

Comparative analysis of NGS and Sanger sequencing methods for HLA typing at a Russian university clinic

Oleg S. Glotov ^{1,2}, Olga V. Romanova ^{1,2}, Yuri A. Eismont ¹, Andrey M. Sarana ^{1,2}, Sergey G. Scherbak ^{1,2}, Elena V. Kuzmich ³, Alexander L. Alyanskiy ³, Natalya E. Ivanova ³, Vera V. Teplyashina ³, Yury A. Serov ³, Ludmila S. Zubarovskaya ³, Boris V. Afanasyev ³

¹ City Hospital №40, Sestroretsk, St. Petersburg, Russia

² Institute of Translation Biomedicine, St. Petersburg State University, St. Petersburg, Russia

³ Raisa Gorbacheva Memorial Research Institute for Pediatric Oncology, Hematology and Transplantation, The First St. Petersburg State I. Pavlov Medical University, St. Petersburg, Russia

Oleg S. Glotov, Ph.D., Laboratory of Genetics, Municipal Hospital №40, Sestroretsk, Borisova St 9, 197706, St. Petersburg, Russia; Laboratory of Genomic and Proteomic Research, Institute of Translational Biomedicine, St. Petersburg State University, Universitetskaya Emb. 7/9, 199034 Saint Petersburg, Russia

Phone: +7 (921) 756 7809

E-mail: olglotov@mail.ru

Citation: Glotov OS, Romanova OV, Eismont YuA et al. Comparative analysis of NGS and Sanger sequencing methods for HLA typing at a Russian university clinic. *Cell Ther Transplant* 2018; 7(4): 72-82

Summary

The database of the World Health Organization (WHO) Nomenclature Committee for Factors of the HLA System (IPD-IMGT/HLA Database) contained information on the nucleotide sequences of 20272 different HLA alleles in September 2018, of which 14800 were HLA class I and 5288 were found for the HLA class II alleles.

Over the last 20 years, the automated Sanger technique is a prevalent approach to genome sequencing in humans, animals, bacteria, and viruses. However, a need for more rapid routine genome screening stimulated novel technologies of multiplex DNA sequencing. These modern methods are depicted as the second-generation approaches (Next-Generation Sequencing, NGS).

The aim of our research was a comparison of two methods and their efficiency evaluation. To achieve our purpose, we selected a group of 35 DNA samples, mainly from potential hematopoietic cells donors, and conducted a comparative analysis by Sanger and NGS method.

NGS method allowed detecting rare or novel variants of alleles. This approach is confirmed to be more sensitive and more cost-effective, especially in large HLA-typing laboratories.

Keywords

Major histocompatibility complex, novel HLA alleles, technological solutions, next-generation sequencing, NGS, Sanger sequencing, hematopoietic cells transplantation, Sequence-Based Typing (SBT).

Introduction

The Major Histocompatibility Complex (MHC) is among the most polymorphic genetic systems in humans. Over last decade, extensive research in HLA (Human Leukocyte Antigens) has revealed hundreds of new HLA alleles through intensive application of immunogenetic sequencing methods, including monoallelic Sanger-sequencing method, or, more recently, next-generation sequencing. In September 2018, the database of the World Health Organization (WHO) Nomenclature Committee for Factors of the HLA System (IPD-IMGT/HLA Database) contained information on the nucleotide sequences of 20272 different HLA alleles, of which 14800 were HLA class I and 5288 founded for the HLA class II alleles [1-3].

During last 20 years, the automated Sanger technique has become a prevalent approach to genome sequencing in humans, animals, bacteria, and viruses. However, a need for more rapid routine genome screening required some novel technologies of multiplex DNA sequencing. It depicts these modern methods as the second-generation approaches (Next-Generation Sequencing, NGS). These technological platforms based on different strategies, regarding unique preparations of DNA templates, their sequencing, registration, retrieval and evaluation of the nucleotide sequences with novel bioinformatics approaches [4]. A principal benefit of the new-generation sequencing is an opportunity get large databases of multiple defined oligonucleotide sequences within a short time period with low costs.

Out of all known HLA loci, the relatively important and most commonly used for transplantation of hematopoietic cells are HLA – A, B, C, DRB1 and DQB1 (Fig. 1).

The American Society for Histocompatibility and Immunogenetics (ASHI) established a catalogue of common and well-documented (CWD) HLA. It is very commonly used now around the world as a great tool for resolving typing ambiguities in tissue transplantation or for checking the universality of any HLA allele in the world [5]. There established catalogues (database). The total number of CWD alleles is similar in the EFI (N = 1048) and ASHI (N = 1031) catalogues [6] (<http://igdawg.org/cwd.html>).

The importance of only Exons 2 and 3 for the Class I and Exon 2 for Class II is very well-known and designated as coding proteins involved in antigen presentation in the major histocompatibility complex (MHC) receptor groove in-between the two helices accommodates peptides and interaction between an alloantibody IgG complex.

HLA alleles having nucleotide sequences that encode the same protein sequence for the peptide binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles) designated by an upper case 'P' which follows the allele designation of the lowest numbered allele in the group. HLA alleles that have identical nucleotide sequences for the exons encoding the peptide binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles) designated by an upper case 'G' which follows the allele designation of the lowest numbered allele in the group.

The first two digits describe the allele family, which often corresponds to the serological antigen carried by the allo-type. The third and fourth digits assigned in the order in which the sequences have been determined. Alleles whose numbers differ in the first four digits must differ by one or more nucleotide substitutions that change the amino-acid sequence of the encoded protein. Alleles that differ only by

