Л.Н. Гумилев атындағы Еуразия ұлттық университетінің ХАБАРШЫСЫ.
BULLETIN of L.N. Gumilyov Eurasian National University.
ВЕСТНИК Евразийского национального университета имени Л.Н. Гумилева.
ISSN: 2616-7034. eISSN: 2663-130X

IRSTI 34.15.23
Review Article

https://doi.org/10.32523/2616-7034-2025-153-4-94-117

HLA system nomenclature and discovery of novel allelic variants in the Kazakh population

A. Turganbekova[1,2], S. Abdrakhmanova[2], W.Y. Almawi*[3,4]

[1]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*
[2]*Scientific and Production Center for Transfusiology, Astana, Kazakhstan*
[3]*Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada*
[4]*Faculty of Sciences, El-Manar University, Tunis, Tunisia*

*(E-mail: aidaturganbekova84@gmail.com, a.saniya@mail.ru, *wassim.almawi@fst.utm.tn)*

Abstract. The human leukocyte antigen (HLA) system is among the most genetically diverse in humans, encompassing over 220 genes that encode immune proteins essential for transplant compatibility, immune regulation, and disease susceptibility. This review outlines the fundamentals of HLA nomenclature, standardized by the World Health Organization (WHO) and curated in the IPD-IMGT/HLA database. We describe the gradual improvements in HLA typing methods, ranging from serological assays to molecular-based techniques, including Sanger sequencing and next-generation sequencing (NGS), as well as interpretation software, and evaluate their strengths and limitations in allele discovery. We discuss allelic variants of HLA genes, methods for sequencing HLA alleles, and their variants. Additionally, we report the identification of four novel HLA alleles in the Kazakh population: *DQB1*03:82*, *C*06:256*, *B*13:150*, and *A*32:95*. All four alleles feature non-synonymous substitutions within peptide-binding domains, suggesting potential immunological relevance. Comparative analysis reveals that NGS enhances allele detection efficiency by 2.8-fold compared to Sanger sequencing (one novel allele per 635 typings vs. 1,773). These findings demonstrate the significant HLA diversity present in Central Asian populations, which remain underrepresented in global databases. The identification of population-specific alleles reveals critical gaps in international donor databases, underscoring the urgent need to expand HLA profiling in ethnically diverse regions to improve transplant outcomes and advance personalized immunotherapy.
Keywords. Allelic variants, HLA typing, IPD-IMGT/HLA database, Kazakhstan, Next-generation sequencing

Introduction

The genes encoding human leukocyte antigens (HLA) are known for their high level of polymorphism, making the systematic classification of HLA genes essential. Located on the short arm of chromosome 6 at position 6p21.3, the HLA complex includes over 220 genes, many of which play crucial roles in immune function. Oversight of HLA allele naming and quality

Received: 08.10.2025. Accepted: 12.12.2025. Available online: 25.12.2025.

**Corresponding author*

control is managed by the WHO Nomenclature Committee for Factors of the HLA System [1]. Comprising both classical (HLA-A, -B, -C, -DRB1, -DQB1, -DPB1) and non-classical loci, the HLA complex displays remarkable genetic diversity that has evolved under pathogen-driven selective pressures. This diversity enables populations to respond to various infectious challenges, although it also makes clinical applications such as transplantation and disease association studies more complex.

Understanding HLA diversity across global populations is essential for several reasons. First, successful hematopoietic stem cell transplants depend on close HLA matching between donors and recipients, yet European and North American populations are overrepresented in donor registries. Second, certain HLA alleles are associated with either increased risk or protection against autoimmune diseases, infectious diseases, and adverse drug reactions, making population-specific HLA analysis important for medical purposes. Third, the extensive polymorphism of HLA genes provides a valuable model for studying balancing selection and host-pathogen coevolution.

Established in 1968, the committee set the initial criteria for nomenclature and has since met regularly, publishing 19 key reports that have progressed from identifying HLA antigens to documenting genes and alleles. Standardization efforts were traditionally supported through the exchange of reagents and reference cells via the International Histocompatibility Workshops. The systematic naming of HLA allele sequences began in 1989, marking a significant milestone in immunogenetics. Since then, the development and maintenance of a centralised sequence database have become essential. The assignment and distribution of allele names are particularly important in clinical settings. Thanks to the efforts of the HLA biinformatics team, in collaboration with the European Bioinformatics Institute (EBI), this information is publicly accessible through both the EBI website (http://www.ebi.ac.uk/ipd/imgt/hla) and http://hla.alleles.org [2-4].

The IPD-IMGT/HLA database compiles newly identified and validated HLA sequences, which undergo expert curation and approval by the WHO Nomenclature Committee. Only sequences meeting strict quality standards are incorporated into the database and associated resources available on the official platforms. Continuous updates ensure that researchers and clinicians worldwide have access to the latest HLA allele data [5].

Molecular genetic diagnostics has become one of the fastest-growing fields in the detection and management of various diseases. The introduction of the polymerase chain reaction (PCR) technology, a method of specific DNA amplification developed over the past two decades, marked a significant turning point in the clinical application of genetic testing [6]. Today, next-generation sequencing (NGS) is widely employed for the diagnosis and prediction of graft rejection in organ and hematopoietic stem cell transplantation. This technology is also used to detect leukocyte antigens associated with specific diseases. In these applications, the patient's polymorphic HLA variants are compared with those of potential donors. A donor with a fully or closely matched HLA genotype is selected to reduce the risk of transplant rejection and improve the overall success of transplantation procedures [8].

History

The role of the major histocompatibility complex (MHC) in allograft rejection was first suggested by Bover [9], who noted that skin grafts between identical twins were not rejected in the same manner as those from genetically dissimilar individuals. The MHC genes implicated in the allograft rejection process were initially characterized in mice by Gorer [10].

Following this, Snell [11] utilized mouse cell lines to further delineate a locus known as H for histocompatibility. Gorer [10] referred to the gene products of locus H as antigens II, leading to the combined designation H-2 for the mouse MHC. The human leukocyte antigen (HLA) system was subsequently identified in the 1950s. Multiple researchers independently found that sera from individuals who had received previous blood transfusions and from multiparous women contained antibodies that agglutinated leukocytes [12]. This finding led to the development of serological typing methods that identified a single locus, which was later divided into two loci: HLA-A and HLA-B [13-16].

Initially, various techniques were employed for serological typing; however, the microlymphocytotoxicity assay became the most common method. It was later observed that when lymphocytes from unrelated individuals matched at the HLA-A and HLA-B loci and were cultured together in a "mixed lymphocyte culture" (MLC) [17,18], they exhibited a strong proliferative response. This led to the discovery of an additional locus initially called HLA-D [19-21]; it was later shown that mismatches at HLA-DR and HLA-DQ contributed to the lymphocyte activation seen in the MLC. Extensive serological studies soon revealed the existence of HLA-C [22]. Later, HLA-DP, originally known as the "secondary B cell" (SB) antigen, was discovered through a secondary stimulation assay called the "primed lymphocyte test" (PLT), which indicated recognition of another HLA molecule distinct from those identified in the primary mixed lymphocyte culture [23,24].

The IPD-IMGT/HLA database and nomenclature committee

Since its launch in 1998, the IPD-IMGT/HLA database has served as a central repository for information on immune system gene polymorphisms. The initial version included 964 HLA allelic variants, and the database has since expanded considerably, becoming a vital reference resource for HLA researchers and clinicians [5]. In addition to storing allele sequences, the database provides comprehensive metadata about the biological source of each sequence and the methods used for sequence validation. Today, it is standard practice for researchers to submit newly identified sequences directly to the IPD-IMGT/HLA database for expert review and official naming before publication.

This approach helps prevent confusion caused by the renaming of already published sequences or the use of multiple identifiers for the same allele [3]. The growing importance of timely reporting of novel HLA allele sequences has led to the regular convening of the WHO Nomenclature Committee for HLA Factors, with annual meetings dedicated to maintaining the consistency and accuracy of HLA nomenclature. To ensure rapid access to newly approved sequences, monthly nomenclature updates are also published online and in scientific journals.

In collaboration with the Imperial Cancer Research Fund (ICRF; now part of Cancer Research UK) and the European Bioinformatics Institute (EBI), a sophisticated Oracle-based database was developed. This system enables users to perform complex queries and access detailed data on sequence features, references, contact information, and official allele designations via a user-friendly graphical web interface. The initial creation of this database was supported by European Union BIOMED1 (BIOMED1-930038) and BIOTECH2 (BIOTECH-960037) grants awarded to the ICRF as part of the International Immunogenetics (IMGT) initiative.

Development and curation of the HLA database were carried out in collaboration with Julia Bodmer (ICRF), James Robinson (formerly at ICRF, now with the HLA Informatics Group), and Peter Parham (Stanford University). The WHO Nomenclature Committee is currently chaired by Professor Steven G.E. Marsh, and its headquarters are located at the Anthony Nolan Research

Institute in London, UK [25]. Since its establishment, the committee has held 18 international meetings, covering a wide range of topics, from classical serological HLA typing methods to advanced molecular genetic techniques, including mixed lymphocyte reactions, PCR-based DNA typing, and next-generation sequencing (NGS). A summary of these international workshops is provided in Table 1, detailing the seminars organized under the auspices of the WHO Nomenclature Committee for HLA.

