. 217, 1046  
Kulakovskiy I.V. 81, 87, 1107  
Kulazhenko M.S. 645  
Kulbachinskiy A. 166, 983  
Kulbatskii D.S. 319  
Kuleshova O. 29  
Kuligina E. 819  
Kulikov A. 882  
Kulikov A.V. 683  
Kulikova A. 166  
Kulikova N. 518, 526  
Kulikova T. 49  
Kulms D. 1007  
Kulyashov M.A. 573, 584  
Kupriyanova E. 499  
Kuprushkin M.S. 984  
Kurgina T.A. 996, 998, 1008  
Kurina A. 26  
Kurochkin N.N. 972  
Kurochkina N.S. 1052  
Kurochkina Y.V. 843, 1085  
Kurova A. 685  
Kussainova A. 761  
Kutumova E.O. 214  
Kutuzov M.M. 996, 998, 1008, 1023  
Kutyркиn V. 896  
Kuziakova O. 407  
Kuzmenko A. 983  
Kuzmichev P. 311  
Kuzmin A. 296  
Kuzmin I.V. 696  
Kuzmin P. 274  
Kuznetsov I. 470  
Kuznetsov N.A. 970, 971, 975, 976, 988,  
1000, 1001, 1018, 1022  
Kuznetsov S. 275  
Kuznetsova A.A. 1001, 1018  
Kuzucu E. 1105  
Kyrova E. 177  
Lachynova M.E. 154  
Lagunin A. 345  
Lagunov T. 123  
Lahiri A. 347  
Laikov A. 534  
Lakhno V.D. 1106  
Lakhova T. 218, 225  
Lalova M.V. 566  
Lamanov A.N. 764  
Lan F.H. 423  
Lantushenko A. 694  
Laprina D.S. 1018  
Larichev K. 621  
Larin A.K. 392  
Larina D. 438  
Larkin D.M. 692, 699, 709  
Lashin S.A. 83, 204, 218, 220, 225, 227, 241,  
498, 575, 578, 580, 582, 593, 792, 952  
Latsis I. 359  
Latyuk E. 523  
Lauc G. 474, 476  
Lavrekha V.V. 655  
Lavrenov A.R. 696  
Lavrik I.N. 206, 986, 989, 1002, 1025  
Lavrik O.I. 969, 972, 981, 984, 990, 994, 996,  
998, 1003, 1008, 1013, 1019, 1023  
Lazarev V. 359, 911  
Lazareva N. 1076  
Lebedev G.S. 1103  
Lebedev I. 688  
Lebedev I.N. 131, 411, 436, 457  
Lebedev S.N. 901, 921  
Lebedkin D.A. 950, 960  
Lednev E.M. 1052  
Lemskaya N.A. 137  
Leonova I.N. 614  
Leonova T.E. 560, 590  
Leonteva O. 390  
Lesani S. 787  
Leslie N. 765  
Lesovoy D. 275  
Letyagin A.Yu. 764, 804, 811, 836  
Levchenko I.N. 221, 223  
Levitsky V.G. 83, 90, 100, 599, 655

Lezhnin S. 562  
 Li H. 425  
 Liao J. 423  
 Liaudanski A. 594  
 Likhachev I.V. 298  
 Likhoshvai V.A. 211  
 Likhovskoi V.A. 630  
 Lin C. 303  
 Liou M. 942  
 Lipchinsky A. 663  
 Lipina T.V. 747  
 Lishai E. 826  
 Lisitsa A.E. 281, 329  
 Lisitskaya L. 983  
 Litti Yu. 524  
 Litunenko D.N. 553  
 Litus E. 779, 818  
 Litvinenko I.A. 921  
 Litvinova A. 1011  
 Litvinova L. 193  
 Lobanova N. 1043  
 Lobaskova M. 932, 937  
 Loboda A. 469  
 Logachev A. 623, 642  
 Logacheva M.D. 697  
 Logofet D.O. 244  
 Lomakina A. 514, 540  
 Lomzov A. 775  
 Lopatkina M.E. 411  
 Lozovskaya E.A. 749  
 Luginina A.P. 262, 265, 294, 300, 302, 303  
 Lukyanchikova V. 115, 117  
 Luparev A.R. 867  
 Lushchay E.A. 630  
 Lushpa V.A. 286, 300, 303  
 Lutova L. 661, 665  
 Luzina O.A. 990  
 Lyamzaev K.G. 1054  
 Lyapina E. 265, 302  
 Lys A. 1005  
 Lysenko E.A. 84, 829  
 Lytkin I.N. 570  
 Lytkin K.F. 630  
 Lyubetsky V.A. 173  
 Lyukmanova E.N. 317, 319  
 Lyupina Y. 108  
  
**M**achulin A. 156  
 Makarikova O. 158  
 Makarova A.V. 166, 1004, 1021, 1026  
 Makarova N. 49  
 Makarova V. 807  
 Makeev V.J. 55, 81, 87  
 Makeeva O. 388  
 Makhnovskii P.A. 84, 829, 1051, 1052  
 Makovka Yu.V. 690  
  