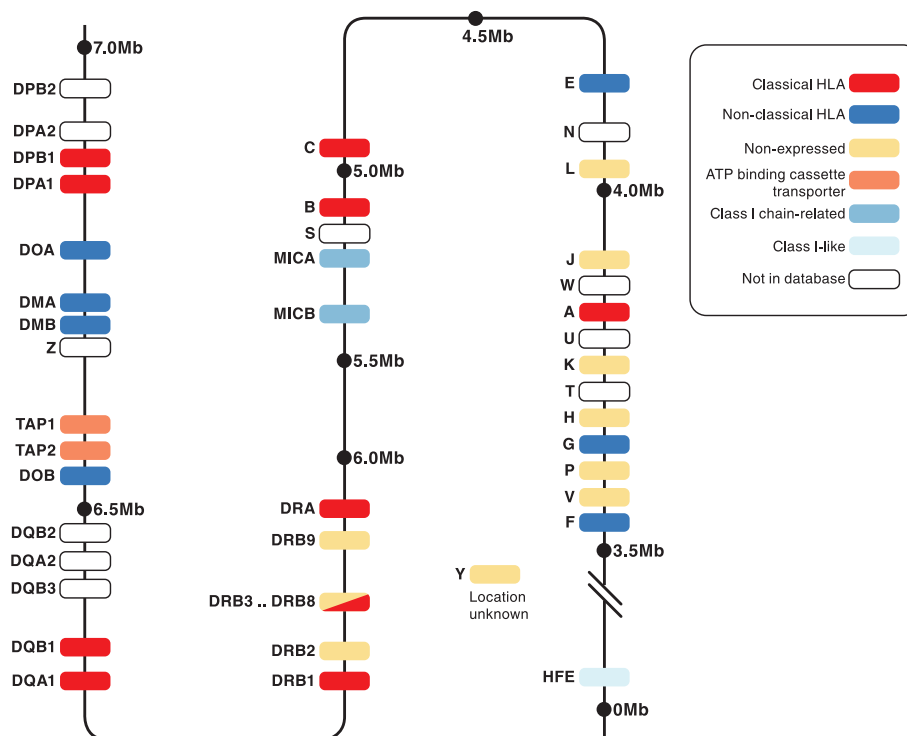


Figure 1. Current mapping of HLA loci on the chromosome 6 [Robinson J et al.] <http://www.hla.alleles.org/alleles/index.html>

synonymous nucleotide substitutions within the coding sequence distinguished by the use of the fifth and sixth digits. Alleles that only differ by sequence polymorphisms in introns or in the 5' and 3' untranslated regions that flank the exons and introns distinguished by the use of the seventh and eight digits [7].

No wonder that the general NGS approach adapted for HLA typing proved to be a breakthrough in molecular biology applications being quite promising to the transplantation clinics and bone marrow donor registries. However, to promote the NGS implementation, we need specialized typing strategies and digital program algorithms. The sequencing costs per single run sharply decreased with NGS approach, which may be accessible to the small size tissue typing laboratories in a sooner time [8].

However, despite higher resolution of NGS [9], it was necessary to conduct a comparative analysis of control samples with "rare" genotypes. It is also important to understand the cost-effectiveness of different methods. Hence, the aim of this pilot study was to evaluate the comparative advantages of using NGS and the Sanger sequencing approach, to identify rare HLA alleles, and to estimate the costs for the both different methods.

Materials and methods

The potential donor's test samples were obtained from the Bone Marrow Donor Registry at the First I. P. Pavlov State Medical University of St. Petersburg, Russia, Raisa Gorbacheva Memorial Institute of Children's Hematology, Oncology, and Transplantation, and various hematology patients undergoing HLA SBT testing for planned allogeneic hematopoietic stem cell transplantation.

Genomic DNA was isolated from peripheral blood leukocytes using MagNA Pure System (Roche Life Science). The target DNA concentration was from 10 to 140 ng/ μ L. Quantity and quality estimation of the isolated DNA was performed with Quantus Fluorometer TM (Promega Corp., USA).

The main steps of the NGS as performed with Illumina platform (MiSeq, USA) using NGSgo protocol were as follows:

1. HLA locus-specific amplification: the complete sequences of HLA genes are amplified with allele-specific primers in a single reaction for each locus using Long-Range DNA polymerase;
2. DNA quantification by Quantus Fluorometer and pooling of amplicons according to the volumes calculated by the NGSgo Pooling Calculation Sheet (provided by GenDx, Netherlands);

3. Double-stranded DNA fragmentation by means of specific fragmentase optimized by its size for the specific HLA locus, end repair, 5' phosphorylation of poly-A and poly-T ends and adapter ligation (Fig. 2);
4. DNA cleanup and size selection with 0.45x SPRI beads using 80% ethanol (Beckman Coulter, AMPure XP);
5. Indexing PCR products using a unique combination of i5 and i7 primers for each sample;
6. Plate-based DNA cleanup and size selection with 0.6x SPRI beads using 80% ethanol (Beckman Coulter, AMPure XP);
7. Plate-based library pooling, library quantification performed using Qubit Fluorometer and loading to the NGS sequencer (MiSeq, Illumina, USA);
8. Next-generation sequencing by MiSeq and data analysis.

The libraries are sequenced on an Illumina NGS platform. The FastQ data can be analyzed with an HLA typing software package to determine the HLA typing (for example, NGS engine). To assign the HLA alleles, the software allows communicating with updated IMGT database (Fig. 3).

The NGS method allows performing sequencing of all exons in the A, B, and C HLA loci and three exons (from second to fourth) of DQB1 and DRB1 (Fig. 4) which, however, has its limitations. The allele imbalances can be observed in some rare cases:

- NGSgo HLA-DRB1: allele imbalances for DRB1*01, DRB1*04, and DRB1*14 alleles can occur in case of imbalanced amplification.
- NGSgo HLA-DRB4: allele imbalances for DRB2 exon 2 and exon 3 can occur in case of imbalanced DRB4 amplification. In the case of an HLA-DRB4 exon 3 amplicon dropout, limit the analysis to exon 2 only.
- NGSgo HLA-DRB3/4/5: allele imbalances for heterozygous DRB3/4/5 samples can occur in case of imbalanced amplicon pooling. Analysis of DRB3/4/5 has been optimized in NGSengine v2.1 (and higher), which applies a split-analysis of the individual DRB3/4/5 loci to improve HLA typing.

The main steps of Sanger sequencing when performed with Applied Biosystems Genetic Analyzer (USA) 3500xl genetic analyzer using PROTRANS HLA SBT Class I and Class II S4 (Hockenheim, Germany, <http://www.protrans.info/nano.cms/en/products/MainCatID/9/>). Single Allele, Allele-Group and Locus Specific Sequencing. Fourteen specific primer mixes pre-pipetted in 8 and 16 well strip, in order of sequencing the Exons 1, 2, 3 and 4 for Class I, Exon 2 for DRB1 and Exon 2, 3 for DQB1, according to the manufacturers' recommendations.



