Human Immunology 82 (2021) 466–477

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Review

The changing landscape of HLA typing: Understanding how and when HLA typing data can be used with confidence from bench to bedside

Lee Ann Baxter-Lowe [a,b,*]

[a] Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, USA
[b] Department of Pathology, University of Southern California, USA

A R T I C L E   I N F O

Article history:
Received 14 March 2021
Revised 26 April 2021
Accepted 29 April 2021
Available online 21 May 2021

A B S T R A C T

Human leukocyte antigen (HLA) genes are extraordinary for their extreme diversity and widespread impact on human health and disease. More than 30,000 HLA alleles have been officially named and more alleles continue to be discovered at a rapid pace. HLA typing systems which have been developed to detect HLA diversity have advanced rapidly and are revolutionizing our understanding of HLA's clinical importance. However, continuous improvements in knowledge and technology have created challenges for clinicians and scientists. This review explains how differences in HLA typing systems can impact the HLA types that are assigned. The consequences of differences in laboratory testing methods and reference databases are described. The challenges of using HLA types that are not equivalent are illustrated. A fundamental understanding of the continual expansion of our understanding of HLA diversity and limitations in some of the typing data is essential for using typing data appropriately in clinical and research settings.

© 2021 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

Contents

1. Introduction: The Bird's eye view . . . 466
2. The landscape is diverse . . . 467
3. The HLA phenotyping path . . . 469
4. The HLA genotyping path . . . 470
5. Challenges created by changing the landscape . . . 473
6. The path of inferred and imputed HLA genotypes . . . 475
7. Summary . . . 475
Declaration of Competing Interest . . . 476
Appendix A. Supplementary data . . . 476
References . . . 476

1. Introduction: The Bird's eye view

The human leukocyte antigen (HLA) system has far reaching clinical and scientific impact including effects in transplantation [1–3], transfusion [4], autoimmune disease [5,6], infectious disease [5], pregnancy [7], adverse drug reactions [8,9], cancer [10], and immunotherapy [10,11]. HLA's importance is reflected in more than 180,000 scientific publications which have contributed to our current understanding of HLA's role in health and disease (Supp Fig. 1). The best studied HLA functions involve the pivotal role of membrane-bound HLA proteins in controlling the specificity of immune responses by serving as ligands for T cell receptors (TCR) and killer-immunoglobulin-like receptors (KIR) [12–14]. The part of the HLA protein that contacts these receptors is referred to as the antigen recognition domain (ARD) or peptide binding domain.

A hallmark of the HLA system is the extraordinary diversity created by the presence of several homologous genes that are hyper-

* Address: Department of Pathology and Laboratory Medicine Children's Hospital Los Angeles, USA.
E-mail address: lbaxterlowe@chla.usc.edu

https://doi.org/10.1016/j.humimm.2021.04.011
0198-8859/© 2021 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