 Maksimenko K. 628  
 Maksimov V. 103  
 Maksimova K.Yi. 1066  
 Maksimova N.R. 405, 409, 436, 445, 450, 459  
 Maksiutenko E.M. 85, 404  
 Malankhanova T.B. 854  
 Malinovskaya N.A. 1070  
 Maljković M. 888  
 Malkov S. 888  
 Malovichko Y.V. 172, 543  
 Maltsev V.P. 836  
 Maltseva D. 333  
 Maluchenko N. 1005  
 Malyavko A.A. 690  
 Malykh S. 932, 937  
 Malyugin E. 50  
 Malyutina S. 1055  
 Manakhov A.M. 808, 828, 847  
 Manolov A.I. 851, 885, 1089  
 Manuvera V. 911  
 Manzeniuk I. 518, 526  
 Marakhonov A. 442  
 Marchenkov A. 528  
 Mardanov A.V. 520, 530, 532  
 Marin E. 265, 294, 302  
 Markel A.L. 678, 689, 713, 825, 842  
 Markelova M. 499  
 Markov A.V. 381, 383, 457  
 Markov O.V. 984  
 Markova E.V. 701, 719, 740  
 Markova Z.G. 411  
 Markovic S. 916  
 Marsova M.V. 500, 564  
 Maryum B. 1117  
 Mashkina E. 803  
 Mashko S.V. 551  
 Maslov D.E. 52  
 Maslov N.A. 729  
 Maslova A. 49, 118  
 Maslova E.A. 770, 771  
 Matskova L.V. 58  
 Matsvay A.D. 348, 428, 867  
 Matushkin Yu.G. 241, 580, 582  
 Matveenko A.G. 85  
 Matveishina E. 87  
 Matyushkina D. 911  
 Matyuta I. 273, 321  
 Mavletova D.A. 500  
 Mavliev F.A. 392  
 Maximov V. 1055  
 Mazur A.Ch. 494  
 Mazurov E. 87  
 Mednikova M.B. 482  
 Mednova I.A. 962  
 Medvedeva A.V. 966  
 Medvedeva Y.A. 87

Meger Y. 53, 694  
 Mehrani M.J.B. 787  
 Melamud M. 785  
 Melamud V. 520, 523  
 Meledin V.G. 548  
 Melikhova A.V. 160, 271  
 Melnikova N.V. 33, 60, 631, 633  
 Menzorov A. 671  
 Merkulova E. 931  
 Merkulova T. 434, 466, 757, 777  
 Merkushova A. 543  
 Meschaninova M. 819  
 Meshcheryakov G. 1107  
 Meshcheryakov N. 671  
 Meshcheryakova E.N. 703  
 Meshcheryakova I.A. 491, 513, 546,  
 554, 555, 557, 570  
 Mezhlumyan E. 929  
 Michurina S.V. 810, 811  
 Miguel A.A. 764  
 Mikhailov K.V. 697  
 Mikhailov S.N. 972  
 Mikhailova A.G. 277  
 Mikhailova A.Y. 937  
 Mikhaleuka A.P. 494  
 Mikhaylov I.M. 348, 428, 867  
 Mikhaylova Y. 518, 526, 845  
 Mikheeva E. 524  
 Mikushina E.S. 988  
 Milakhina E. 992  
 Milakhina N. 944, 967  
 Milicevic M. 897  
 Milicevic O. 916  
 Milyaeva P.A. 696  
 Mineev K.S. 286, 287, 291, 293, 300, 303  
 Mingazheva E.T. 427  
 Minina J.M. 672  
 Minushkina L. 1095  
 Mirhoseini S. 305  
 Mironova V.V. 65, 655, 662, 666  
 Mirosnichenko M.I. 883  
 Mirosnichenko S.M. 810, 811  
 Mirzaei M. 305  
 Mirzaeva S. 860  
 Mishchenko E. 812  
 Mishchenko T.A. 742, 856  
 Mishin A.V. 262, 265, 283, 294, 302  
 Mishra V. 1009  
 Mishuk Y. 594  
 Mistro G. 1007  
 Mitić N. 888  
 Mitkevich V.A. 375  
 Mitrofanova I. 35  
 Mitroshina E. 742, 954  
 Mohamed O. 887  
 Moiseenko A. 127  
 Mokrushnikov P.V. 813  
 Molobekova K.A. 717  
 Molodin V.I. 486  
 Monakhova M. 1011  
 Montero J. 344  
 Moor N.A. 996  
 Mordvinov V.A. 52, 351, 799, 826  
 Morgunova G.V. 1057  
 Morozov A. 528  
 Morozova K.N. 349, 438, 681, 854  
 Morozova M.V. 946  
 Moshensky D. 162  
 Moshkin M.P. 349, 668, 674  
 Moshkin Y.M. 668, 674  
 Moskalenko S.E. 85  
 Moskalensky A.E. 553, 793, 815, 827, 852  
 Moskalev A. 1049  
 Motov V.V. 293  
 Moysak G.I. 961  
 Mozheiko E.A. 54, 122, 411  
 Mozyleva Y.A. 197, 232  
 Mukhin A.M. 83, 220, 225, 227, 593  
 Mukhina K. 88  
 Mulkidjanian A.Y. 1054  
 Muntyan V.S. 152  
 Muraleva N.A. 1059, 1080  
 Mursalimov S. 657  
 Murtaza I. 789  
 Murugova T. 339  
 Musabirov A. 983  
 Mushtaq I. 789  
 Mustafin Kh.Kh. 482  
 Mustafin R. 932, 937  
 Mustafin Z.S. 50, 204, 227, 593, 792  
 Muterko A.F. 610, 635  
 Mutin A. 169  
 Mutovina A. 715, 929  
 Muvingi M. 535  
 Myakinkov I. 120  
 Myasnikova E. 235  
 Nabiullina R.M. 392  
 Nadezhdin K.D. 287  
 Natashin P.V. 307  
 Naumenko K.N. 996, 1008  
 Naumenko V.S. 717, 928, 934, 935, 964  
 Nazarenko L.P. 411  
 Nazarenko M.S. 381, 383, 394, 400, 825, 863  
 Nazarova G. 675  
 Nazina T.N. 511, 521, 537  
 Nebesnykh I.A. 139  
 Nebrat V. 816  
 Nechaeva A. 520  
 Nechepurenko Yu. 212  
 Nedoluzhko A. 484  
 Nefedova L.N. 696