Table 1

International seminars of the WHO nomenclature committee for HLA

Seminar	Year	Committee Chair	Location	Seminar topic
1	1964	DB Amos	Durham, USA	Determination of the specificity of antigens Hu-1, LA, and Four
2	1965	JJ van Rood	Leiden, The Netherlands	Testing of mixed lymphocyte culture.
3	1967	R Ceppellini	Torino, Italy	Family studies; HLA in kidney transplantation
4	1970	PI Terasaki	Los Angeles, USA	Determination of the specificity of 27 HLA-A, HLA-B, and HLA-C
5	1972	J Dausset	Evian, France	Typing of 49 populations worldwide
6	1975	F Kissmeyer-Nielsen	Aarhus, Denmark	Description of the characteristics of Dw.
7	1977	WF Bodmer	Oxford, UK	Determination of the characteristics of DR1-7; HTC testing
8	1980	PI Terasaki	Los Angeles, USA	Determination of MB (DQ) and MT (DR52/53); HLA in transplantation and diseases.
9	1984	EA Albert/W Mayr	Munich, Germany Vienna, Austria	New features of class I and II; HLA class II in kidney transplantation
10	1987	B Dupont	Princeton, USA	Creation of RFLP; T-cell clones; GTK methods; Biochemistry of 1D IEF, 2D gels; Creation of a panel of homozygous cell lines.
11	1991	T Sasazuki/K Tsuji/M Aizawa	Yokohama, Japan	DNA PCR typing of HLA class II; Anthropology.
12	1996	D Charron	St Malo/Paris, France	DNA PCR typing of HLA class I; Anthropology.
13	2002	J Hansen	Victoria, Canada Seattle, USA	Virtual DNA analysis; Identification of SNP markers; Anthropology; Disease association; GWAS (Genome-Wide Association Studies)
14	2005	J McCluskey	Melbourne, Australia	MHC and anthropology; Disease; Infectious disease; GWAS (Genome-Wide Association Studies); Cancer; KIR (Killer Immunoglobulin-like Receptors); Cytokine genes.

15	2008	M Gerbase de Lima/ME Moraes	Buzios/Rio de Janerio, Brazil	Anthropology; GWAS (Genome-Wide Association Studies); Informatics.
16	2012	SGE Marsh/D Middleton	Liverpool, UK	NGS (Next-Generation Sequencing), GWAS (Genome-Wide Association Studies)
17	2012	M Fernandez-Viña	Asilomar, USA	NGS
18	2021	S Heidt/E Spierings	Amsterdam, The Netherlands	Topic under clarification

Note: *(website source https://hla.alleles.org/nomenclature/workshops.html)

The International ImMunoGeneTics Information System (IMGT) is a comprehensive set of databases and bioinformatics tools dedicated to immunogenetics and immunoinformatics. It focuses on the sequences of V (variable), D (diversity), J (joining), and C (constant) genes, which are crucial components of the adaptive immune system. IMGT was established in June 1989 by Marie-Paule Lefranc, an immunologist at the University of Montpellier. The initiative was formally introduced at the 10th Human Genome Mapping Workshop, where the V, D, J, and C regions were officially recognized as genes. The first IMGT database, IMGT/LIGM-DB, was created to compile nucleotide sequences of human immunoglobulins and T-cell receptors, later expanding to include sequences from various vertebrate species. The project was launched at the Laboratoire d'Immunogénétique Moléculaire at the University of Montpellier and supported by the French National Centre for Scientific Research (CNRS) [26].

Given that T-cell receptors and immunoglobulins are generated through somatic recombination of nucleotide segments, the genomic annotation of these gene regions poses unique challenges. To address this, IMGT-NC, the nomenclature committee, was formed in 1992 to provide standardized terminology. This subcommittee is officially recognized by the International Union of Immunological Societies [1]. In addition to its databases, IMGT offers several key tools:

• IMGT/Collier-de-Perles: provides a two-dimensional graphical representation of receptor amino acid sequences.

• IMGT/mAb-DB: a curated database of monoclonal antibodies.

• IPD-IMGT/HLA database: a critical resource for HLA allele data maintained by the HLA Informatics Group, was developed in part through IMGT's efforts and remains integrated within its broader infrastructure [2,25].

Polymorphisms

The HLA system is the most polymorphic genetic system in the human genome. HLA polymorphisms were first identified phenotypically by observing acceptance or rejection of tissue and/or through reactions with specific alloantibodies (serological typing methods). Later, molecular typing methods revealed HLA polymorphisms that range from a single nucleotide change to the loss or gain of an entire genetic region. Initially, HLA polymorphisms were identified using serological and cell proliferation assays [14-16, 21]. These methods were primarily used to characterise the HLA system; however, despite their widespread use, they have notable limitations in accuracy and reproducibility [26]. Moreover, alloimmune antisera are often limited in supply, and both serological and cellular HLA typing require live cells. Overall, a key limitation of serological HLA typing is its inability to detect minor polymorphic differences that can activate CD4+ or CD8+ T lymphocytes.

A general characteristic of the HLA genes is that the distal membrane domains are highly polymorphic. In contrast, the proximal membrane domains, the transmembrane, and cytoplasmic domains have very low or no polymorphisms. The heavy chain of the HLA class I molecule is composed of 3 extracellular domains, and both the α and β chains of the HLA class II molecule contain 2 extracellular domains (Figure 3). Of note, the HLA genes, like all eukaryotic genes, contain both coding (exons) and non-coding (introns) regions. The HLA class I genes contain 8 exons, while the HLA class II genes contain 6 or 7 exons [27,28]. The widespread application of molecular typing techniques has enabled the characterization of thousands of HLA alleles [29]. The current number of HLA alleles is shown in Table 4 and can also be found on the IPD-IMGT/HLA Database [30,31].

Analysis of the nucleotide sequences of the HLA genes indicates that most of the polymorphisms are found in exons 2 and 3 of the HLA class I genes and in exon 2 of the HLA class II genes. These exons encode for the distal membrane domains called the "peptide-binding region" [2, 29-31]. It has been observed that most nucleotide polymorphisms within the "peptide-binding regions" involve changes that induce a change in the corresponding amino acid sequence (non-synonymous substitutions) and have a high level of correlation with phenotype differences detected by serological and cellular methods [32]. However, serological equivalents are not available for all described alleles [2, 30, 33], and it is difficult to predict the serological specificities of selected alleles with polymorphisms corresponding to more than one antigenic group [34,35] (Table 2).

It has been observed that most of the polymorphisms are restricted to certain segments of the gene, known as variable regions. Allele pairs associated with the same serotype (e.g., A*02:01, A*02:02) differ only by a few nucleotides. At the same time, distinguishing sequences are observed in alleles of other serotypes, indicating the patchwork nature of HLA polymorphisms. The significant HLA polymorphism probably evolved from the existence of a few allelic lineages, followed by short segmental exchanges, to increase the number of alleles at a given location. It appears that most of the HLA polymorphisms were generated by this mechanism. Then, selected natural selection events must have been necessary for new alleles to reach a significant frequency in the population. Nevertheless, it is worth noting that many alleles arise from single-point mutations.

With an understanding of the significant level of HLA polymorphism and the improvement of molecular techniques, several molecular HLA typing methods have been developed [44-46]. These methods have focused on detecting polymorphisms in exons 2 and 3 of HLA class I genes and in exon 2 of HLA class II genes. The application of these molecular techniques has led to the development of accurate and reproducible HLA typing methods suitable for clinical use [30-32, 36]. The wide application of these methods has led to the identification of many new alleles that were previously undetectable with the serological and cellular methods. The molecular HLA typing methods are widely used and take advantage of the efficiency of DNA amplification by polymerase chain reaction (PCR). The more widely used molecular HLA typing methods currently used in histocompatibility laboratories are: (1) amplification with sequence-specific primers (SSP), (2) hybridization with sequence-specific oligonucleotide probe hybridization (SSOPH), and (3) direct analysis of the DNA sequence (sequence-based typing, SBT) by means of Sanger sequencing or next-generation sequencing (NGS).

Table 2

HLA specificities 1 identified by serological and molecular methods (as of September 2022)

Gene	Serology	Proteins	Alleles	Null alleles
HLA-A	28	4450	7644	397
HLA-B	60	5471	9097	318
HLA-C	10	4218	7609	330
HLA-DRA1	0	5	43	0
HLA-DRB1	21	2203	3389	115
HLA-DRB3	1	334	446	22
HLA-DRB4	1	144	223	25
HLA-DRB5	1	142	187	23
HLA-DQA1	0	244	508	13
HLA-DQB1	9	1455	2330	102
HLA-DPA1	0	233	491	21
HLA-DPB1	6	1325	2221	113

Note:[1] Obtained from the IPD-IMTG/HLA Database [30, 31, 36].

Naming of HLA alleles

G-codes are used to identify ambiguous HLA allele types that share identical nucleotide sequences in the peptide-binding domains. Specifically, they refer to exons 2 and 3 for class I HLA genes and exon 2 for class II genes. These alleles are grouped under a "G" designation, which includes three fields and must contain at least six digits, such as A*01:01:01G [39,40]. Groups may include alleles with unsequenced genomic regions, as well as those obtained through whole-gene sequencing that differ only in silent (synonymous) nucleotide substitutions. When full sequencing is not available, alternative alleles within the same group may be used to infer the missing parts. A detailed list of such alleles, including information about unsequenced regions and substituted sequences, can be found in the ambiguous typing files available in the IMGT/HLA database. If a sequence is revised or removed and only one allele remains in a G group, the G-code designation is kept and can be expanded later if new alleles with identical nucleotide sequences in the peptide-binding domains are identified.

P-codes, on the other hand, are used to report alleles that encode identical antigen-binding domains at the protein level. For class I HLA alleles, this includes proteins encoded by exons 2 and 3, while for class II, it refers to exon 2. The P-code follows the two-field nomenclature of the allele with the lowest number in the group and must include at least four digits, for example, A*01:01P. A full list of alleles grouped by P-code can be found in the downloadable file HLA_nom_p.txt at https://hla.alleles.org [39,40]. Because exon boundaries often intersect codons, meaning one base of a codon may be in one exon and the remaining two in another, such partial codons are excluded from P group comparisons. For example, in the analysis of HLA-A, codons 1 and 183 are excluded since they straddle the boundaries between exons 1/2 and 3/4, respectively.