Figure 2. Gene library preparation with NGSgo – LibX and NGSgo – IndX referred from <https://www.gendx.com> [10]

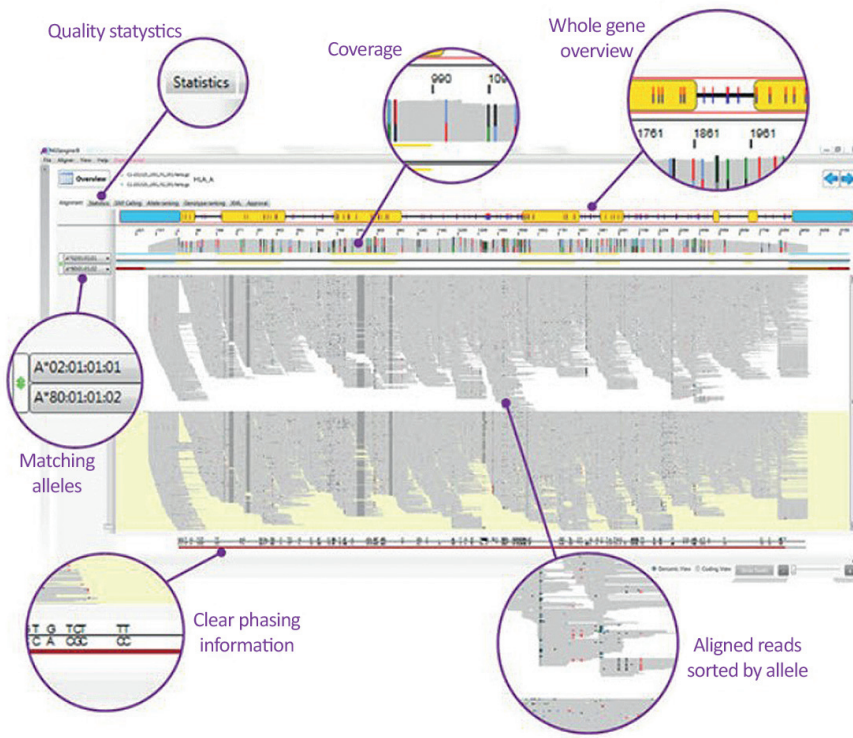


Figure 3. Data analysis software (NGSengine, <https://www.gendx.com>)

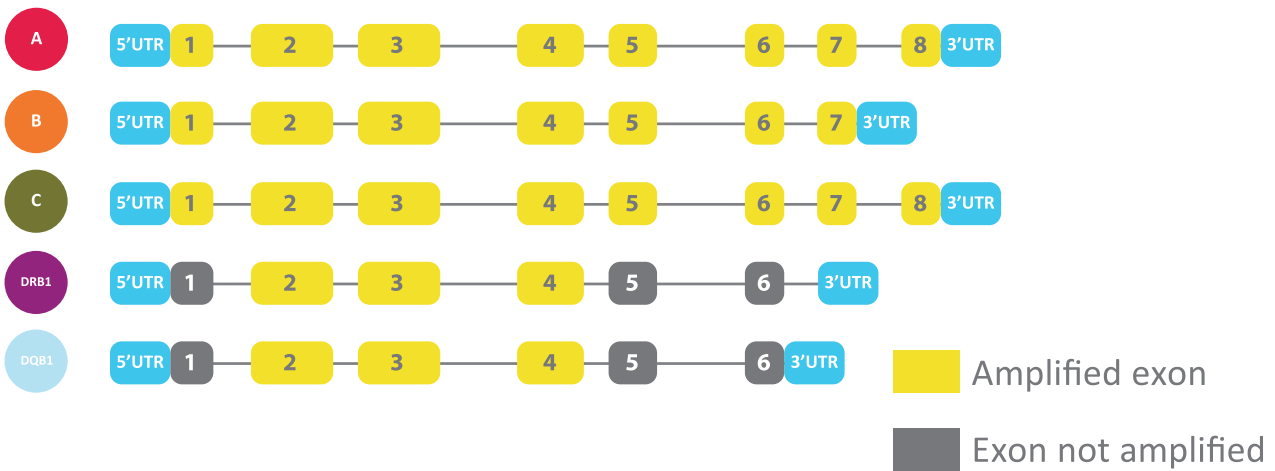


Figure 4. Target generation with NGSgo – AmpX [10]

Direct automated fluorescent DNA sequencing was performed by a 24-channel automated capillary electrophoresis system, and fluorescent detection of DNA fragments using an Applied Biosystems GA3500xl Genetic Analyzer. Capillary electrophoresis proceeded in the POP-7 polymer under denaturation conditions. The data on nucleotide sequences were retrieved at a stationary computer in the Data Collection program, then having been analyzed by Protrans SEQUENCE PILOT software (Hockenheim, Germany).

DNA amplification kits for Sanger sequencing are designed to provide high-resolution identification of alleles of the human HLA-A, -B, -C, -DRB1, -DQB1 genes.

Results

The aim of our pilot study was a comparison of two methods and an evaluation of their effectiveness. To achieve our purpose, we selected a group of 35 persons (see Materials and methods), and conducted analysis by Sanger and NGS method in parallel. The NGS method allowed detecting rare variants of alleles when performing data analysis with NGSengine software (Fig. 5).

We have conducted two sets of experiments. Mean coverage in the first experiment was 881x – (1010x, 897x, 768x, 807x, 923x, respectively for A-, B-, C-, DRB-, DQB- loci), and 992x

in a second experiment (1194x, 856x, 698x, 1001x, 1346x, respectively for A-, B-, C-, DRB-, DQB- loci).

Mean percentage of aligned reads to the total read number was 96.5% in the first set (DRB locus, 92.6%, other loci, >97.2%). In the second set, an appropriate percentage of aligned reads to the total read number was 95.0% (DRB locus, 91.6%, other loci, >95.5%) (Fig. 6). This metrics shows a

high quality of the sequencing that was performed according to the manufacturer's instructions.

To perform a more detailed analysis of each sample, the NGSengine software contains the sections of «typing results» and «visualization», where the coverage for different regions may be registered in more details, or a nucleotide position of interest should be found (Fig. 7).