- Neganova I.E. 217, 1046  
 Nekrashevich N. 26  
 Nekrasov A.N. 269  
 Nekrasova M. 707  
 Nemashkalova E. 779, 818  
 Nemerov E.V. 394  
 Nemtseva E.V. 281, 329  
 Nenasheva N.V. 55, 484  
 Nesmelov A. 48, 92  
 Nesterov M.A. 609  
 Netesov S.V. 912  
 Neverova G.P. 246, 248, 257  
 Nevinsky G.A. 366, 785, 841  
 Nezametdinova V.Z. 507  
 Nikiforov A. 35  
 Nikiforov S.B. 749  
 Nikitin A. 396  
 Nikitin N. 1124  
 Nikitina E.A. 966  
 Nikitina E.S. 965  
 Nikitina T. 457  
 Nikolaev Yu.A. 532  
 Nikolaeva A. 273, 321  
 Nikolaeva I. 409, 459  
 Nikolaichik Y. 73, 106  
 Nikolaychuk K. 1037  
 Nikulich I.F. 801, 1034  
 Nilov D. 1005  
 Nimaev V.V. 836, 849  
 Niyazova T.K. 755, 880  
 Nizamutdinov I. 688  
 Nizhnikov A.A. 172, 543, 626  
 Nizovtseva I. 562  
 Nolde D.E. 267, 317  
 Nostaeva A.V. 470, 476  
 Nosulchak V.A. 630  
 Novak D.D. 839  
 Novakovskiy R.O. 631, 633  
 Novikov K. 913  
 Novikova A. 1021  
 Novikova D.D. 662  
 Novikova P.Y. 627  
 Novopashina D. 819  
 Nuriddinov M. 117, 118, 122, 411  
 Nurislamov A. 123  
 Nurmukanova V. 428  
 Nurullin L. 71  
 Nushtaeva A.A. 839  
 Nyunkov P.A. 566
- O'Brien P.** 179  
 Obanina N.A. 821  
 Odorskaya M.V. 500  
 Ogienko A.A. 721  
 Oguienko A. 983  
 Ohno N. 657
- Okhriemenko I. 302, 309, 311, 337  
 Okhtienko A. 983  
 Okunev A. 1086  
 Olekhnovich E.I. 505  
 Olina A. 983  
 Olkov N.M. 961  
 Omelyanchuk N.A. 65, 655, 658  
 Ontiveros N.J. 344  
 Onufriev M. 1076  
 Oparina N.Y. 79  
 Orbant M.O. 668  
 Orekhov Ph. 302  
 Orlov Y.L. 44, 79, 197, 199, 232, 271, 305, 323, 425, 766, 773, 781, 823, 1103  
 Orlova N.G. 44, 197, 232, 823  
 Oshchepkov D. 83, 432, 554, 555, 689, 825, 863  
 Osipova L.P. 419, 452  
 Oslovsky V.E. 972  
 Osypov A. 88  
 Otomo T. 455  
 Ovsyannikova A. 430  
 Ovsyukova M.V. 731, 732, 744, 751
- Pak S.** 238  
 Pakharukov Y.V. 351  
 Pakharukova M.Y. 52, 351, 799, 826, 862  
 Palyanov A. 1109  
 Panfilov M.A. 793, 815, 827  
 Panikov N.S. 586  
 Pankova O. 437  
 Pankova S. 484  
 Panov A.V. 1060  
 Panova V. 313  
 Paramonik A. 841  
 Paramonov A.S. 319  
 Parshin D. 229  
 Parshukova D. 785  
 Parui A. 1009  
 Pasko A. 579  
 Pastré D. 1019  
 Pasyukova E.G. 1062, 1069  
 Patrakova E. 992  
 Patrushev M.V. 609  
 Pavlenko A. 911, 1089  
 Pavlenko M.V. 670  
 Pavlenko O.N. 921  
 Pavlova A. 1011  
 Pavlova E. 557  
 Pavlova O.A. 390  
 Pavlova O.N. 540  
 Pavlović-Lažetić G. 888  
 Pavlovsky E.N. 804  
 Pelevina A.V. 530  
 Peltek S.E. 491, 513, 546, 554, 555, 557, 559, 568, 570