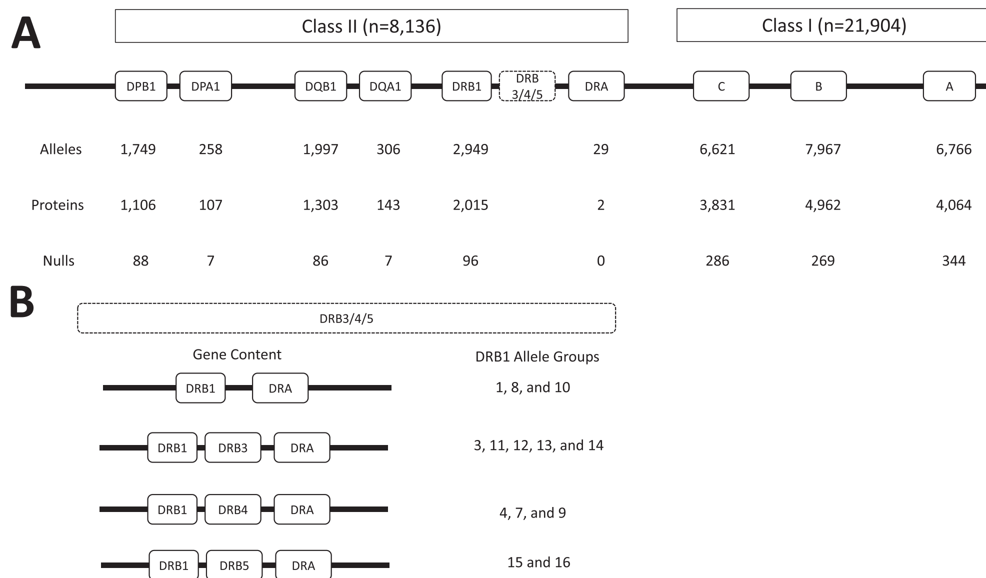

Fig. 1. HLA Alleles and Proteins. A. For HLA-A, -B, -C, DRB1, DQA1, DQB1, DPA1, and DPB1 the number of alleles, proteins and null alleles are shown. B. Some chromosomes have HLA-DRB3, -DRB4, and DRB5 genes. Each allele group (listed on the right) has a particular gene content, but rare exceptions have been reported.

polymorphic [15,16]. The classical Class I HLA genes, designated HLA-A, -B, and C, encode proteins that are non-covalently bound to β2 microglobulin. The classical Class II genes (HLA-DRB1, -DRB3, -DRB4, -DRB5, -DQA1, -DQB1, -DPA1, and -DPB1) encode heterodimers with two HLA subunits (alpha and beta). The structures of the Class I and Class II proteins have many similar features, the most notable being a cleft that binds peptide antigens. Complexes formed by HLA proteins with antigenic peptides bound in their ARDs are monitored by TCR selected to have exquisite sensitivity for the presence of abnormal peptides caused by disease states such as malignancy and infection [17,18]. A key difference between Class I and Class II proteins is that the majority of antigenic peptides bound by Class I HLA proteins are from inside the cell while those binding Class II proteins are from outside the cell [19,20].

HLA genes are the most polymorphic genes in the human genome [16] with 30,522 alleles currently recognized by the WHO Nomenclature Committee for Factors of the HLA System (Fig. 1, http://hla.alleles.org/nomenclature/stats.html). The official repository for curated HLA sequences is the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/). This database is updated every three months and the most recent version, 3.44.0, 2021-04-20, contains the sequences (partial or full-length) of 30,522 HLA alleles (Fig. 1) [15]. The quarterly updates include new alleles, corrections in sequences listed in prior versions of the database, changes in allele names such as a suffix change, and removed alleles (usually attributed to errors in sequencing). The number of sequences in the reference database is expected to increase *ad infinitum* because evolutionary biologists estimate that millions of HLA alleles exist in the population and there is evidence that *de novo* mutations in the germ line will continue to introduce novel HLA alleles into the population [21,22].

HLA typing uses laboratory data to assign a phenotype or genotype and these HLA types play an essential role in understanding how small differences in the HLA genes and the proteins they encode underlie HLA's role in health and disease [16]. It can be challenging to use HLA typing data because typing methods, the reagents used for typing, and the sequence databases used to interpret typing data are constantly changing, and these changes can influence typing results. This review highlights key factors to consider when using HLA typing data. Examples are provided to illustrate how differences in HLA typing that have occurred over many years can have significant impact.

2. The landscape is diverse

HLA genes are widely recognized as the most polymorphic genes in the human genome [16]. Analysis of HLA's variation has revealed (1) evidence for intra- and inter-genic recombination, (2) variation at more than 40% of nucleotides located in coding regions, (3) many alleles that differ by a single nucleotide, (4) large homopolymer regions and tandem repeats of variable length, and (5) insertions and deletions [15,22,23]. Although recombination events have contributed to HLA diversity, the rates of recombination in this region of the genome are relatively low and show population-specific recombination sites [24]. For this reason, HLA genes are often inherited as haplotypes which contain the Class I and Class II genes on a single maternal or paternal chromosome [25,26].

The WHO Nomenclature Committee of Factors for the HLA System is responsible for naming HLA alleles. The Committee's system for naming HLA alleles has been revised over time as knowledge of HLA polymorphism has improved [15,27]. The current nomenclature system uses two to four fields of numbers to name each allele (Fig. 2). The first two fields of numbers are used to name an HLA protein with a unique amino acid sequence. The first of these fields originates from serological typing assignments and has evolved into defining a family of alleles which have some similarity to the alleles initially assigned to the field. The second field distinguishes alleles within the family that encode unique amino acid sequences. The third field is used to name alleles that encode the same amino acid sequence but have silent mutations in the coding region. The fourth field is used to name variations in introns, 3′ untranslated regions, and 5′ untranslated regions. Alleles with abnormal expression have a suffix consisting of a letter that describes expression variants. For example, the suffix "N" is used for null alleles which are particularly important because these alleles are not expressed as an HLA protein [28]. Today, approximately 4% of the HLA alleles that have been named are not expressed http://hla.alleles.org/nomenclature/stats.html).