The allotype of an allele is represented by the numbers preceding the first column, which typically correspond to serological antigen types. The second field identifies subtypes, assigned sequentially as DNA sequences are discovered. In cases where two alleles differ in the first or second field, this implies they differ by at least one nonsynonymous substitution, a change

in the nucleotide sequence that alters the resulting amino acid. The third field distinguishes alleles that differ only by synonymous (silent) mutations within coding regions. In contrast, a fourth field is used when alleles differ only in non-coding regions, such as introns or the 5′/3′ untranslated regions (UTRs) adjacent to exons and introns (Figure 1).

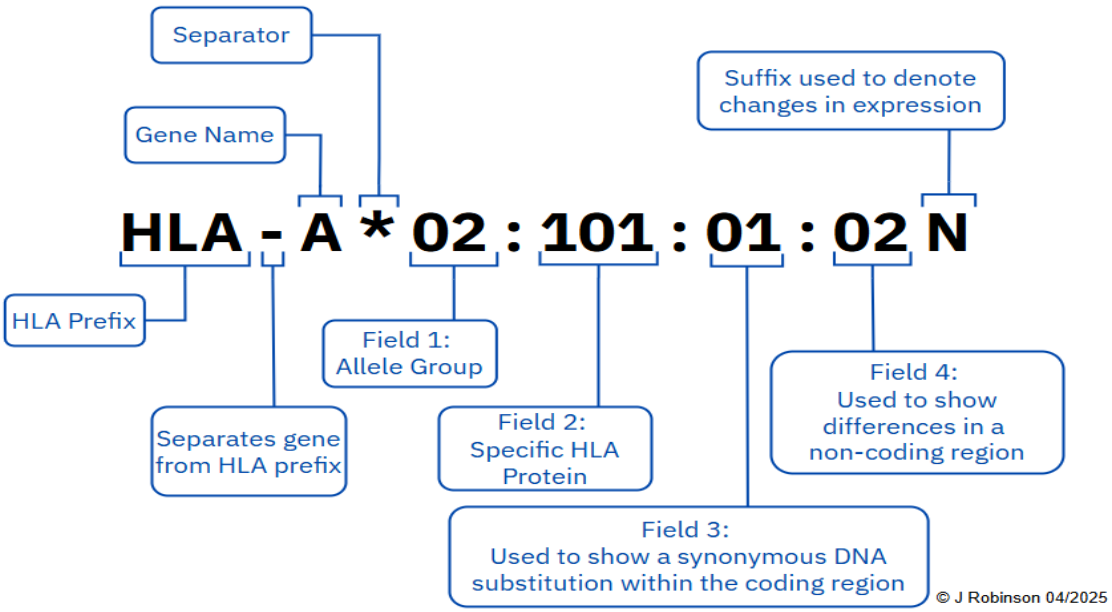

Figure 1. Nomenclature of the HLA alleles

In addition to the unique allele number, there are additional optional suffixes that can be added to the allele to indicate its expression status. Alleles that are not expressed, i.e. 'Null' alleles, are assigned the suffix 'N'. Alleles that are alternatively expressed may have suffixes L, S, C, A, or Q (Table 3).

Table 3

Designation of alleles by additional suffixes *

Suffix	Interpretation
N	An allele that is not expressed on the cell surface.
L	An allele that has been shown to have "low" cell surface expression compared to normal levels.
S	An allele that encodes a protein expressed as a soluble, "secreted" molecule, but not present on the cell surface.
C	Assigned to alleles that produce proteins present in the "cytoplasm," rather than on the cell surface.
A	Indicates "deviant" expression, when it is doubtful that the protein is actually expressed.
Q	Used when the expression of the allele is "questionable," considering that a mutation observed in the allele has been shown to affect normal expression levels in other alleles.

Note: * Taken from www.hla.alleles.org, Anthony Nolan Research Institute, London, UK.

According to the WHO Nomenclature Committee for Histocompatibility and the *IMGT/HLA* database, there are three principal levels of resolution used in *HLA* typing [4]:

• Low-resolution typing: molecular HLA typing results defined by the first field of the allele nomenclature *(e.g., A*31, B*07)*, which often correspond to the serological equivalents of *HLA* antigens [5,36].

• High-resolution typing: typing results that include both the first and second fields of the nomenclature *(A*31:01, B*07:02, etc.)*, and distinguishes alleles encoding identical peptide-binding domains, excluding those that are not expressed as surface proteins. The antigen-binding domain includes domains 1 and 2 of the α polypeptide chain (class I), and domain 1 of the α and β chains (class II).

• Intermediate-resolution typing: based on results of the first field of the nomenclature, irrespective of the specific position between low- and high-resolution distinctions. This includes allele groups such as *DRB1*11:01/11:09/11:28,* which may be represented using *NMDP* (National Marrow Donor Program) codes, such as 11:BYCC [6,33].

The HLA system consists of a highly diverse group of genes and their associated molecules, which are crucial for immune regulation, transplant compatibility, and transfusion success. These antigens are encoded by genes found within the *MHC* on the short arm of chromosome 6. HLA molecules are essential for differentiating self from non-self, triggering immune responses to antigenic stimuli, and coordinating both cell-mediated and humoral immunity (Figure 2).

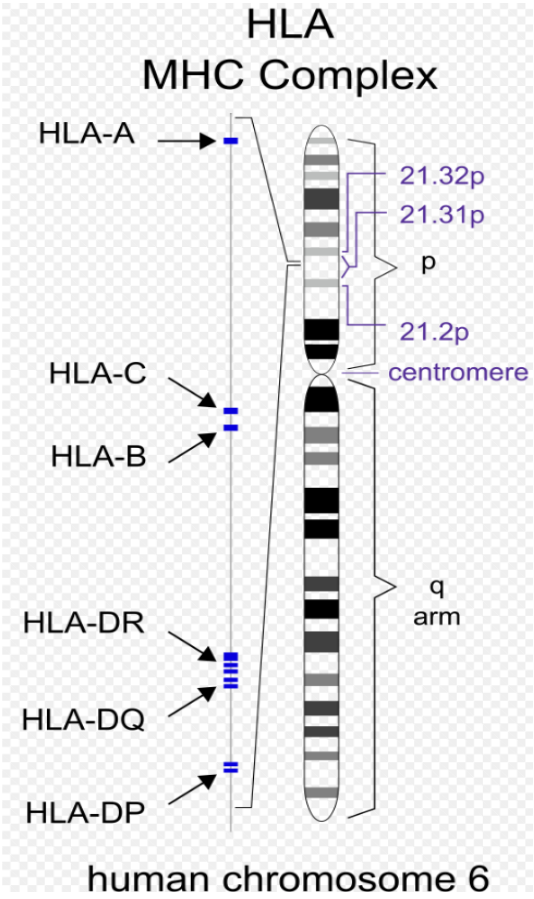

Figure 2. The HLA complex on chromosome 6 comprises the 3-6 kb class I region and the 4-11 kb HLA class II region. The HLA class III is not part of the polymorphic HLA system

The products of HLA genes are crucial in triggering various immune responses by presenting both self and non-self peptide antigens to T lymphocytes. Most polymorphisms in class I HLA genes are found in exons that encode the peptide-binding groove and regions involved in interacting with the T cell receptor. This high level of polymorphism is seen as an evolutionary adaptation, boosting immune defence by increasing the diversity of peptides that can be presented to T cells. Additionally, the codominant expression of HLA genes further enhances this diversity, as both maternal and paternal alleles are expressed, broadening the range of peptides that can be recognized and presented [7,8,35].

Class I MHC molecules are heterodimeric molecules, comprising a single heavy α-chain (~45 kDa) that is non-covalently associated with the invariant 12 kDa β2-microglobulin, and are expressed on the surface of all nucleated cells and platelets (Figure 4). The heavy α-chain is organized into three extracellular domains (α1, α2, α3), a transmembrane segment, and a cytoplasmic tail. Of these, the α1 and α2 domains are or particular importance, as they form an elongated groove through their α-helical structures, which functions as the binding site for processed antigens. This antigen-binding complex is essential for immunogenic recognition by T cells. The α1 and α2 domains are encoded by exons 2 and 3 of the HLA-A, HLA-B, and HLA-C genes, as illustrated in Figure 3 [8,34].

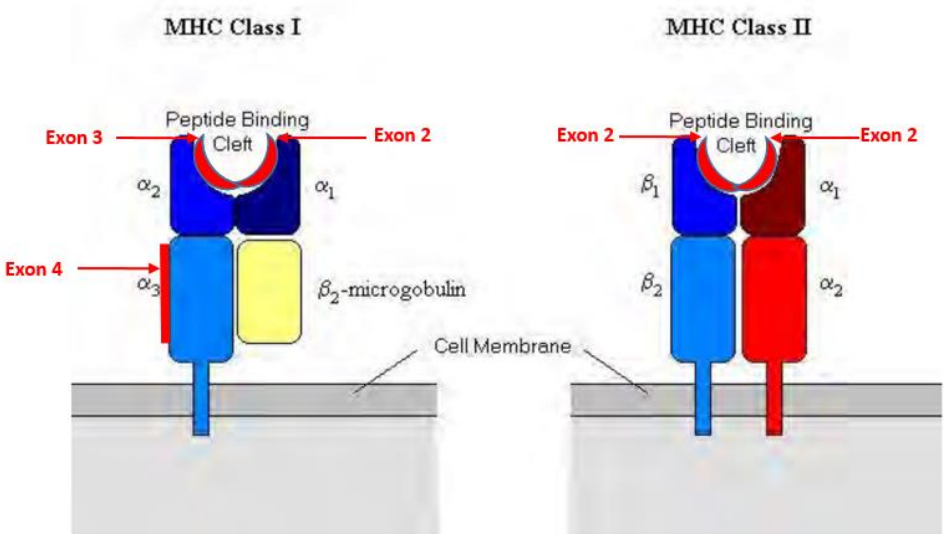

Figure 3. Structure of HLA class I and class II molecules. HLA class I molecules consist of one polymorphic heavy chain (α) associated with a non-polymorphic β2-microglobulin (β2m). The HLA class II molecules consist of two disulphide-linked polymorphic α and β chains

Conversely, class II molecules are mainly expressed on the surface of B lymphocytes, activated T cells, monocytes, macrophages, and dendritic cells, where they are vital in directing the interactions between T lymphocytes and antigen-presenting cells during immune responses. Structurally, class II molecules are heterodimers composed of non-covalently linked α- and β-chains [9,34]. Each chain features an extracellular region folded into two domains, connected to the transmembrane domain by a short linker peptide. Unlike class I MHC molecules, the peptide-binding site of class II MHC consists of domains from both chains, especially the α1 and β1 domains. Therefore, exon 2 of both the α and β genes is central to forming this binding groove.