- Pendina A.A. 417  
Penenko A. 1111  
Penkin L.N. 885  
Penzar D. 174  
Pereira J.C. 167  
Perelman P.L. 113  
Perelmutter V. 437  
Peretolchina T. 677  
Pereverzev I. 1119  
Perfl'ev R. 628  
Permyakov S. 779, 818  
Permyakova E.S. 808, 828, 847  
Pershina A. 853  
Petrakova V. 914  
Petrenko D. 277  
Petrova A. 654, 660  
Petrova D. 528  
Petruseva I. 1013  
Petrushanko I.Yu. 271, 375  
Petrushin I.S. 56, 163, 608  
Petukhova D. 409, 459  
Peunov D.A. 1063  
Pevzner I.B. 1065  
Piatkova M. 407  
Pichkur E. 317  
Pilipenko A.S. 486  
Pilipenko I.V. 486  
Pilsner J.R. 68  
Pimenov N.V. 530, 532  
Pintus S.S. 234, 573  
Pisareva N.V. 1070  
Plashchinskaia D. 275  
Pletenev I. 111, 124  
Plotnikov E.Y. 1065  
Plotnikov N. 484  
Plotnikov V. 118  
Pnev S. 1124  
Podgornaya O. 140  
Podkolodnaya O.A. 230, 1096  
Podkolodnyy N.L. 230, 1096  
Pogodaeva T. 514, 540  
Pogodina N. 484  
Pokatova O. 355  
Poleksic A. 1112  
Polenov N.I. 417  
Poliakov A. 484  
Polonikov A. 400, 415, 421  
Poloznikov A. 954  
Polshvedkina O. 415  
Poltorachenko V. 1021  
Poluboyarova T.V. 703  
Poluektova E.U. 500  
Polunin D.A. 142, 874  
Polyakova V.A. 867  
Polyansky A.A. 267  
Polyudina R. 628  
Ponomarenko M.P. 432, 638, 689, 825, 863  
Ponomarenko P. 689, 825  
Ponomarev D.V. 351  
Ponomareva L. 421  
Popik V. 546, 554, 555, 559  
Popkov V.A. 1065  
Popkova T.P. 749  
Popov A. 49  
Popov A.A. 1013  
Popov D.V. 84, 234, 829, 1051, 1052  
Popov P. 302  
Popov V. 273  
Popov V.N. 1065  
Popova A. 158  
Poroikov V.V. 341, 345, 352, 364  
Poroshina A.A. 250  
Pospelova T. 103  
Posukh O.L. 770, 771  
Potapov D. 628  
Potapova N.A. 474  
Potokina E.K. 630, 646  
Potsenkovskaia E. 661, 665  
Povshenko O.V. 802  
Povkhova L.V. 631, 633  
Pozdnyakov D.V. 486  
Pozhvanov G. 663  
Predeus A.V. 385  
Presta L. 1007  
Pristyazhnyuk I.E. 438, 440  
Prokofiev V. 830, 832  
Prokofieva D.S. 427  
Prokopov D.Yu. 113, 137, 167, 179  
Pronozin A. 50, 57  
Proskura A.L. 834, 948  
Proskurnyak L. 675  
Proskuryakova A.A. 165, 167  
Prostova M. 166  
Pupyshev A.B. 727, 731, 736, 744  
Pushkova E.N. 33, 631, 633  
Pushmina E. 163  
Puzyrev V.P. 394  
Pyakillya B.I. 354  
Pyrkova A.Yu. 880  
Pyshnyi D. 775  
  
Qiu P. 423  
  
Rachkovskaya L.N. 810, 811, 836  
Rachkovsky E.E. 811, 836  
Raditsa V.V. 90  
Rajoka M.S.R. 362  
Rakhmetullina A.K. 880  
Rakitina T. 277, 321, 335  
Rakitko A. 484  
Rapoport I.B. 171  
Rasskazov D. 64, 432