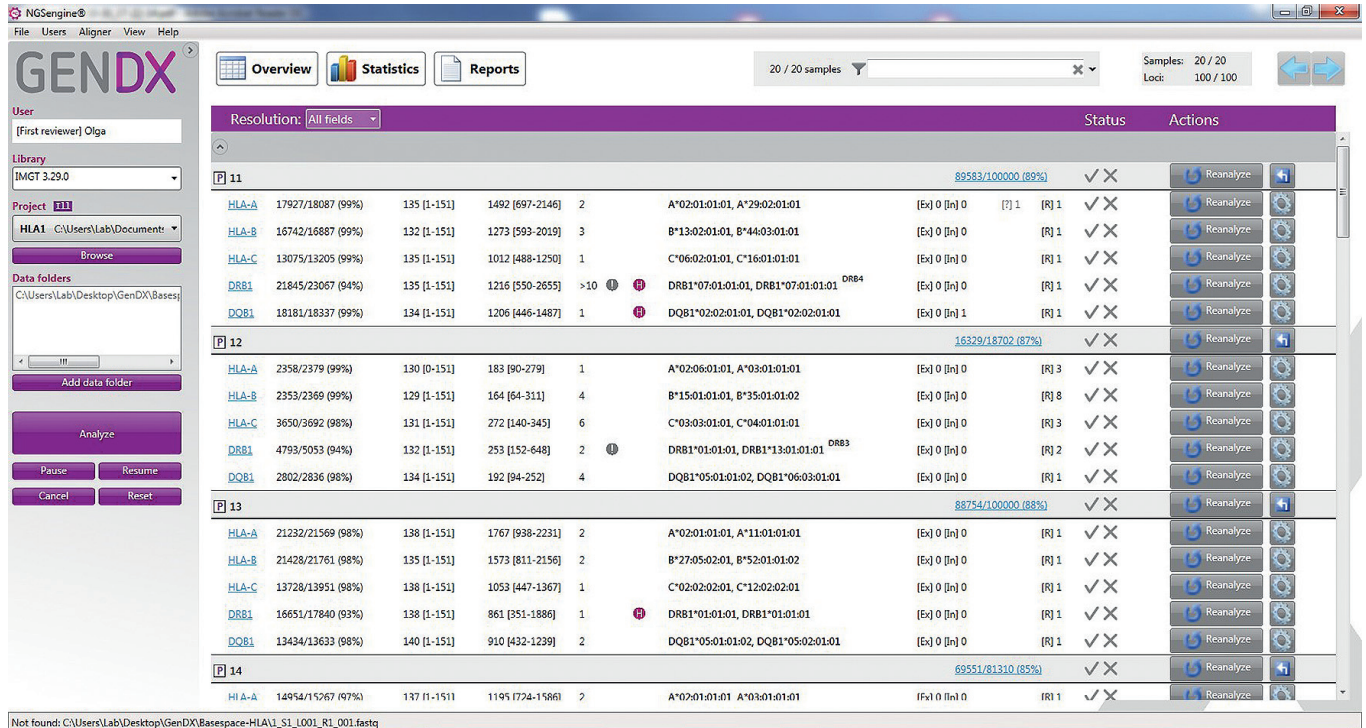


Figure 5. A typical data evaluation table presented by NGSengine software (Genome Diagnostics, Netherlands)

It presents information for each locus (HLA-A, -B, -C, -DRB1, -DQB1) for single samples. Data on total read number and percentage of aligned reads for the given locus, mean read length, mean coverage, alleles identified and presence of synonymous substitutions in coded [Ex] and it also displays non-coding [In] regions.



Figure 6. Statistics for samples (percentage aligned reads from total number of reads and number of reads mapped to the reference per strand) in NGSengine software

Typing result

	Allele 1	Allele 2
HLA-A	02:01:01:01	02:01:01:01
HLA-B	40:02:01:01	44:27:01
HLA-C	02:02:02:01	07:04:01:01
DRB1	11:01:01:01	16:01:01
DQB1	03:01:01:02	05:02:01:01

Allele ambiguities

Allele	CWD
A*02:01:01:01	No
: A*02:01:01:16	
A*02:01:01:01	No
: A*02:01:01:16	
B*40:02:01:01	No
: B*40:02:01:02	
: B*40:02:01:03	
C*07:04:01:01	No
: C*07:04:01:03	
DRB1*11:01:01:01	No
: DRB1*11:01:01:02	
: DRB1*11:01:01:03	
: DRB1*11:01:01:04	
DQB1*03:01:01:02	No
: DQB1*03:01:01:03	
: DQB1*03:01:01:09	
: DQB1*03:01:01:16	
: DQB1*03:01:01:18	

Visualization

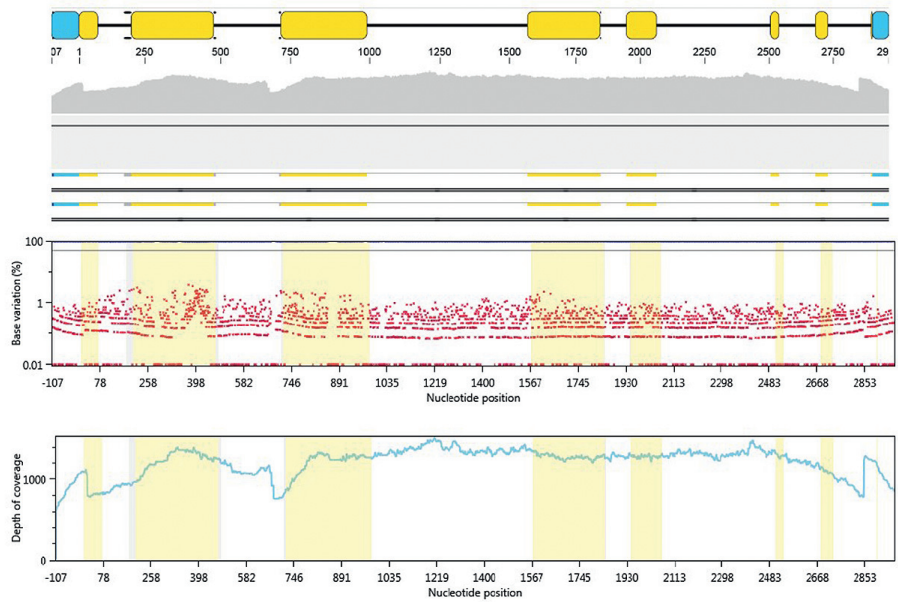


Figure 7. The results of the NGS sequencing all alleles for sample "3" and visualization of HLA-A locus for sample "3" in NGSengine software

It displays typing results for all HLA loci assayed. The cases of ambiguous results shown as *Allele Ambiguities*. In our series, no ambiguities were detectable for any locus. The figure visualization allows us to look at the visual segment (it shows exons in yellow). It indicates the sequencing coverage of the given locus below (marked gray). The vertical ticks seen at appropriate points of HLA loci in cases of synonymous nucleotide substitutions.