The structure of classical HLA genes, schematically shown in Figure 3, illustrates the exons responsible for encoding the peptide-binding region of the antigen-binding site (highlighted in red). Other exons encoding the transmembrane domain and cytoplasmic tail do not contribute to antigen presentation and are therefore considered of clinical relevance only when non-expressed (null) alleles are present. Such alleles may result from premature stop codons within exons or from mutations at splice sites that disrupt normal mRNA processing [41,42].

The current number of known HLA alleles can be found on the official HLA nomenclature committee website and in the IPD-IMGT/HLA database [25, 32, 43]. As noted previously, most polymorphisms in class I HLA genes are located in exons 2 and 3, while in class II genes, they are mainly found in exon 2. These regions encode the membrane-distal domains that form the peptide-binding site [30]. Importantly, most nucleotide polymorphisms within these regions lead to nonsynonymous amino acid substitutions, which strongly correlate with phenotypic variation observed through serological and cellular typing methods [30, 31]. However, serological equivalents have not been established for all HLA alleles, and in some cases, allelic polymorphisms may correspond to multiple antigenic groups, complicating the prediction of serological specificity [41].

Methods for obtaining HLA allele sequences

Sanger sequencing, also known as Sequence-Based Typing (SBT), is widely recognized as one of the most precise techniques for HLA typing, as it allows for the direct determination of nucleotide sequences. Until the advent of newer technologies, the most commonly employed method for sequence-based HLA typing was the chain termination technique, developed by Dr. Frederick Sanger in the 1970s [44]. This approach relies on the random incorporation of four dideoxyribonucleotide triphosphates (ddNTPs), ddATP, ddCTP, ddGTP, and ddTTP, each tagged with a unique fluorescent dye. During the second phase of the process, the DNA fragments terminated by ddNTPs are separated by size using gel electrophoresis.

To determine the nucleotide sequence, the gel is read from bottom to top, as the smallest fragments (which migrated the farthest) represent the nucleotides closest to the 5' end of the DNA strand. DNA synthesis by DNA polymerase proceeds in the 5' to 3' direction, starting from a specific primer. Therefore, each fragment ends with a ddNTP that corresponds to a particular position in the original sequence. A computerized detection system was introduced to analyze the gel, with a laser exciting the fluorescent dyes at the ends of the fragments. The final output is a chromatogram, which presents a series of fluorescent peaks, each representing one nucleotide in the 5’ to 3’ direction of the DNA strand (Figure 4).

Sanger sequencing is commonly employed to obtain clinically relevant HLA exon sequences, particularly exons 2 and 3 for class I genes and exon 2 for class II genes, which are critical for transplantation. Depending on the design of primers in commercial reagent kits and the specific sequencing requirements, additional exons may also be sequenced. For example, the “Protrans S4” kit (Protrans medizinische diagnostische Produkte GmbH) enables the sequencing of exons 1 through 4 for class I HLA genes, full exon 2 for HLA-DRB1*, and exons 2 and 3 for HLA-DQB1* of class II.

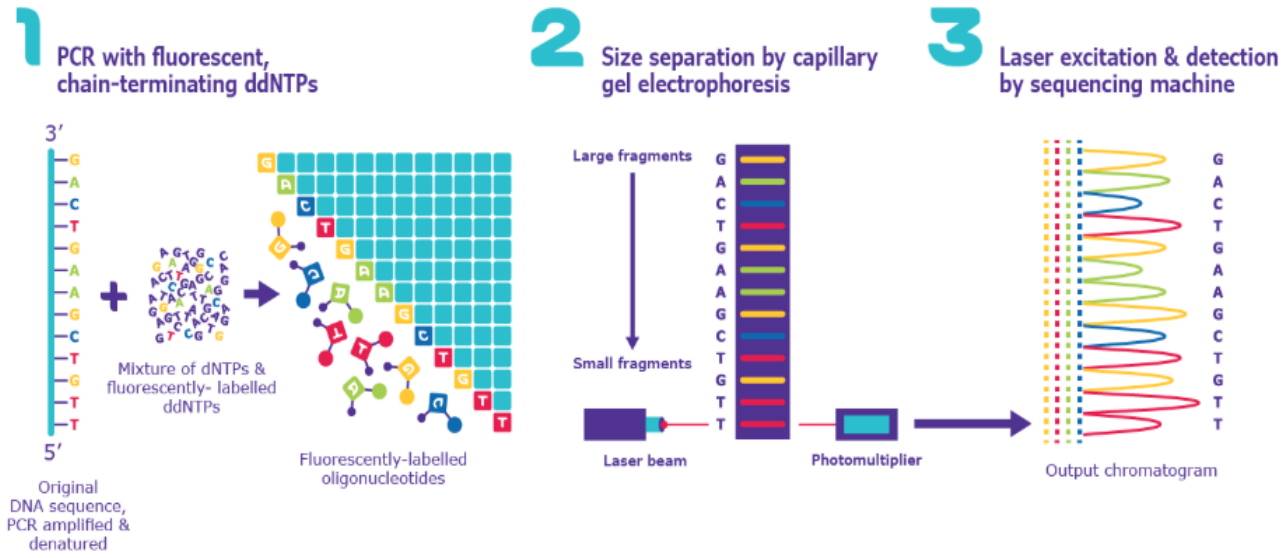

Figure 4. Three basic steps of automated sanger sequencing

Despite its high accuracy, Sanger-based HLA typing has several limitations, including low throughput and high operational costs. A notable drawback is its inability to phase heterozygous nucleotide positions, which often results in ambiguous typing that requires additional, labor-intensive testing to resolve. These limitations are increasingly being addressed by next-generation sequencing (NGS) technologies, specifically short-read NGS (second-generation sequencing). These methods rely on the high-throughput parallel sequencing of clonally amplified short DNA fragments (typically 250–800 base pairs in length). Several commercial NGS-based HLA typing kits are now available, offering comparable accuracy to Sanger sequencing, with the added advantages of efficient multiplexing and the ability to sequence both class I and class II HLA genes relevant to clinical practice. A standard NGS workflow typically involves DNA extraction, library preparation, and sequencing (Figure 5) [45].

Library preparation involves a series of additional steps, such as DNA fragmentation, end-repair, adapter ligation, and size selection. Fragmentation is crucial for producing DNA fragments within the optimal size range for a specific NGS platform. Methods used include sonication, transposase-mediated “tagmentation”, or thermal fragmentation using divalent metal ions. Shorter DNA fragments generally provide higher sequencing accuracy, while longer fragments offer valuable phasing information over greater genomic distances. After fragmentation, DNA ends are repaired to facilitate the ligation of NGS-compatible adapters. These adapters contain both platform recognition sequences and barcodes, enabling sample multiplexing, the simultaneous sequencing of multiple samples in a single run.

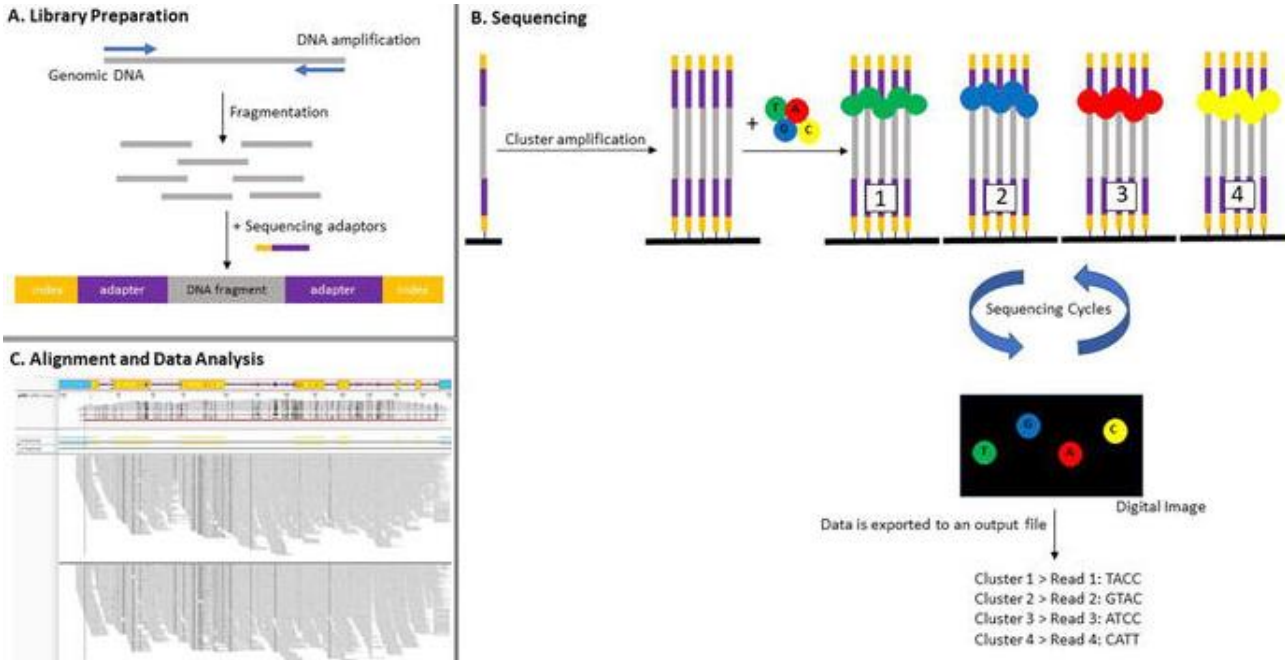

Figure 5. Steps of next-generation sequencing (NGS)-based HLA typing. **(A)** A sequencing library is generated by fragmenting amplified genomic DNA and attaching platform-specific adapters, which include index sequences (barcodes), to both ends of each DNA fragment. **(B)** The prepared library is then loaded into a flow cell, where DNA fragments bind to the surface and are amplified to form clonal clusters. Sequencing begins with the addition of fluorescently labelled nucleotides and other reagents. Each cycle incorporates one nucleotide per cluster, and the emitted fluorescence is recorded. The wavelength and intensity of the light emitted by each cluster identify the incorporated base. This sequencing-by-synthesis cycle is repeated over multiple rounds. **(C)** The resulting sequencing reads are then aligned to reference sequences from the IPD-IMGT/HLA database [4,5], allowing for the identification of nucleotide variations between the reference genome and the newly sequenced DNA