Ratushnyak A.S. 834, 948  
 Ravin N.V. 530, 532  
 Rechkunova N.I. 994  
 Redina O.E. 678, 713, 825, 842  
 Reshetnikov V. 42, 715, 929  
 Reveguk I. 407, 442  
 Revel-Muroz A. 125  
 Revva P.M. 206, 231  
 Rezvykh A.P. 965  
 Richter M. 1025  
 Richter V.A. 819, 839, 1025  
 Riemann S. 1007  
 Rihan F.A. 887, 895, 915  
 Rinchinov O. 1086  
 Rippa A.L. 768  
 Ritter G. 1014  
 Rodic A. 897, 916  
 Rodionov E. 437  
 Rodnyy A.Ya. 717  
 Rogachev A. 302  
 Romanenko S.A. 137, 167  
 Romanishin A. 407  
 Romanov V. 491  
 Romanova E.V. 697  
 Romanova V. 534  
 Romanyukha A.A. 905, 913, 923  
 Romashchenko A.V. 194, 349, 679, 681, 724  
 Romashov G.A. 699, 709  
 Roshina N.V. 1062  
 Rotskaya U.N. 62  
 Roumiantseva M.L. 152  
 Rozanov A. 557  
 Rozantseva V.V. 560  
 Rozhmina T.A. 33, 60, 595, 600, 623,  
 631, 642  
 Rubanov L.I. 173  
 Rubel M. 355  
 Rubets V. 535  
 Rubina K.A. 860, 956, 1016  
 Rubtsova D.V. 670  
 Rudnitskaya E.A. 1032, 1045, 1063, 1066  
 Rudyak V.Ya. 813  
 Rudych P. 931  
 Rumyantseva Yu.V. 801, 1067  
 Rusinov I. 184, 576  
 Russkikh N. 1113  
 Ryabchenko L.E. 560, 590  
 Ryabikov A. 1055  
 Ryabova A. 77, 94  
 Ryabova A.S. 214  
 Ryabova A.V. 92  
 Ryabushkina Y. 715, 929  
 Ryazansky S. 166  
 Ryazantsev G. 1113  
 Rybina O.Y. 1062, 1069  
 Rykova E. 434  
 Rymar O. 430  
 Rysenkova K. 956, 1016  
 Ryumina E. 127  
 Ryzhikov A. 331  
 Ryzhkova A. 115  
 Ryzhkova T.V. 379  
 Rzhechitskiy Y. 169  
 Sadova A. 283  
 Sadovnikova I.S. 1065  
 Sadovskaya A. 357  
 Saenko S. 518, 526  
 Safronova N. 265  
 Sagynbay Sh. 755  
 Saifitdinova A.F. 131, 390  
 Saik O. 837, 890  
 Saiqa A. 1117  
 Salakhutdinov N.F. 735, 751, 990  
 Salina E.A. 31, 609, 610, 614, 621, 628, 635  
 Salmin V.V. 1070  
 Salmina A.B. 746, 1070  
 Salnikov P. 129, 134, 148  
 Salom I. 897, 899, 916  
 Saltykov M. 252  
 Samoilov A.E. 619, 885, 892  
 Samsonova A.A. 595, 600, 623, 642,  
 762, 1121  
 Samsonova M. 66  
 Samsonova M.G. 595, 600, 623, 642, 685  
 Sanina N.A. 1035  
 Sannikova A.V. 636  
 Sannikova T.E. 905, 918, 923  
 Saparbaev M. 975  
 Sapronova A. 929  
 Saprygin A.E. 931, 936, 939, 944, 950  
 Saranchina A. 169  
 Saryglar A. 419  
 Sauter T. 1007  
 Savchenko S.V. 764  
 Savina M. 666  
 Savinkov R. 871  
 Savinkova L.K. 432, 638, 689  
 Savinkova M.M. 839  
 Savitskaya V. 1011  
 Savkin I.V. 719  
 Savostyanov A.N. 931, 936, 944, 950, 952,  
 960, 967, 1105  
 Savvina M. 436  
 Savyuk M.O. 856, 954  
 Say A.V. 348, 428  
 Saydakova S.S. 721  
 Sazhenova E. 457  
 Schegoleva A. 437  
 Schelkunov M. 49  
 Schindler S. 1007  
 Schweigert I. 992

Sedykh S.E. 366, 841  
Sekacheva M.I. 425  
Selezneva O. 851  
Seliverstov A.V. 173  
Semenova D. 901  
Semenova E. 521, 537  
Semenova E.A. 392  
Semenova E.V. 1060  
Semina E.V. 860, 956, 1016  
Semke A.V. 958  
Senchurova S.I. 1018  
Serdyukova N.A. 137, 672  
Serebrennikova M. 359  
Serebrova V. 453  
Seregin A. 619  
Seregin A.A. 958  
Serenko E.V. 701  
Sergeeva E.M. 610, 621  
Sergeeva S. 491  
Sergeyev O. 68  
Sergushichev A. 99  
Serov O.L. 440, 672  
Seryakov A. 1119  
Seryapina A.A. 842  
Serykh A.E. 810, 811  
Shabalkina A.V. 287, 291, 293, 303  
Shaburova E.V. 1113, 1115  
Shadrin N. 694  
Shadrina A. 478  
Shagimardanova E.I. 48, 71, 92, 180, 379, 829  
Shaikhutdinov N.M. 92, 180  
Shakhtshneider E. 430  
Shakirov E.V. 636  
Shakirova R.R. 197, 199, 232  
Shalaginova I.G. 58, 722  
Shalik I.K. 95  
Shananin A. 920  
Shaposhnikov M. 1049  
Sharakhov I. 117  
Sharapov S.Zh. 474, 476  
Sharapova M.B. 679, 681, 686  
Sharipov R.N. 77, 94, 214  
Sharipova M.R. 636, 645  
Sharypova E.B. 432, 638, 825  
Shatarnova T. 594  
Shatskaya N.V. 146, 640  
Shatunova E. 843  
Shayakhmetova L.Sh. 366  
Shaydullina A.D. 427  
Shaytan A.K. 315  
Shchepkin D. 46  
Shcherban A. 628  
Shedko E. 491, 546, 555, 557, 559, 568  
Shekhovtsov S.V. 171, 703  
Shelenkov A. 518, 526, 845  
Sheng G. 573  
Shenkarev Z.O. 317, 319  
Shenoy P.U. 873  
Shepeleva D.V. 690  
Shepelin D.D. 551  
Sherbakov D.Y. 677, 697  
Sheremetieva M.E. 560  
Sheremetyeva I.N. 670  
Shevchenko A.U. 58  
Shevchenko A.V. 830, 832  
Shevelev O. 686  
Shevelyova M. 779, 818  
Shevtsov M.B. 265, 294, 300, 302  
Shigapova L.Kh. 180  
Shikhevich S.G. 689, 690, 825  
Shikov A.E. 172, 543  
Shilenok I. 398  
Shilkin E. 166, 1021, 1026  
Shilova N.V. 131, 411  
Shilova S. 321  
Shilovsky G.A. 173  
Shipova A.A. 513, 557, 568, 570  
Shipulin G.A. 348, 428, 867  
Shirazi E.A. 323  
Shishkina O.D. 146  
Shishlenin M. 589, 1097  
Shitik E.M. 95  
Shklyar A. 1071  
Shkodenko L. 355  
Shlyakhtun V. 491, 557  
Shmakov N.A. 593, 640  
Shmakova A. 1016  
Shnaider T.A. 438, 440  
Shnyrova A. 274  
Shokina V. 911  
Shtansky D.V. 828  
Shtern E.N. 960  
Shtratnikova V. 68  
Shtykova E. 335  
Shubenkova O. 540  
Shubin A. 253  
Shulepko M.A. 319  
Shulinski R.S. 494  
Shulpin M. 484  
Shumkov O.A. 849  
Shunkina D. 193  
Shunkov M. 487  
Shupletsov M. 544  
Shuvaev A.N. 1070  
Shvaikovskaya A.A. 961  
Shvetsov A.V. 878, 1123  
Sidorenko A.D. 662  
Sidorenko A.V. 490  
Sidra L. 1117  
Sigorskih A. 174  
Sigova E.A. 33, 60