A functionally relevant factor that is often overlooked is that HLA-DQ and -DP proteins are heterodimers that are both encoded

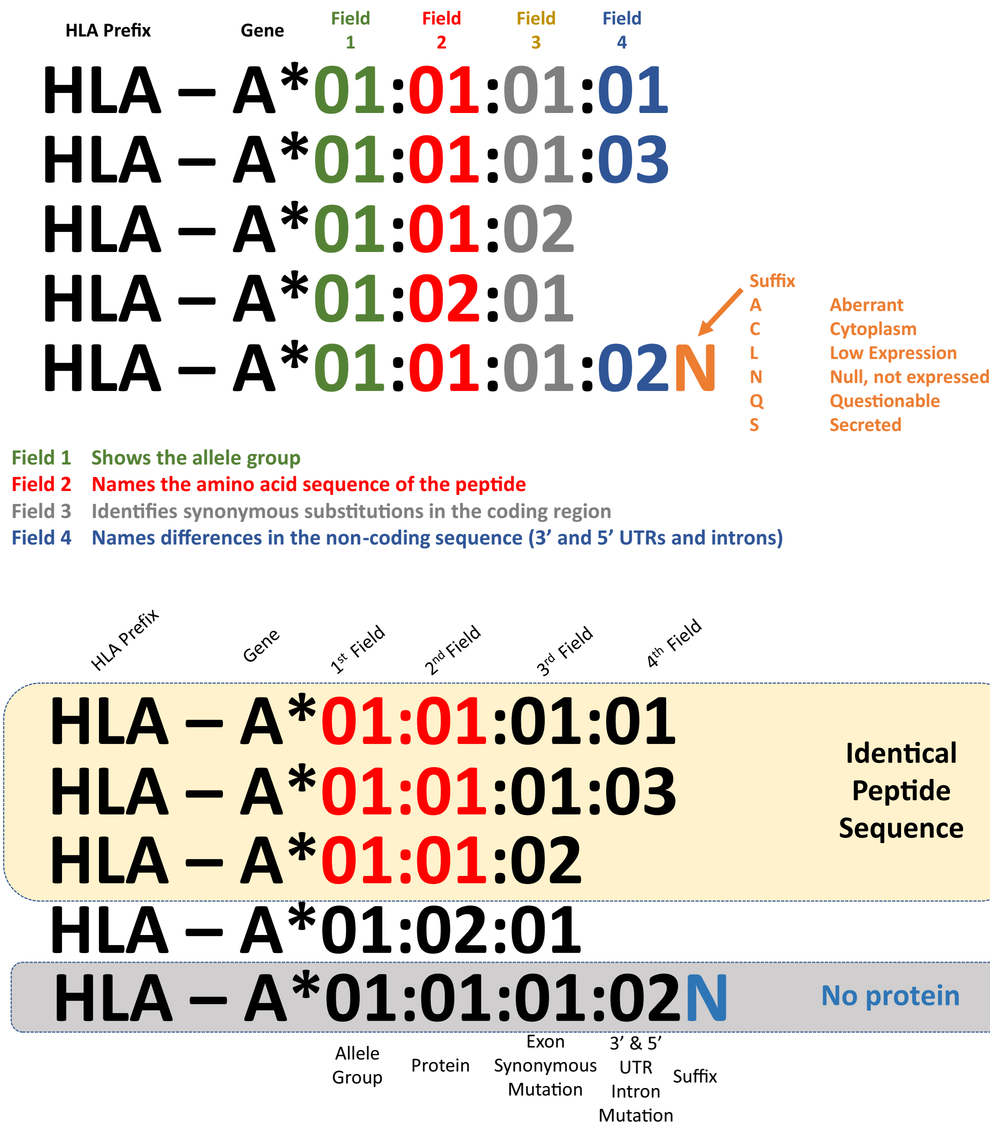

Fig. 2. Nomenclature for HLA alleles since 2010. Top. All allele names begin with a prefix (HLA), a separator (*), and the HLA locus. HLA alleles are named using two, three, or four "fields" which are separated by colons (:). The third and fourth fields are only used to name non-coding differences. If protein expression is likely to be abnormal, a suffix is added to the end of the name. Bottom. Alleles that have the same first two fields but differences in the third or fourth field encode the same peptide sequence. Null alleles are not expressed as the peptide defined by the first two fields.