Size selection, performed after adapter ligation, enriches DNA fragments of a desired size and removes residual contaminants, enhancing sequencing efficiency. This can be achieved via bead-based or electrophoretic methods. Bead-based methods enable concurrent DNA concentration, whereas electrophoresis offers higher precision. Alternatively, on-bead tagmentation protocols streamline the process by combining fragmentation, adapter ligation, and normalization into a single step. A PCR amplification step follows, adding platform-specific adapters and sample-identifying barcodes. This streamlined workflow enables library generation in under 90 minutes, with less than 15 minutes of hands-on time, making it both efficient and scalable for clinical and research applications [46].

Given the large volume of data generated by NGS, efficient bioinformatics analysis and data management are essential for its successful application in HLA laboratories. The initial step in sequencing data analysis is carried out by the instrument itself, which performs base-calling for each clonally amplified DNA fragment. During this process, quality control procedures such as read filtering and trimming are also applied. The sequencing data, along with associated quality metrics, is stored in a FASTQ file format. For HLA genotyping, specialized commercial software programs are available for the final analysis phase.

Due to the high polymorphism of HLA genes, aligning sequences to the human reference genome is often inadequate for precisely identifying the specific HLA alleles in a patient's

sample. Instead, alignment is conducted against the IPD-IMGT/HLA Database, which contains sequences of all currently known HLA alleles [30,31]. Coverage is another crucial quality metric at this stage, comprising both depth of coverage (the number of times a base is sequenced) and breadth of coverage (the percentage of the reference genome covered). It is important to note that coverage may not be uniform across the amplicon, and insufficient coverage in critical regions such as exons can affect the accuracy of HLA typing results. Commercial HLA typing software generally includes filters to ensure the minimum coverage required for reliable genotyping. However, in some cases, lower thresholds may be acceptable, such as when polymorphisms between alleles at a locus are phased or when low-coverage regions, like introns and untranslated genomic areas, do not impact typing accuracy. A major concern in HLA data analysis is ensuring balanced allele representation to avoid allele dropout, which can result from preferential amplification due to technical factors or the patient's disease state, where one allele may be lost (loss of heterozygosity).

Recently, third-generation sequencing technologies have been developed, enabling the direct sequencing of DNA fragments over 10 kb from the genome. Although earlier versions of these methods had limitations, recent improvements have significantly enhanced their accuracy. NGS technologies enable the resolution of polymorphisms at various stages, eliminating ambiguities and providing high-resolution HLA typing without the need for re-testing, thereby offering a potential solution to previous challenges in HLA typing. Many HLA typing kits, such as Holotype HLA (Combion Biocomputing Ltd., Hungary), focus on amplifying long PCR fragments of HLA genes and perform sequencing on platforms like Illumina MiSeq. These kits amplify genes such as A, B, C, DQA1, and DQB1 across their entire coding region, including parts of the untranslated 5' and 3' regions. In contrast, DRB1 is amplified from intron 1 to intron 4, and DPB1 is amplified from intron 1 to intron 3 (Figure 6).

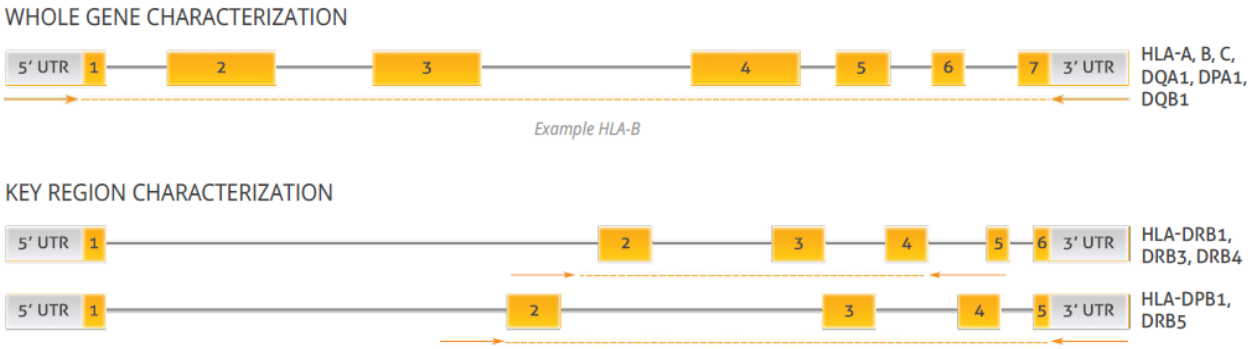

Figure 6. Gene coverage by NGS sequencing method

The WHO Nomenclature Committee for Factors of the HLA System releases an HLA antigen database every three months at the request of the Working Committee on Information Technology of the World Marrow Donor Association (WMDA) for the interpretation of sequencing in various computer programs. These files comprise the IMGT/HLA database, which documents the official HLA nomenclature, the relationships between serologically defined antigens, and the relationships between HLA allele sequences [30, 31, 36]. The database comprises five files, each with a short header containing four lines of information: the file name, the date of file creation (YYYY-MM-DD), the file format (URL), and the author. The files contain information about all

current and retired HLA antigens and alleles, sorted by locus and antigen/allele number. For HLA antigen names assigned before November 1987, dates are assigned approximately.

Conditions for accepting new allelic sequences

To accept DNA sequences as a new gene variant and assign an official name, the following conditions must be met:

• Material used for sequencing: These must be clearly specified.

• Sequencing direction: to be conducted in both directions, using forward and reverse primers.

• PCR amplification requirements: If sequencing is done using PCR-amplified material in the forward direction, the products must come from two separate PCR reactions.

• Heterozygous individuals: If one allele is new, it must be sequenced separately from the other allele. Sequencing both alleles together (as in SBT typing) is insufficient to assign an official name to a new allele.

• Primer-derived sequences: these should not be included in the submitted data.

• Confirmation through other methods: If the new sequence contains a novel mutation or an uncharacterized nucleotide combination, it must be confirmed by DNA typing methods (PCR-SSOP, PCR-SSP). This might require new mutation-specific probes or primers which should be documented.

• Database registration: A registration number must be obtained for the new sequence. Sequences can be submitted to online databases such as:

o EMBL: www.ebi.ac.uk/Submissions/index.html

o GenBank: www.ncbi.nlm.nih.gov/Genbank/submit.html

o DDBJ: www.ddbj.nig.ac.jp/sub-e.html

• Sequence length requirements: though preferred, full-length sequences are not required. Minimum sequence requirements are exons 2 and 3 for class I genes and exon 2 for class II genes.

• Intron or non-coding differences: The entire gene sequence, including coding and non-coding regions, must be sequenced if the new sequence differs only in introns or non-coding regions. In the absence of a full-length sequence of the closest related allele, it may also need to be sequenced and presented before an official name can be assigned.

• Manuscript submission: A manuscript describing the new sequence is recommended to be submitted for publication. Draft copies can be sent to the database via email or fax.

• Tumor-derived sequences: Sequences obtained solely from tumor material are generally not accepted by the nomenclature committee.

• Complete HLA type: A complete HLA type for HLA-A, -B, and -DRB1 genes must be provided for any material where a new allele is identified. The sample must also have a second allele at the locus of interest in a heterozygous individual.

• Repository submission: DNA or other materials, preferably cell lines, should be submitted to a public repository or remain in the originating laboratory. The WHO Nomenclature Committee will maintain documentation of this material.

• Online submission: Sequences should be submitted to the WHO Nomenclature Committee using their online submission tool at www.ebi.ac.uk/imgt/hla/subs/submit.html. Researchers must complete a questionnaire comparing their new sequence to related known alleles [47,48].