Silachev D.N. 1065  
Silnikov V. 819  
Simonenko S.Yu. 271  
Simonova V. 665  
Singatulina A. 1019  
Siniauskaya M. 594  
Sinkkonen L. 1007  
Sitnikova N.A. 847  
Sivkina A. 325  
Sivtsev A. 462  
Sizikov A. 785  
Skakov I. 68  
Skakun V. 106  
Skitchenko R. 407  
Sklyarov A.N. 948  
Skoblov M. 407, 442  
Skolotneva E.S. 610  
Skorik V. 1111  
Skripileva A.I. 616  
Skryabin N. 462  
Slobodchikov A. 617  
Slovareva O. 535  
Slynko N.M. 570  
Smagin A.A. 836  
Smagin D.A. 738  
Smagin M.A. 849  
Smirnov A.V. 440, 671  
Smirnov D. 1072  
Smirnova A. 438  
Smirnova E. 283  
Smirnova K.V. 747  
Smirnova L.P. 783, 958, 962  
Smolenskaya S.E. 678  
Smolnikova M. 443  
Smykov A. 644  
Snytnikova O.A. 946  
Sobolev A. 125  
Sobolev B. 345  
Soboleva O. 402  
Sobolevskaya E.V. 849  
Sofija M. 897  
Sofronova V. 445  
Sokolova D. 521, 537  
Sokolova O.S. 325, 327  
Soldatov V. 398  
Soldatova M. 398  
Solodneva E.V. 705  
Soloviev N.O. 988  
Solovieva A.O. 808, 828, 847  
Soloviova A. 26  
Solovyev M. 125, 502  
Son A. 234  
Sonets I. 502  
Sorokin I.E. 683  
Sorokina N. 97, 140  
Sosnovskaia M. 909  
Speranskaya A.S. 619, 885, 892  
Spirin S. 174, 184, 313, 576  
Spirov A. 235  
Sripa B. 826  
Srivatsan R. 75  
Stanin V. 623, 642  
Stanova A.K. 668, 684  
Starchevskaya M.E. 40, 62  
Starikova E. 851  
Starodubtseva E.S. 852  
Starodumov I. 562  
Stavrovskaya A.V. 564  
Stefanova N.A. 717, 1045, 1066, 1074  
Stein D. 1072  
Stepanchuk Y.K. 129, 131, 411  
Stepanichev M. 1076  
Stepanov G. 841  
Stepanov V.A. 388, 413, 447, 453  
Stepanova M. 707  
Stetsenko I.F. 348, 428  
Stojanović B. 888  
Stojku S. 899  
Stolpovsky Y.A. 705  
Stolyarenko A. 1021  
Studitsky V. 325, 1005  
Stupina A.A. 573  
Stupnikov A. 68  
Sudakov N.P. 749, 941  
Sukhachev V. 360  
Sukhanova M. 969, 1008, 1019  
Sukhanova X.V. 448  
Sukhih I. 557  
Sukhinina E. 853  
Sukhomlinov D. 1119  
Sukhomyasova A.L. 405, 409, 436, 445, 450, 459  
Sukhov V. 99  
Sukovatyi L.A. 281, 329  
Suldina L.A. 854  
Suri P. 464, 467  
Surkova S. 685  
Surnin P. 589  
Surovtseva M.A. 802  
Suslov D. 663  
Suslov V.V. 182, 689  
Suvorov A.N. 68, 504  
Svarovskaya M. 453  
Svishcheva G.R. 473, 705  
Svitich V. 562  
Symonenko A.V. 1062  
Sysoeva V.Yu. 860  
Tabakmakher V. 333  
Tabassum S. 362  
Tabikhanova L.E. 419, 452  
Tahirov T.H. 1004