Table 1. Comparison of allele sequenced by NGS and Sanger's method – 100% homology results (there are no differences in 2nd and 3rd exons sequences)

Sample number	Locus	NGS approach		Sanger technique	
		Allele 1	Allele 2	Allele 1	Allele 2
1	HLA-A	01:01:01:01	02:01:01:01	01:01	02:01
	HLA-B	08:01:01:01	27:02:01:01	08:01	27:02
	HLA-C	02:02:02:03	07:01:01:01	02:02	07:01
	DRB1	03:01:01:01	08:01:01	03:01	08:01
	DQB1	02:01:01	04:02:01:01	02:01	04:02
2	HLA-A	02:01:01:01	66:01:01:01	02:01	66:01
	HLA-B	07:02:01:01	07:05:01:01	07:02	07:05
	HLA-C	07:02:01:03	15:05:02	07:02	15:05
	DRB1	10:01:01:01	15:01:01:01	10:01	15:01
	DQB1	05:01:01:02	06:02:01:01	05:01	06:02
3	HLA-A	02:01:01:01	02:01:01:01	02:01	02:01
	HLA-B	40:02:01:01	44:27:01	40:02	44:27
	HLA-C	02:02:02:01	07:04:01:01	02:02	07:04
	DRB1	11:01:01:01	16:01:01	11:01	16:01
	DQB1	03:01:01:02	05:02:01:01	03:01	05:02
5	HLA-A	01:01:01:01	11:01:01:01	01:01	11:01
	HLA-B	08:01:01:01	35:01:01:02	08:01	35:01
	HLA-C	04:01:01:05	07:01:01:01	04:01	07:01
	DRB1	03:01:01:01	15:01:01:01	03:01	15:01
	DQB1	02:01:01	06:02:01:01	02:01	06:02
6	HLA-A	01:01:01:01	02:01:01:01	01:01	02:01
	HLA-B	15:01:01:01	38:01:01	15:01	38:01
	HLA-C	03:03:01:01	12:03:01:01	03:03	12:03
	DRB1	15:01:01:01	13:01:01:01	13:01	15:01
	DQB1	06:02:01:01	06:03:01:01	06:02	06:03
7	HLA-A	23:01:01:01	23:01:01:01	23:01	23:01
	HLA-B	49:01:01:01	50:01:01:01	49:01	50:01
	HLA-C	06:02:01:02	07:01:01:01	06:02	07:01
	DRB1	03:01:01:01	11:01:01:01	03:01	11:01
	DQB1	02:01:01	03:01:01:02	02:01	03:01
8	HLA-A	03:01:01:01	23:01:01:01	03:01	23:01
	HLA-B	35:01:01:02	44:03:01:01	35:01	44:03
	HLA-C	04:01:01:01	04:01:01:01	04:01	04:01
	DRB1	01:01:01	07:01:01:01	01:01	07:01
	DQB1	02:02:01:01	05:01:01:02	02:02	05:01

Sample number	Locus	NGS approach		Sanger technique	
		Allele 1	Allele 2	Allele 1	Allele 2
9	HLA-A	01:01:01:01	25:01:01:01	01:01	25:01
	HLA-B	18:01:01:02	37:01:01	18:01	37:01
	HLA-C	06:02:01:01	12:03:01:01	06:02	12:03
	DRB1	15:01:01:01	15:01:01:01	15:01	15:01
	DQB1	06:02:01:01	06:02:01:01	06:02	06:02
10	HLA-A	02:01:01:01	24:02:01:01	02:01	24:02
	HLA-B	35:02:01:01	44:02:01:01	35:02	44:02
	HLA-C	04:01:01:06	05:01:01:02	04:01	05:01
	DRB1	11:01:01:01	11:137	11:01	11:137
	DQB1	03:01:01:02	03:01:01:02	03:01	03:01
11	HLA-A	03:01:01:01	31:01:02:01	03:01	31:01
	HLA-B	35:01:01:02	44:03:02	35:01	44:03
	HLA-C	04:01:01:01	07:06	04:01	07:01/07:06/07:18/07:343/07:419/07:458
	DRB1	11:01:01:01	15:02:01:01	11:01	15:02
	DQB1	03:01:01:02	06:01:01	03:01	06:01
12	HLA-A	02:01:01:01	26:01:01:01	02:01	26:01
	HLA-B	08:01:01:02	13:02:01:01	08:01	13:02
	HLA-C	06:02:01:01	07:02:01:01	06:02	07:02
	DRB1	03:01:01:01	07:01:01:01	03:01	07:01
	DQB1	02:01:01	02:02:01:01	02:01	02:02
13	HLA-A	11:01:01:01	33:01:01:01	11:01	33:01
	HLA-B	14:02:01:01	35:03:01:01	14:02	35:03
	HLA-C	08:02:01:01	12:03:01:01	08:02	12:03
	DRB1	01:02:01	04:08:01	01:02	04:08
	DQB1	03:04:01	05:01:01:01	03:04	05:01
14	HLA-A	02:01:01:01	29:02:01:01	02:01	29:02
	HLA-B	13:02:01:01	44:03:01:01	13:02	44:03
	HLA-C	06:02:01:01	16:01:01:01	06:02	16:01
	DRB1	07:01:01:01	07:01:01:01	07:01	07:01
	DQB1	02:02:01:01	02:02:01:01	02:02	02:02
15	HLA-A	02:01:01:01	29:02:01:01	02:01	29:02
	HLA-B	13:02:01:01	44:03:01:01	13:02	44:03
	HLA-C	06:02:01:01	16:01:01:01	06:02	16:01
	DRB1	07:01:01:01	07:01:01:01	07:01	07:01
	DQB1	02:02:01:01	02:02:01:01	02:02	02:02
16	HLA-A	02:06:01:01	03:01:01:01	02:06	03:01
	HLA-B	15:01:01:01	35:01:01:02	15:01	35:01
	HLA-C	03:03:01:01	04:01:01:01	03:03	04:01
	DRB1	01:01:01	13:01:01:01	01:01	13:01
	DQB1	05:01:01:02	06:03:01:01	05:01	06:03
17	HLA-A	02:01:01:01	11:01:01:01	02:01	11:01
	HLA-B	27:05:02:01	52:01:01:02	27:05	52:01
	HLA-C	02:02:02:01	12:02:02:01	02:02	12:02
	DRB1	01:01:01	01:01:01	01:01	01:01
	DQB1	05:01:01:02	05:02:01:01	05:01	05:02
18	HLA-A	02:01:01:01	03:01:01:01	02:01	03:01
	HLA-B	07:02:01:01	35:03:01:01	07:02	35:03
	HLA-C	04:01:01:01	07:02:01:03	04:01	07:02
	DRB1	11:01:01:01	13:01:01:01	11:01	13:01
	DQB1	03:01:01:02	06:03:01:01	03:01	06:03
19	HLA-A	02:01:01:01	30:01:01	02:01	30:01
	HLA-B	13:02:01:01	27:05:02:01	13:02	27:05
	HLA-C	02:02:02:01	06:02:01:01	02:02	06:02
	DRB1	01:01:01	07:01:01:01	01:01	07:01
	DQB1	02:02:01:01	05:01:01:02	02:02	05:01
20	HLA-A	03:01:01:01	68:01:27	03:01	68:01
	HLA-B	35:01:01:02	35:01:01:02	35:01	35:01
	HLA-C	03:03:01:01	04:01:01:01	03:03	04:01
	DRB1	01:01:01	08:01:01	01:01	08:01
	DQB1	04:02:01:01	05:01:01:02	04:02	05:01
21	HLA-A	03:01:01:01	24:02:01:01	03:01	24:02
	HLA-B	13:02:01:01	35:01:01:02	13:02	35:01
	HLA-C	04:01:01:01	06:02:01:01	04:01	06:02
	DRB1	01:01:01	07:01:01:01	01:01	07:01
	DQB1	02:02:01:01	05:01:01:02	02:02	05:01
22	HLA-A	01:01:01:01	02:01:01:01	01:01	02:01
	HLA-B	08:01:01:01	44:27:01	08:01	44:27
	HLA-C	07:01:01:01	07:04:01:01	07:01	07:04
	DRB1	03:01:01:01	16:01:01	03:01	16:01
	DQB1	02:01:01	05:02:01:01	02:01	05:02
23	HLA-A	01:01:01:01	02:01:01:01	01:01	02:01
	HLA-B	08:01:01:01	13:02:01:01	08:01	13:02
	HLA-C	06:02:01:01	07:01:01:01	06:02	07:01
	DRB1	03:01:01:01	07:01:01:01	03:01	07:01
	DQB1	02:01:01	02:02:01:01	02:01	02:02