New alleles identified in the Kazakh population

From 2011 to 2022, four new HLA alleles at the *HLA-A, -B, -C,* and *-DQB1* loci were identified in individuals of Kazakh nationality in Kazakhstan. These were *DQB1*03:82*, *C*06:256*, *B*13:150,* and *A*32:95 [49-52],* identified using capillary sequencing (SBT method). Initial typing of blood samples was performed using SBT sequencing for the *HLA-A, -B, -C, -DRB1,* and *-DQB1* loci with Protrans S4 technology (Protrans, Hockenheim, Germany). During typing at the PCR product stage, haplotypes were separated, resulting in a heterozygous sequence. The nucleotide sequence was obtained using BigDye v1.1 Terminator Reagent (Applied Biosystems, Foster City, CA) and analyzed on a 3730XL Genetic Analyzer (Applied Biosystems). Sequencing was conducted in both forward and reverse directions for exons 2, 3, and 4 of the HLA-A, B, and C loci; for exon 2 of DRB1; and for exons 2 and 3 of DQB1. The results were analyzed using SeqPilot software (JSI Medical Systems, Germany, version 3.35.0):

1. DQB1*03:82: Identified in a patient with acute myeloid leukemia, this allele differs from DQB1*03:01 at exon 2 at position 223, where adenine (A) replaces guanine (G), resulting in an amino acid substitution from Cysteine to Tyrosine. Inheritance was confirmed through family analysis. The allele is registered in the EMBL and IMGT/HLA databases (HWS10018423) and has been named HLA-DQB1*03:82 [49].
2. C*06:256: Discovered while typing a potential hematopoietic stem cell donor, this allele differs from C*06:02:01:01 by a C to A substitution at exon 3, leading to an amino acid change from Glutamine to Lysine (Q→K). Identification was confirmed by NGS. The new allele is registered under the number HWS10027913 and has been named HLA-C*06:256 [50].
3. B*13:150: Found in a child with acute leukemia, this allele differs from B*13:02:01 by an A to C substitution at exon 2 that causes an amino acid change from Methionine to Leucine (M→L). Family analysis confirmed inheritance of the variant from the mother and brother. It is registered in the database under number HWS10060741 and has been named HLA-B*13:150 [51].
4. A*32:95: Identified in a potential stem cell donor, this allele differs from A*32:01:01 by a C to A substitution at exon 2 at position 28 that leads to an amino acid change from Lysine to Threonine (K→T). The allele is registered in the database under number HWS10027001 and has been named HLA-A*32:95 [52].

Following the implementation of the NGS sequencing technology, our laboratory identified an additional ten new sequences with changes in clinically significant exons (exons 2 and 3), as well as four sequences in clinically insignificant exons, such as exons 1, 4, 5, and 8 [53, 54]. Despite the short time of using NGS technology, the identification of new sequences has increased. The frequency of detecting new allelic variants using capillary sequencing is one new allele per 1,773 typings; whereas with NGS technology, detection of new alleles occurs every 635 typings.

The 2.8-fold increase in detection efficiency after implementing NGS technology results from NGS's superior ability to phase heterozygous positions across larger genomic areas and to sequence intronic regions where additional polymorphisms may be found. Additionally, NGS workflows enable simultaneous typing of multiple loci with high coverage depth, reducing ambiguities that often affect Sanger-based approaches. All four identified alleles feature non-synonymous substitutions within exons encoding peptide-binding domains (exons 2 and 3 for class I; exon 2 for class II). These amino acid changes: Cys→Tyr (DQB1*03:82), Gln→Lys (C*06:256), Met→Leu (B*13:150), and Lys→Thr (A*32:95), occur at positions that may influence peptide repertoire and T cell receptor recognition. Nevertheless, functional validation

through peptide-binding assays and T cell activation studies remains essential to determine the immunological impact of these substitutions.

Discussion

During this study, previously unreported HLA class I and II allelic varFor example, the "Protrans S4" kit (Protrans medizinische diagnostische Produkte GmbH) enables the sequencing of exons 1 through 4 for class I HLA genes, full exon 2 for HLA-DRB1*, and exons 2 and 3 for HLA-DQB1* ofnary pressures related to pathogen antigen recognition [55]. According to the IMGT/HLA Database (version 3.57.0, January 2025), there are currently over 37,000 different HLA gene alleles registered, including 8,123 alleles for HLA-A, 9,276 for HLA-B, and 6,475 for HLA-DRB1, with this number continually increasing [2].

In our analysis, new variants were identified, including B*51:01:XX and DRB1*13:XX, which contain single-nucleotide substitutions in coding exons that could potentially alter the peptide-binding domain structure. Such mutations can significantly affect antigen presentation and are associated with an increased risk of developing autoimmune and infectious diseases [56].

While African and South Asian populations have historically been considered the primary reservoirs of HLA diversity due to longer evolutionary timescales and larger effective population sizes [57], our findings indicate that substantial unique ancient migration waves and prolonged isolation of certain sub-ethnic groups persist in Central Asian populations. The Kazakh population occupies a geographic crossroads that has historically been traversed by migration waves connecting Europe, East Asia, and Central Asia. This positioning likely facilitated admixture between diverse ancestral groups, potentially generating novel allelic combinations through recombination events. Furthermore, subsequent isolation of certain sub-ethnic groups within Kazakhstan's vast territory may have preserved rare variants through genetic drift. The four novel alleles identified in this study show no exact matches in the Allele Frequency Net Database (AFND), which aggregates HLA data from over 1,500 worldwide [58].

This absence could reflect either recent mutational origin, preservation of ancient variants in isolated populations, or simply the historical undersampling of Central Asian groups in global HLA studies. Distinguishing among these scenarios requires expanded sampling with dense geographic coverage and phylogenetic analysis of allelic relationships. Comparing the obtained alleles with global panels revealed that several commonly occurring alleles in our study sample are absent from most standard diagnostic panels used in bone marrow donor typing. This may complicate donor matching for Kazakh recipients. According to the World Marrow Donor Association (WMDA), as of 2024, there are over 41 million potential donors registered in international registries; however, a significant portion is represented by European and North American populations. This creates a disparity and reduces the effectiveness of finding HLA-compatible donors for recipients from Central Asia [59].

From a practical standpoint, identifying new alleles highlights the need to adapt HLA typing panels to local ethnic characteristics. This is particularly important in the context of developing biomedical products, such as personalized vaccines, immunotherapeutic agents, and transplant matching algorithms. Moreover, accounting for rare and unique alleles can enhance the accuracy of population genetic and epidemiological models while minimizing the risks of transplant rejection and immunological complications [60]. Thus, our findings expand existing knowledge about HLA gene diversity and underscore the importance of deep local analysis within global biobanks and registries.

Several limitations are noteworthy. The sample sizes for individual novel alleles remain small (n = 1-4 carriers), which hampers the accurate estimation of allele frequencies and restricts understanding of linkage disequilibrium patterns within the Kazakh population [61-63]. The functional attributes of the identified amino acid substitutions were not experimentally validated, for example, through peptide-binding assays, thermal stability measurements, or T cell activation studies. Additionally, our sequencing strategy focused on clinically relevant exons (exons 2-3 for class I; exon 2 for class II), potentially missing variants in promoter regions or intronic enhancers that could influence expression levels. Furthermore, the comparison of detection rates between Sanger and NGS technologies may be confounded by temporal differences in sample composition, typing indications (e.g., patient versus donor typing), and operator experience [64]. Future plans include increasing the sample size and conducting functional tests to assess the impact of identified mutations on antigen binding, which could help advance personalized medicine in the region.

Conclusions

Our study identified the novel HLA alleles that contain non-synonymous substitutions within peptide-binding domains: *HLA-DQB1*03:82, HLA-C*06:256, HLA-B*13:150*, and *HLA-A*32:95* in the Kazakh population. This addition to the IPD-IMGT/HLA database enhances our understanding of HLA diversity in Central Asia, a region that has historically been underrepresented in global immunogenetic studies. The transition from Sanger sequencing to NGS-based typing in our laboratory increased the efficiency of novel allele detection by 2.8-fold, further demonstrating the superiority of NGS for HLA characterization, which stems from its ability to resolve phase ambiguities and sequence extended genomic regions with high throughput. Several aspects must be considered when identifying a new allele based on DNA sequence. It is essential to have knowledge of the structure of a similar allele, and the location (exon, intron) of the nucleotide substitution and its relation to the codon. It is necessary to describe the protein sequence and determine whether this substitution will lead to an amino acid change. These and related factors are needed when submitting a new gene to the nomenclature committee and for its registration.

Kazakhstan's multi-ethnic composition (comprising over 124 ethnic groups) and its location along migration routes support ongoing efforts to identify additional unique HLA variants. These population-specific data are crucial for improving transplant donor matching for Kazakh recipients, who are currently underserved by registries dominated by European and North American data, for developing personalized immunotherapeutic strategies that consider local HLA diversity, and for studying disease associations in Central Asian populations. Future research will focus on expanding sample sizes, validating the immunological effects of identified substitutions through peptide-binding and T-cell assays, and establishing collaborations to enhance Central Asian representation in international HLA registries and biobanks. These initiatives will significantly enhance our understanding of HLA genetic diversity in Central Asia and promote the clinical use of HLA typing in personalized medicine.

Author Contributions

A.T. – conceptualization and writing the draft of the manuscript; **S.A. –** discussion of main findings and project administration; **W.Y.A. –** supervision and editing of the article.

Funding

This study did not receive funding from government or non-government sources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Compliance with ethical standards

Not applicable, as it is a review article.