Takhirova Z. 932  
Talyzina I.A. 300, 303  
Tamazian G. 1121  
Tamozhnikov S.S. 931, 936, 939, 944, 952  
Taranik A.V. 921  
Tarasenko A. 799  
Tarasov K. 539  
Tarasov O.V. 176  
Tarasova A.A. 482  
Tarasova O. 341, 364  
Taskaev S. 819  
Tatarova L.E. 570  
Telegina D.V. 801, 1030, 1032, 1077  
Ten M.N. 749  
Tenditnik M.V. 731, 732, 735, 744, 751  
Testic S. 177  
Tihomirov S. 129  
Tikhonova M.A. 727, 729, 731, 732,  
735, 736, 744, 751, 961  
Tikhonova P. 911  
Tikhvinsky D. 229  
Timmers P. 1079  
Timofeev V. 277  
Timofeeva A.M. 366, 841  
Timofeeva V. 335  
Timofeyev M. 169  
Timoshchuk A. 474  
Tishakova K. 179  
Titov A.I. 1123  
Titov I.I. 104, 590, 882, 1090  
Tiys E. 1079  
Tkachenko A.A. 404  
Tkachev S.E. 180  
Tkachuk V. 956  
Toiber D. 1072  
Tolmacheva E. 457  
Tolstokulakov N.Yu. 804  
Tomilin M.A. 486  
Tonevitsky A. 333  
Toschakov S.V. 609  
Tourova T. 521  
Trapezov O. 707  
Trapezov R.O. 486  
Travina A. 70  
Tregubchak T.V. 62  
Tretyakova I.S. 793  
Trifonov V.A. 113, 167, 179  
Trifonova E. 453  
Trofimova M. 578  
Trojanovskiy K. 956  
Trunov A.N. 802, 832  
Trusov N. 920  
Tsentelovich Y.P. 703, 946  
Tsepilov Y.A. 464, 467, 473,  
476, 478, 710, 1079  
Tsukanov A.V. 90, 100, 655  
Tsybko A.S. 928, 934, 935, 964  
Tsybul'ko E.A. 1062  
Tsygankova O.I. 549, 550  
Tsybmal V.V. 566  
Tsyupka V. 35, 644  
Tuchinov B.N. 804, 1124  
Tufail A. 367, 370  
Tulupov A. 194  
Tumbas M. 897  
Tur D. 686  
Turnaev I.I. 154, 182  
Turubanova V.D. 856  
Tverdokhlebl N.N. 230  
Tvorogova V. 661, 665  
Tyakht A. 125, 502  
Tyapkina O. 71  
Tymanyanyan V.G. 263  
Tyshkovskiy A. 1047  
Tyugashev T.E. 1022  
Tyumentsev M.A. 1045, 1063, 1080  
Tyumentseva M. 526  
Ubogoeva E. 666  
Ukrainitsev A.A. 996, 998, 1023  
ul Haq W. 362  
Ulianov K. 133  
Ulianov S. 125, 135, 502  
Unguryan V. 331  
Uppe V. 64  
Ushakov V. 557  
Ustrokhanova D. 466  
Utkin Ya.A. 137  
Uvarova Y.E. 491, 513, 555, 557, 559,  
568, 570  
Vagaitseva K. 388  
Vakhrameev D.D. 300, 302, 303  
Valeev E.S. 129, 131, 411  
Valeeva E.V. 392  
Valeeva L.R. 636, 645  
Valova G. 194  
Valova Ya.V. 427  
Varizhuk A. 317, 911  
Varlamov M.E. 839  
Vasilev F. 455  
Vasil'eva I. 998  
Vasiliev G.V. 146, 555, 609, 640  
Vasiliev P. 125, 502  
Vasilieva A.R. 555, 557, 568, 570  
Vasilieva E.V. 287, 291  
Vasiljeva O. 390  
Vasilyev S. 457  
Vasilyeva A.R. 783, 962  
Vasilyeva O. 457  
Vasyutkin S.A. 236  
Vatlin A.A. 500

Vdovina I. 421  
Vechkapova S.O. 834  
Vedekhina T. 911  
Vedunova M.V. 742, 856, 954, 964, 1043  
Veljković A. 888  
Vepkhvadze T.F. 84, 1052  
Vergasova E. 484  
Vergounov E. 950  
Vershinin A.V. 144, 602  
Vertyshev A. 234  
Veselovsky V.A. 505, 1089  
Vetchinnikova L.V. 646  
Vigovskiy M.A. 1052  
Vihorev A.V. 616  
Vilar M. 291  
Vilinski-Mazur K. 579  
Vinogradov K.S. 867  
Vinogradova O.L. 84, 829  
Vinokurov N. 546  
Vishnevsky O. 101, 104, 1126  
Vlad A.I. 923  
Vladimirova E. 438  
Vladimirova M.E. 152  
Vlasenkova R. 858  
Vodiasova E. 53, 64, 644  
Voinova V.Y. 411  
Volcho K.P. 733, 735, 751  
Volkov V.A. 630, 646  
Volkova N.A. 710  
Volobueva M. 184  
Volodyaev I.V. 221, 223  
Volokh O. 325  
Volynkin V.A. 630  
Volynsky P. 303, 311, 333, 335  
Vorob'ev A.Yu. 793, 815, 827  
Vorobev R. 437  
Vorobyeva M. 775, 819, 843  
Voronina E.N. 452  
Voropaeva E.N. 103, 197  
Vorotelyak E.A. 768  
Vorozheykin P. 104  
Voskoboiev M. 491, 557  
Vulf M. 193  
Vyatkin Y.V. 62, 1113  
Vychyk P. 73, 106  
Vysotski E.S. 307