EXPERIMENTAL STUDIES

Sample number	Locus	NGS approach		Sanger technique	
		Allele 1	Allele 2	Allele 1	Allele 2
24	HLA-A	02:01:01:01	03:01:01:05	02:01	03:01
	HLA-B	07:02:01:01	57:01:01	07:02	57:01
	HLA-C	06:02:01:01	07:02:01:03	06:02	07:02
	DRB1	01:01:01	07:01:01:01	01:01	07:01
	DQB1	03:03:02:01	05:01:01:02	03:03	05:01
26	HLA-A	02:01:01:01	03:01:01:01	02:01	03:01
	HLA-B	07:02:01:01	18:01:01:02	07:02	18:01
	HLA-C	07:01:01:01	07:02:01:03	07:01	07:02
	DRB1	15:01:01:01	16:01:01	15:01	16:01
	DQB1	05:02:01:01	06:02:01:01	05:02	06:02
27	HLA-A	01:01:01:01	24:02:01:01	01:01	24:02
	HLA-B	08:01:01:01	18:01:01:02	08:01	18:01
	HLA-C	07:01:01:01	07:01:01:01	07:01	07:01
	DRB1	03:01:01:01	11:04:01	03:01	11:04
	DQB1	02:01:01	03:01:01:02	02:01	03:01
28	HLA-A	02:01:01:08	23:01:01:01	02:01	23:01
	HLA-B	38:01:01	49:01:01:01	38:01	49:01
	HLA-C	07:01:01:01	12:03:01:01	07:01	12:03
	DRB1	11:01:01:01	13:01:01:01	11:01	13:01
	DQB1	03:01:01:02	06:03:01:01	03:01	06:03
CET1	HLA-A	24:02:01:01	24:02:01:01	24:02/24:353	-
	HLA-B	44:02:01:01	44:03:01:02	44:02	44:03
	HLA-C	05:01:01:02	02:02:02:01	05:01/05:145	02:02
	DRB1	01:02:01	13:10	01:02/01:83	13:10
	DQB1	05:01:01:01	06:03:01:01	05:01	06:03
CET2	HLA-A	24:07:01	26:01:01:01	24:07	26:01
	HLA-B	07:06:01	15:02:01	07:05/07:06	15:02
	HLA-C	07:02:01:01	08:01:01:01	07:02	08:01
	DRB1	11:05	14:04:01	11:05	14:04
	DQB1	05:03:01:01	06:02:01:01	05:03/15:149	06:02
CET3	HLA-A	11:01:01:01	24:02:01:01	11:01/11:263	24:02/24:353
	HLA-B	40:06:01:02	55:01:01	40:06	55:01/55:85
	HLA-C	03:03:01:01	15:02:01:01	03:03/03:227/03:341/03:357	15:02/15:87
	DRB1	14:04:01	14:54:01:01	14:04:01	14:01/14:54
	DQB1	05:03:01:01	05:03:01:01	05:03/15:149	-
CET4	HLA-A	02:01:01:01	02:64:01	02:01/02:665/02:686/02:689	02:64
	HLA-B	39:06:02:03	51:01:01:01	39:06	51:01/51:193/51:224
	HLA-C	01:02:01:01	12:03:01:01	01:02/01:85/01:127/01:142	12:03/12:143/12:167
	DRB1	14:05:01:01	16:01:01	14:05	16:01
	DQB1	05:02:07	05:03:01:01	05:02	05:03
CET5	HLA-A	02:01:01:01	02:01:01:01	02:01/02:665/02:686/02:689	-
	HLA-B	07:02:01:01	44:02:01:01	07:02/07:61	44:02
	HLA-C	05:01:01:02	07:02:01:03	05:01/05:145	07:02/07:50/07:349/07:566/07:592/07:594/07:595/07:596
	DRB1	04:01:01:02	15:01:01:01	04:01	15:01/15:141/15:145/15:146
	DQB1	03:02:01:01	06:02:01:01	03:02/03:32/03:85/03:190/03:245/03:247/03:251/03:263/03:265	06:02/06:47/06:84/06:109/06:111/06:116/06:117/06:127/06:188/06:200/06:219/06:224/06:225/06:226/06:227/06:228/06:237/06:240
CET7	HLA-A	33:03:01	68:02:01:01	33:03	68:02/68:163
	HLA-B	08:01:01:01	58:01:01:01	08:01/08:173/08:183	58:01
	HLA-C	03:02:02:01	03:04:01:02	03:02	03:04/03:358/03:359
	DRB1	13:04	13:04	13:04	-
	DQB1	03:19:01	03:19:01	03:19	-
CET8	HLA-A	02:05:01:01	74:01:01	02:05	74:01
	HLA-B	15:03:01:02	15:10:01	15:03	15:10
	HLA-C	02:10:01:01	02:10:01:02	02:10	-
	DRB1	11:01:02	11:04:02	11:01	11:04
	DQB1	03:19:01	05:02:01:01	03:01/03:09/03:19/03:21/03:22/03:24/03:29/03:35/03:42/03:49/03:50/03:51/03:94/03:115/03:116/03:164/03:165/03:169/03:182/03:191/03:196/03:198/03:206/03:241/03:243/03:246/03:353/03:264/03:266	05:02/05:14/05:17/05:35/05:36/05:37/05:46/05:47/05:57/05:102/05:106/05:136
CET9	HLA-A	33:03:01	33:03:01	33:03	
	HLA-B	44:03:02	56:01:01:03	44:03	56:01
	HLA-C	01:02:01:01	07:06	01:02/01:85/01:127/01:142	07:01/07:06/07:18/07:343/07:419/07:458/07:591
	DRB1	03:01:01:01	07:01:01:01	03:01/03:124/03:132/03:137	07:01/07:34/07:72/07:79
	DQB1	02:01:01	02:02:01:01	02:01	02:02/02:97
CET10	HLA-A	34:02:01	36:01	34:02	36:01
	HLA-B	35:01:01:02	53:01:01	35:01/35:332	53:01
	HLA-C	04:01:01:01	04:01:01:01	04:01/04:40/04:82/04:226	
	DRB1	11:01:02	12:01:01:01	11:01	12:01/12:06/12:10/12:17
	DQB1	05:01:01:02	06:02:01:01	05:01	06:02