References

1. Nomenclature for T-cell receptor (TCR) gene segments of the immune system. WHO-IUIS Nomenclature Sub-Committee on TCR Designation.. Bulletin of the World Health Organization 2016;71(1):113-115. World Health Organization. https://iris.who.int/handle/10665/261601

2. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020;48(D1):D948-D955. https://doi.org/10.1093/nar/gkz950

3. Robinson J, Barker DJ, Marsh SGE 25 years of the IPD-IMGT/HLA Database. HLA. 2024;103(6):e15549. https://doi.org/10.1111/tan.15549

4. Robinson J, Soormally AR, Hayhurst JD, Marsh SGE. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. Hum Immunol. 2016;77:233-237. https://doi.org/10.1016/j.humimm.2016.01.020

5. Nomenclature for Factors of the HLA System. Nomenclature of HLA alleles [Electronic resource]. Available at: https://hla.alleles.org/nomenclature/naming.html

6. Kahn J, Kahn S. Molecular diagnostics: A review of the current state and future directions. Clin Chem Lab Med. 2018;56(4):569-578. https://doi.org/10.1515/cclm-2017-0662

7. Anzar I, Sverchkova A, Samarakoon P, et al. Personalized HLA typing leads to the discovery of novel HLA alleles and tumor-specific HLA variants. HLA. 2022;99(4):313-327. https://doi.org/10.1111/tan.14562

8. NMDP Allele Code List in Numerical Order [Electronic resource]. Available at: https://network.nmdp.org/about-us

9. Bover KH. Homoisotransplantation von Epidermis bei eineiigen Zwillingen. Beiträge zur Klinischen Chirurgie. 1927;141:442-447

10. Gorer PA. The genetic and antigenic basis for tumor transplantation. J Pathol and Bacteriol 1937;44:691-697. https://doi.org/10.1002/path.1700440313

11. Snell GD. Methods for the study of histocompatibility genes. J Genet. 1948;49(2):87-108. https://doi.org/10.1007/BF02986826

12. Payne R. The development and persistence of leukoagglutinins in parous women. Blood. 1962;19:411-424. https://doi.org/10.1182/blood.V19.4.411.411

13. Terasaki PI, Mandell M, Vandewater J, Edgington TS. Human blood lymphocyte cytotoxicity reactions with allogenic antisera. Ann NY Acad Sci. 1964;120:322-334. https://doi.org/10.1111/j.1749-6632.1964.tb34731.x

14. Terasaki PI, McClelland JD. Microdroplet assay of human serum cytotoxins. Nature. 1964;204:998-1000. https://doi.org/10.1038/204998b0

15. Terasaki PI, Mickey MR. Histocompatibility-transplant correlation, reproducibility, and new matching methods. Transplantation Proceedings. 1971;3(2):157-171.

16. Terasaki PI, Rich NE. Quantitative determination of antibody and complement directed against lymphocytes. J Immunol. 1964;92:128-138. https://doi.org/10.4049/jimmunol.92.1.128

17. Bach FH, Amos DB. Hu-1: Major histocompatibility locus in man. Science. 1967;156(3781):1506-1508. https://doi.org/10.1126/science.156.3781.1506

18. Bach FH, Kisken WA. Predictive value of results of mixed leukocyte cultures for skin allograft survival in man. Transplantation. 1967;5(4, Suppl):1046-1052.

19. Williams RC Jr, Emmons JD, Yunis EJ. Studies of human sera with cytotoxic activity. J Clin Invest. 1971;50(7):1514-1524. https://doi.org/10.1097/00007890-196707001-00039

20. Yunis EJ, Amos DB. Three closely linked genetic systems relevant to transplantation. Proc Nat Acad Sci USA. 1971;68(12):3031-3035. https://doi.org/10.1073/pnas.68.12.3031

21. Yunis EJ, Plate JM, Ward FE, Seigler HF, Amos DB. Anomalous MLR responsiveness among siblings. Transplant Proc. 1971;3(1):118-120.

22. Thorsby E, Sandberg L, Lindholm A, Kissmeyer-Nielsen F. The HLA system: Evidence of a third sub-locus. Scand J Haematol. 1970;7(3):195-200. https://doi.org/10.1111/j.1600-0609.1970.tb01887.x

23. Shaw S, Duquesnoy RJ, Smith PL. Population studies of the HLA-linked SB antigens. Immunogenet. 1981;14(1–2):153-162. https://doi.org/10.1007/BF00344308

24. Shaw S, Johnson AH, Shearer GM. Evidence for a new segregant series of B cell antigens that are encoded in the HLA-D region and that stimulate secondary allogenic proliferative and cytotoxic responses. J Exp Med 1980;152(3):565-580. https://doi.org/10.1084/jem.152.3.565

25. IPD-IMGT/HLA Database. European Bioinformatics Institute [Electronic resource]. Available at: https://www.ebi.ac.uk/ipd/imgt/hla

26. Noreen HJ, Yu N, Setterholm M, et al. Validation of DNA-based HLA-A and HLA-B testing of volunteers for a bone marrow registry through parallel testing with serology. Tissue Antigens. 2001;57(3):221-229. https://doi.org/10.1034/j.1399-0039.2001.057003221.x

27. Klein J, Sato A. The HLA system. First of two parts. N Engl J Med. 2000;343(10):702-709. https://doi.org/10.1056/NEJM200009073431006

28. Klein J, Sato A. The HLA system. Second of two parts. N Engl J Med. 2000;343(11):782-786. https://doi.org/10.1056/NEJM200009143431106

29. Leen G, Stein JE, Robinson J, Maldonado Torres H, Marsh SGE. The HLA diversity of the Anthony Nolan register. HLA. 2021;97(1):15-29. https://doi.org/10.1111/tan.14127

30. HLA Nomenclature Database. Available from: https://hla.alleles.org/

31. IPD-IMGT/HLA Database. Available from: https://www.ebi.ac.uk/ipd/imgt/hla/

32. Nunes E, Heslop H, Fernandez-Vina M, et al. Definitions of histocompatibility typing terms. Blood. 2011;118(23):e180-3. https://doi.org/10.1182/blood-2011-05-353490

33. Noreen HJ, Yu N, Setterholm M, et al. Validation of DNA-based HLA-A and HLA-B testing of volunteers for a bone marrow registry through parallel testing with serology. Tissue Antigens. 2001;57:221-229. https://doi.org/10.1034/j.1399-0039.2001.057003221.x

34. Müller CA, Engler-Blum G, Gekeler V, Steiert I, Weiss E, Schmidt H. Genetic and serological heterogeneity of the supertypic HLA-B locus specificities Bw4 and Bw6. Immunogenetics. 1989;30(3):200-7. https://doi.org/10.1007/BF02421207

35. Voorter CE, van der Vlies S, Kik M, van den Berg-Loonen EM. Unexpected Bw4 and Bw6 reactivity patterns in new alleles. Tissue Antigens. 2000;56(4):363-370. https://doi.org/10.1034/j.1399-0039.2000.560409.x

36. Barker DJ, Maccari G, Georgiou X, et al. The IPD-IMGT/HLA Database. Nucleic Acids Res. 2023;51(D1):D1053-D1060. https://doi.org/10.1093/nar/gkac1011

37. Smith AG, Pereira S, Jaramillo A, et al. Comparison of sequence-specific oligonucleotide probe vs next generation sequencing for HLA-A, B, C, DRB1, DRB3/B4/B5, DQA1, DQB1, DPA1, and DPB1 typing: Toward single-pass high-resolution HLA typing in support of solid organ and hematopoietic cell transplant programs. HLA. 2019;94(3):296-306. https://doi.org/10.1111/tan.13619

38.Bravo-Egana V, Sanders H, Chitnis N. New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing. Hum Immunol. 2021;82(7):478-487. https://doi.org/10.1016/j.humimm.2021.01.010

39. Allen ES, Yang B, Garrett J, Ball ED, Maiers M, Morris GP. Improved accuracy of clinical HLA genotyping by next-generation DNA sequencing affects unrelated donor search results for hematopoietic stem cell transplantation. Hum Immunol. 2018;79(12):848-854. https://doi.org/10.1016/j.humimm.2018.10.008

40. Lind C, Ferriola D, Mackiewicz K, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum Immunol. 2010;71(10):1033-1042. https://doi.org/10.1016/j.humimm.2010.06.016

41. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. Science. 1996;272(5258):67-74. https://doi.org/10.1126/science.272.5258.67

42. Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 1988;335(6186):167-170. https://doi.org/10.1038/335167a0

43. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43(Database issue):D423-D431. https://doi.org/10.1093/nar/gku1161

44. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-5467. https://doi.org/10.1073/pnas.74.12.5463

45. Furuno M, Kasukawa T, Saito R, et al. CDS annotation in full-length cDNA sequence. Genome Res. 2003;13(6B):1478-1487. https://doi.org/10.1101/gr.1060303

46. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021;82(11):801-811. https://doi.org/10.1016/j.humimm.2021.02.012

47. Sanchez-Mazas A, Nunes JM. PGAE HLA Consortium of the 18th International HLA and Immunogenetics Workshop. The most frequent HLA alleles around the world: A fundamental synopsis. Best Pract Res Clin Haematol. 2024;37(2):101559. https://doi.org/10.1016/j.beha.2024.101559

48. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010;75(4):291-455. https://doi.org/10.1111/j.1399-0039.2010.01466.x

49. Turganbekova A, Burkitbaev Z, Abdrakhmanova S, et al. Detection of a novel allele, HLA DQB1*03:82 in a patient of Kazakh nationality with acute myeloid leukaemia. Tissue Antigens. 2014;84:63.

50. Turganbekova A, Baimukasheva D, Zhanzakova Z, Saduakas Z, Parkhomenko I, Abdrakhmanova S. Characterization of the novel HLA-C*06:256 allele, identified in the Republic of Kazakhstan. HLA. 2020. https://doi.org/10.1111/tan.13844

51. Turganbekova A, Zhanzakova Z, Parkhomenko I, et al. The characteristics of the new HLA-B*13:150 allele, which was Identified in the patient of Kazakh nationality with acute leukemia. HLA. 2018;322(4):322.