Wajid A.A. 1117  
Wang X. 303  
Wang Y. 648  
Wettberg E. 66  
Wiebe D. 65, 599  
Williams F.M.K. 464, 467  
Wilson J. 1079  
Wohlfromm F. 1025

Xu Y.J. 423

Yablonskaya M.I. 411  
Yakhnenko A. 186, 539  
Yakimenko V.V. 180  
Yakovleva A. 405, 459  
Yakovleva D. 665  
Yamileva K.I. 860  
Yan A. 129, 134  
Yan A.Z. 423  
Yanenko A.S. 490, 560, 590  
Yang W.J. 423  
Yankovsky N.K. 461  
Yaremko A. 535  
Yarkov R. 964  
Yarmolinskaya M.I. 417  
Yastrebova E.S. 836  
Yatsevich K. 26  
Yavorsky N.I. 548  
Yevshin I.S. 77, 94, 214, 573  
Yoshitake Y. 645  
Yudin N.S. 692, 699, 709  
Yudina M. 1092  
Yudkin D.V. 95  
Yudkina A.V. 982, 1026  
Yukhtanov D. 1080  
Yunes R.A. 490, 500, 507  
Yunusova A.M. 120, 440  
Yurchenko A.A. 699  
Yurikova O.Yu. 880  
Yurtchenko D. 765  
Yusefovich A.P. 892  
Yusipov I. 1043  
Yusubalieva G. 396  
Yuzik E.I. 798

Zablotskaya S. 355  
Zadorozhny A.V. 491, 513, 557  
Zagryadskaya Yu. 311, 337  
Zaigraev M.M. 319  
Zakharenko A.L. 972, 981, 990  
Zakharenko A.S. 540  
Zakharova Y. 528  
Zakluta A.S. 965  
Zakrevsky D. 992  
Zalevsky A.O. 300, 303  
Zalomaeva E.S. 966  
Zapara T.A. 948  
Zaparina O. 799, 826, 862  
Zarubin A. 388, 457, 462  
Zarubin M. 339, 539  
Zatsepin O. 901  
Zatsepina O.G. 965, 1033  
Zavarvin E. 931, 967  
Zavestovskaya I. 373

Zavjalov E.L. 736, 811  
Zaytseva O.O. 467, 476  
Zelentsova E.A. 703  
Zemlyanskaya E.V. 65, 83,  
599, 655, 658, 662, 666  
Zemskaya N. 1049  
Zemskaya T. 514, 540  
Zgoda V.G. 829, 1051  
Zhaivoron A. 470  
Zhalsanova I.Zh. 394, 462  
Zhanaeva S.Y. 961  
Zhang D. 423  
Zhao M. 976  
Zharkov D.O. 977, 982, 987, 1026, 1027  
Zharkov T.D. 984  
Zhatchenko S.A. 214  
Zhdanov V.S. 248, 257  
Zhdanova O.L. 246, 248, 255, 257  
Zhegalova I. 133, 135  
Zheleznyak A.R. 375  
Zheng D.Z. 423  
Zhigalina D. 457  
Zhigunov A.V. 646  
Zhouravleva G.A. 85, 176  
Zhukova I. 388  
Zhukova N. 388  
Zhukova S. 619  
Zhurakovskaya A. 108  
Zhuravlev A.V. 966  
Zhuravleva E. 524  
Zigic D. 897, 916  
Zinovieva N.A. 710  
Ziobron A. 66  
Zlobin A.S. 710  
Zlobin V.I. 866  
Zlydneva N. 665  
Zolotareva K. 432, 825, 863  
Zolotovskaya E. 188  
Zorkoltseva I. 473, 478  
Zorov D.B. 1065  
Zorova L.D. 1065  
Zoruk P. 851  
Zozulya S.A. 744  
Zubairova U. 1111  
Zubkova E.V. 259  
Zubritskiy A. 87  
Zuev A. 457  
Zuev D. 724  
Zvereva M. 956, 1011  
Zverkov O.A. 173  
Zvonareva T. 1128  
Zyatkov N. 909  
Zytsar M.V. 770, 771

Научное издание

**BIOINFORMATICS OF GENOME REGULATION  
AND STRUCTURE/SYSTEMS BIOLOGY**

The Thirteenth International Multiconference

Abstracts

*Printed without editing*

**БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ  
ГЕНОМА/СИСТЕМНАЯ БИОЛОГИЯ**

Тринадцатая международная мультиконференция

Тезисы докладов

*Публикуется в авторской редакции*

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

---

Подписано к печати 30.06.2022. Электронное издание

---

Федеральный исследовательский центр  
«Институт цитологии и генетики Сибирского отделения Российской академии наук»  
630090, Новосибирск, проспект Академика Лаврентьева, 10