Factors contributing to the costs arising for the in-depth sequencing

The reagents for the entire HLA-sequencing process include those used for routine pre-analytic steps (e.g., DNA extraction, quality assessment, and initial low-resolution typing step). Additional expenditures are subject to some ambiguities, due to different prices for reagents and equipment offered by distinct manufacturers. Moreover, it should be addressed that all the commercial NGS platforms offer their closed-type systems, thus causing broad variations in prices for the entire NGS procedure *per* single DNA sample, strongly depending on the annual capacity of the given HLA typing laboratory.

However, even considering maintenance costs (about 10% equipment cost), enrolling third-party core facilities or shared equipment, the Sanger sequencing (220 K) proves to be twice more expensive than NGS (variable, but still less than Sanger technique), as shown elsewhere [11]. Hence, the sample preparation costs remain the same whereas the sequencing tends to decrease as discussed in [9].

Calculations of economic efficiency for HLA typing in Russia by Sanger technique *versus* NGS were among our major tasks. Therefore, we have performed a pricing for the sequencing kits at a company providing reagents to this purpose. The request was made twice (February 2017 and August 2018). The reagent price in Euros did not change sufficiently. Of note, the sequencing kits by Sanger are produced for 25 or 100 tests, whereas NGS kits are offered for 24 and 96 tests.

Clear benefit of Sanger approach is that a single locus may be sequenced in the sample, being, however, economically ineffective when using NGS technology.

Hence, we have compared the panels for 100 tests covering five main HLA loci, i.e., AlleleSEQR HLA-A PCR/Sequencing Mix, AlleleSEQR HLA-B PCR/Sequencing Mix, AlleleSEQR HLA-C Plus PCR/Sequencing Mix, AlleleSEQR HLA-DRB1 PCR/Sequencing Mix, AlleleSEQR HLA-DQB1 PCR/Sequencing Mix). Each set was purchased for 6250 Euros. Hence, the total cost of locus-specific reagents for 5 loci was 31250 Euros, thus providing 312.50 Euros per 1 human DNA sample (ca. 24,000 roubles as per October 2018). The prime costs should also include disposables for the core sequencing procedure. E.g., if performing 100 tests for 5 HLA loci, we run 2500 reactions with a 24-channel Applied Biosystems GA3500xl Genetic Analyzer ABI 3500. To start the process, the following general items are needed: 26 plates for the gene analyzer (MicroAmp 96 Well Reaction Plate); three universal polymers. For capillary electrophoresis, POP-7 for 960 samples, five Formamide packs (25 mL each), containers with anode and cathode buffers etc., at a total price of 4500 Euros. Hence, the prime costs of Sanger reagents, when sequencing 5 main loci at the full-load regimen and usage of a 100-test kit, makes about 35750 Euros (ca. 2681250 roubles), excluding costs for pipette tips, microtubes, gloves and other inexpensive disposables. That means ca. 27000 roubles *per* one human DNA sample.

When applying NGS approach, the number of samples taken into analysis is quite sufficient, since even a high-throughput

MiSeq machine may perform sequencing of up to 269 samples in parallel, using a standard 4.5-Gb cartridge.

To calculate costs of comparable NGS analysis, we have chosen a reagent set for sequencing of 96 samples which included the following items: NGSgo[®]-AmpX HLA-A, B, C, DRB1, DQB1, NGSgo[®]-LibrX Library Preparation (2 kits), NGSgo[®]-IndX Adapter & Indices (4x24) RUO Illumina, Agencourt AMPure XP 5 mL Kit, GenDx LongRange polymerase (3 kits), MiSeq Reagent Kit v2. A total sum for typing 5 loci made ca. 14 800 Euros, thus comprising 155 Euros (>10000 Roubles) *per* sample.

However, in case of spared use of the reagents for sample preparation by two-fold decrease in reaction volume (as proven by our experience), the prime costs *per* a single test dropped to 90 Euros. The self-cost is here provided without accessory disposables.

Hence, the costs of sequencing reagents, even without their sparing, is sufficiently cheaper when using NGS technologies. Moreover, the procedure takes 2-fold less time for its performance than the Sanger technique.

Conclusion

Comparison of allele sequenced by NGS and Sanger's method yielded 100% homology results. Hence, our work is in accordance with previously published data [9], which demonstrate the advantage and efficiency of NGS, as compared to Sanger sequencing.

NGS-based HLA analysis is performed with a 100% reliability, and well fits the tasks of HLA typing in unrelated donors, in concordance with EFI and ASHI policies. This work process well corresponds to the working schedules for medium- and high-capacity laboratories, thus being potentially attractive to the donor registries. Recently introduced next-generation sequencing techniques have a facilitating potential for the high-resolution genotyping *via* a decrease of general ambiguity of end results, like as due to more extended sequencing regions. In near future, the NGS approaches will be an effective and cost-effective technology when evaluating histocompatibility parameters and immunogenetic interactions.