52. Turganbekova A, Ramilyeva I, Baimukasheva D, Burkitbayev Z, Abdrakhmanova S. The characteristics of the new HLA-A*32:95 allele, which was identified in the donor of the National Register of stem cells of the Republic of Kazakhstan. HLA. 2017, 2017;90(2):112-113. https://doi.org/10.1111/tan.13028

53. Clark PM, Duke JL, Ferriola D, et al. Generation of full-length class I human leukocyte antigen gene consensus sequences for novel allele characterization. Clin Chem. 2016;62(12):1630-1638. https://doi.org/10.1373/clinchem.2016.260661

54. Brown NK, Kheradmand T, Wang J, Marino SR. Identification and characterization of novel HLA alleles: Utility of next-generation sequencing methods. Hum Immunol. 2016;77(4):313-316. https://doi.org/10.1016/j.humimm.2016.01.001

55. Danchin EG, Pontarotti P. Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet. 2004;20(12):587-591. https://doi.org/10.1016/j.tig.2004.09.009

56. Apps R, Qi Y, Carlson JM, et al. Influence of HLA-C expression level on HIV control. Science. 2013; 340(6128):87-91. https://doi.org/10.1126/science.1232685

57. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernández-Viña MA. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. Hum Immunol. 2001;62(9):1009-1030. https://doi.org/10.1016/s0198-8859(01)00298-1

58. Gonzalez-Galarza FF, McCabe A, Santos EJMD, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 2020;48(D1):D783-D788. https://doi.org/10.1093/nar/gkz1029

59. Petersdorf EW, Anasetti C, Martin PJ, et al. Limits of HLA mismatching in unrelated hematopoietic cell transplantation. Blood. 2004;104(9):2976-2980. https://doi.org/10.1182/blood-2004-04-1674

60. Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens. 2012;80(1):1-11. https://doi.org/10.1111/j.1399-0039.2012.01881.x

61. Turganbekova A, Abdrakhmanova S, Masalimov Z, Almawi WY. Genetic diversity and ethnic tapestry of Kazakhstan as inferred from HLA polymorphism and population dynamics: A comprehensive review. Genes (Basel). 2025;16(3):342. https://doi.org/10.3390/genes16030342

62. Hajjej A, Abdrakhmanova S, Turganbekova A, Almawi WY. Diversity of HLA-A, -B, -C, -DRB1, and -DQB1 alleles and haplotypes in Kazakhstani Tatar population and genetic relatedness to other populations. Gene. 2024;896:148062. https://doi.org/10.1016/j.gene.2023.148062

63. Hajjej A, Abdrakhmanova S, Turganbekova A, Almawi WY. Distribution of HLA Class I and Class II alleles and haplotypes in German and Uzbek minorities in Kazakhstan, and relationship to other populations. HLA. 2020;96(5):615-620. https://doi.org/10.1111/tan.14057

64. Messaoudi SA, Al Sharhan NA, Alharthi B, et al. Detection of genetic mutations in patients with breast cancer from Saudi Arabia using Ion AmpliSeq™ Cancer Hotspot Panel v.2.0. Biomed Rep. 2022;16(4):26. https://doi.org/10.3892/br.2022.1509

Номенклатура HLA системы и открытие новых аллельных вариантов в казахской популяции

А. Турганбекова[1,2], С. Абдрахманова[2], В.Й. Алмави*[3,4]

[1]Евразийский национальный университет им. Л.Н. Гумилёва, Астана, Казахстан
[2]Научно-производственный центр трансфузиологии, Астана, Казахстан
[3]Университет Брок, Сент-Катаринс, Онтарио, Канада
[4]Университет Эль-Манар, Тунис, Тунис

Аннотация. Система человеческих лейкоцитарных антигенов (HLA) является одной из наиболее генетически полимерфных в организме человека, которая включает более 220 генов, кодирующих иммунные белки, играющие ключевую роль в трансплантационной совместимости, регуляции иммунного ответа и предрасположенности к заболеваниям. В данном обзоре представлены основы номенклатуры HLA, стандартизированной Всемирной организацией здравоохранения (ВОЗ) и поддерживаемой в базе данных IPD-IMGT/HLA. Описана эволюция методов типирования HLA - от серологических тестов до молекулярных технологий, включая секвенирование по Сэнгеру и секвенирование нового поколения (NGS), а также использование специализированного программного обеспечения для интерпретации данных. Проведена оценка их преимуществ и ограничений при выявлении новых аллелей. В работе рассмотрены аллельные варианты генов HLA, методы их секвенирования и анализа. Кроме того, сообщается

об идентификации четырёх новых аллелей HLA в казахской популяции: *DQB1*03:82, C*06:256, B*13:150 и A*32:95.* Все четыре аллеля содержат несинонимичные замены в пептидсвязывающих доменах, что указывает на их возможную иммунологическую значимость. Сравнительный анализ показал, что технология NGS повышает эффективность обнаружения новых аллелей в 2,8 раза по сравнению с секвенированием по Сэнгеру (один новый аллель на 635 типирований против 1773). Полученные результаты демонстрируют значительное генетическое разнообразие HLA среди населения Центральной Азии, которое остается недостаточно представленным в международных базах данных. Выявление специфичных для популяции аллелей подчеркивает необходимость расширения HLA-профилирования в этнически разнообразных регионах для улучшения исходов трансплантации и развития персонализированной иммунотерапии.
Ключевые слова: аллельные варианты, HLA-типирование, база данных IPD-IMGT/HLA, Казахстан, секвенирование нового поколения (NGS)

HLA жүйесі номенклатурасы және қазақ популяциясында жаңа аллельдік түрлерінің анықталуы

А. Турганбекова[1,2], С. Абдрахманова[2], В.Й. Алмави*[3,4]
[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
[2]Трансфузиологияның ғылыми-өндірістік орталығы, Астана, Қазақстан
[3]Брок университеті, биологиялық ғылымдар кафедрасы, Сент-Катаринс, Онтарио, Канада
[4]Эль-Манар университеті, жаратылыстану факультеті, Тунис, Тунис

Аңдатпа. Адам лейкоцитарлық антигендер жүйесі (HLA) - адам ағзасындағы ең генетикалық тұрғыда әртүрлі жүйелердің бірі. Ол 220-дан астам генді қамтиды және трансплантациялық сәйкестікке, иммундық жауаптың реттелуіне және әртүрлі ауруларға бейімділікке жауапты иммундық ақуыздарды кодтайды. Бұл шолуда Дүниежүзілік денсаулық сақтау ұйымы (ДДҰ) бекіткен және IPD-IMGT/HLA деректер базасында сақталатын HLA номенклатурасының негіздері баяндалады. Сондай-ақ, HLA типтеу әдістерінің дамуы — серологиялық талдаулардан бастап Сэнгер секвенирлеуі мен жаңа буын секвенирлеуіне (NGS) дейінгі технологиялық жетілдірулер қарастырылған. Деректерді интерпретациялауға арналған бағдарламалық құралдардың мүмкіндіктері мен олардың артықшылықтары мен шектеулері талданады. Зерттеу барысында HLA гендерінің аллельдік вариантты, олардың секвенирлеу әдістері мен талдау тәсілдері сипатталады. Сонымен қатар, қазақ популяциясында төрт жаңа HLA аллелі анықталды: *DQB1*03:82, C*06:256, B*13:150* және *A*32:95.* Бұл аллельдердің барлығында пептид байланыстыру домендерінде мағынасын өзгертетін ауысулар барын және бұл олардың иммунологиялық маңыздылығын көрсетуі мүмкін. Салыстырмалы талдау нәтижелері көрсеткендей, NGS технологиясын қолдану жаңа аллельдерді анықтау тиімділігін 2,8 есеге арттырады (әрбір 635 типтеуге бір жаңа аллель, ал Сэнгер әдісінде – 1773 типтеуге бір жаңа аллель). Бұл нәтижелер Орталық Азия халықтарындағы HLA жүйесінің жоғары генетикалық әртүрлілігін айғақтайды. Аталған аймақтар халықаралық деректер базасында жеткіліксіз қамтылғандықтан, этникалық тұрғыдан алуан түрлі популяцияларда HLA-профильдеуді кеңейту қажеттілігі туындайды. Бұл өз кезегінде трансплантация нәтижелерін жақсартуға және жекелендірілген иммунотерапияны дамытуға ықпал етеді.
Түйін сөздер: Аллельдік түрлер; HLA типтеу; IPD-IMGT/HLA деректер базасы; Қазақстан; жаңа буынды секвенирлеу (NGS)

Сведения об авторах:

Турганбекова Аида Аскаркызы – руководитель лаборатории иммунологического типирования тканей, РГП на ПХВ «Научно-производственный центр трансфузиологии», ул. Керей, Жанибек хандар, 10, 020000, Астана, Казахстан.

Абдрахманова Сания Алишевна – кандидат медицинских наук, Председатель Правления РГП на ПХВ "Научно-производственный центр трансфузиологии", ул. Керей, Жанибек хандар, 10, 020000, Астана, Казахстан.

Алмави Васим Йосиф – автор-корреспондент, доктор медицинских наук, профессор кафедры биологических наук, Университет Брок, 1812 Сэр Айзек Брок Уэй, Сент-Катаринс, Онтарио L2S 3A1, Канада.

Авторлар туралы мәліметтер:

Турганбекова Аида Аскарқызы – ШЖҚ МҚК «Трансфузиология ғылыми-өндірістік орталығы» иммунологиялық тіндерді типтеу зертханасының меңгерушісі, Керей, Жәнібек хандар көшесі 10, 020000, Астана, Қазақстан.

Абдрахманова Сания Алишевна – медицина ғылымдарының кандидаты, ШЖҚ МҚК «Трансфузиология ғылыми-өндірістік орталығы» басқарма төрағасы РГП на ПХВ "Научно-производственный центр трансфузиологии", Керей, Жәнібек хандар көшесі 10, 020000, Астана, Қазақстан.

Алмави Васим Йосиф – хат-хабар авторы, медицина ғылымдарының докторы, биология ғылымдары кафедрасының профессоры, Брок Университеті, 1812 Сэр Айзек Брок Уэй, Сент-Катаринс, Онтарио L2S 3A1, Канада.

Authors' information:

Turganbekova Aida – head of the laboratory of immunological tissue typing, State Enterprise on the Right of Economic Management “Scientific and Production Center for Transfusiology”, Kerey and Zhanibek Khandar 10 St., 020000, Astana, Kazakhstan.

Sania Abdrakhmanova – Candidate of Medical Sciences, Chairman of the Board of the State Enterprise on the Right of Economic Management “Scientific and Production Center for Transfusiology”, Kerey and Zhanibek Khandar 10 St., 020000, Astana, Kazakhstan.

Wassim Y. Almawi – corresponding author, Doctor of Medical Sciences, Professor of the Department of Biological Sciences, Niagara Region, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1 Canada.