Conflict of interest

The authors have declared no conflicting interests.

Funding

This research was funded by Russian Science Foundation grant №14-50-00069.

References

1. Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*. 2015; 43:D423-431.
2. Marsh SGE, Albert ED, Bodmer W, Bontrop RE, Dupont B, Erlich HA, Fernández-Viña M, Geraghty DE, Holdsworth

R, Hurley CK, Lau M, Lee KW, Mach B, Maiers M, Mayr WR, Müller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, Svejgaard A, Terasaki PI, Tiercy JM, Trowsdale J. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010; 75(4):291-455.

3. Marsh SGE. Nomenclature for factors of the HLA system, update July 2017. *Human Immunol*. 2017; 78(11-12):758-761.

4. Holcomb CL, Höglund B, Anderson MW, Blake LA, Böhme I, Egholm M, Ferriola D, Gabriel C, Gelber SE, Goodridge D, Hawbecker S, Klein R, Ladner M, Lind C, Monos D, Pando MJ, Pröll J, Sayer DC, Schmitz-Agheguian G, Simen BB, Thiele B, Trachtenberg EA, Tyan DB, Wassmuth R, White S, Erlich HA. A multi-site study using high-resolution HLA genotyping by next generation sequencing. // *Tissue Antigens*. 2011;77(3):206-217. doi: 10.1111/j.1399-0039.2010.01606.x.

5. Mack SJ1, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, Moraes ME, Pereira SE, Kempenich JH, Reed EF, Setterholm M, Smith AG, Tilanus MG, Torres M, Varney MD, Voorter CE, Fischer GF, Fleischhauer K, Goodridge D, Klitz W, Little AM, Maiers M, Marsh SG, Müller CR, Noreen H, Rozemuller EH, Sanchez-Mazas A, Senitzer D, Trachtenberg E, Fernandez-Vina M. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens*. 2013 Apr;81(4):194-203. doi: 10.1111/tan.12093.

6. Sanchez-Mazas, J. M. Nunes, D. Middleton, J. Sauter, S. Buhler, A. McCabe, J. Hofmann, D. M. Baier, A. H. Schmidt, G. Nicoloso, M. Andreani, Z. Grubic, J.-M. Tiercy, K. Fleischhauer. Common and well-documented HLA alleles over all of Europe and within European sub-regions: A catalogue from the European Federation for Immunogenetics. *HLA Volume 89, Issue 2, February 2017 Pages 104-113*.

7. S. G. E. Marsh, E. D. Albert, W. F. Bodmer, R. E. Bontrop, B. Dupont, H. A. Erlich, M. Fernandez-Vina, D. E. Geraghty, R. Holdsworth, C. K. Hurley, M. Lau, K. W. Lee, B. Mach, M. Maiers, W. R. Mayr, C. R. Müller, P. Parham, E. W. Petersdorf, T. Sasazuki, J. L. Strominger, A. Svejgaard, P. I. Terasaki, J. M. Tiercy & J. Trowsdale. Nomenclature for factors of the HLA system, 2010 *Tissue Antigens* 75, 291-455.

8. Kuzmich EV, Alyanskiy AL, Tyapushkina SS, Nasredinova AA, Ivanova NE, Zubarovskaya LS, Afanasyev BV. Identification of the new HLA-B*44:02:45, DQB1*02:85, DQB1*06:210, DRB1*01:01:30 alleles by monoallelic Sanger sequencing. *Cell Ther Transplant*. 2018. 7(1):62-66.

9. Serov YA, Barkhatov IM, Klimov AS, Berkos AS. Current methods and opportunities of next-generation sequencing (NGS) for HLA typing // *Cell Ther Transplant*. 2016; 5(4): 63-70. doi: 10.18620/ctt-1866-8836-2016-5-4-63-70.

10. <https://www.gendx.com>

11. Baxter-Lowe LA. Tailoring NGS for smaller volume labs. *Proc. 42nd ASHI Annual Meeting. Abstract: Sept 28, 2016*.

Сравнительный анализ методов секвенирования NGS и по Сэнгеру при HLA-типировании в российской университетской клинике

Олег С. Плотов^{1,2}, Ольга В. Романова^{1,2}, Юрий А. Эйсмонт¹, Андрей М. Сарана^{1,2}, Сергей Г. Щербак^{1,2}, Елена В. Кузьмич³, Александр Л. Алянский³, Наталья Е. Иванова³, Вера В. Тепляшина³, Юрий А. Серов³, Людмила С. Зубаровская³, Борис В. Афанасьев³

¹ Городская больница №40, Сестрорецк, Санкт-Петербург, Россия

² Институт трансляционной биомедицины, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

³ НИИ детской онкологии, гематологии и трансплантологии им. Р. М. Горбачевой, Первый Санкт-Петербургский государственный медицинский университет, Санкт-Петербург, Россия

Резюме

База данных Всемирной организации здравоохранения (ВОЗ) Комитета по номенклатуре факторов системы HLA (база данных IPD-IMGT/HLA) на сентябрь 2018 г. содержала информацию о нуклеотидных последовательностях 20272 различных аллелей HLA, из которых 14800 были аллелями HLA класса I, а 5288 – класса II.

На протяжении последних 20 лет при секвенировании генома человека, животных, бактерий и вирусов преобладает автоматизированная технология Сэнгера. Однако необходимость более быстрого скрининга генома стимулировало развитие новых технологий мультиплексного секвенирования ДНК. Эти современные методы обозначаются как подходы следующего поколения (Next-Generation Sequencing, NGS).

Целью нашего исследования было сравнение двух этих методов и оценка их эффективности. Чтобы достичь этой цели, мы выбрали группу из 35 образцов ДНК, в основном – потенциальных доноров гемопоэтических клеток, и провели сравнительный анализ по Сэнгеру и методом NGS. Метод NGS позволяет выявлять редкие или новые варианты аллелей. Этот подход подтвержден в качестве более чувствительного и более экономичного, особенно в больших лабораториях по HLA-типированию.

Ключевые слова

Главный комплекс гистосовместимости, новые аллели HLA, технологические решения, секвенирование следующего поколения, NGS, секвенирование по Сэнгеру, трансплантация гемопоэтических клеток, типирование по сиквенсам ДНК.