by polymorphic genes (Fig. 3). These heterodimers have two subunits (alpha and beta) that form the right and left sides of the HLA protein. Although the complete HLA protein and the ARD are formed by the combination of these two subunits, most scientific publications using HLA-DQ or -DP genotyping data analyze the HLA-DQB and HLA-DQA genotypes independently rather than the heterodimeric proteins that are expressed. The problem is illustrated by determining HLA matching between transplant recipients and their donors. When HLA genotypes for individual genes are considered, there are two alleles for each gene resulting in 0, 1, or 2 HLA matches for each gene. However, as shown in Fig. 3, combinatorial diversity can generate up to four different HLA proteins (i.e., 0, 1, 2, 3, or 4 matches). When four proteins are expressed, all are functionally relevant and need to be considered in clinical and research settings. Although it is possible for an individual

with two HLA-DQA and two HLA-DQB alleles to express up to four different HLA-DQ or -DP heterodimers, there are some constraints on pairing because some combinations are unstable [29,30].

For decades, literature has focused on HLA diversity located in the ARD, but recent research suggests that variation in other parts of the protein and the entire length of HLA genes is functionally important [31–33]. Protein variation outside the ARD can impact trafficking, levels of expression, signaling, and interaction with other membrane proteins [33,34]. Non-coding differences impact levels of HLA expression and regulate expression of other genes [31,35–37]. Recent discoveries suggest that the non-coding variation could have physiological consequences that exceed those of the coding variation [36]. As our understanding of HLA diversity and its functional relevance continues to evolve, our perspective of HLA typing data is likely to change. Until the role of variation in full-length HLA genes is better understood, it is reasonable to use caution in exclusive use of data from coding sequences, particularly for investigation involving an association between HLA and disease.

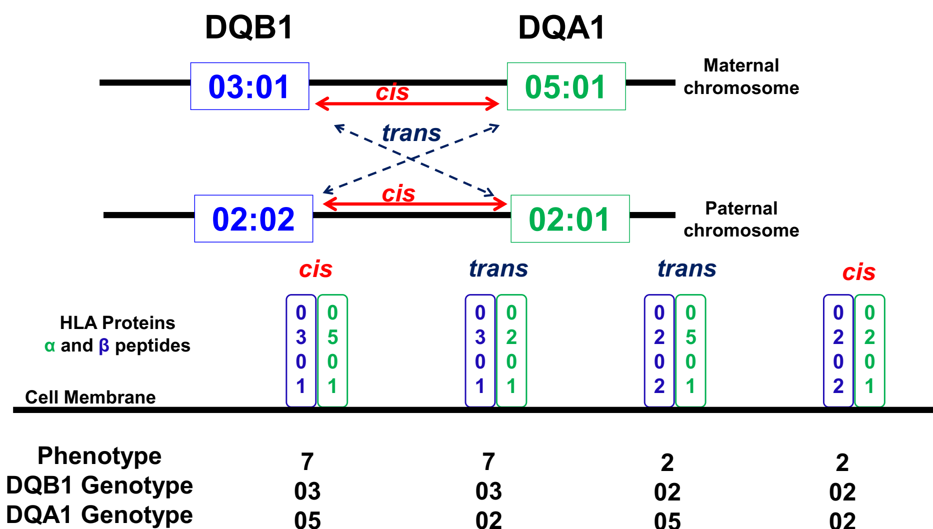

Fig. 3. HLA-DQ Genes and Proteins. Top. Each chromosome has a DQA1 gene and DQB1 gene which encode HLA peptides that combine to produce a heterodimeric protein (encoded in *cis*). The gene products of the HLA genes encoded on the other chromosome can also combine with these HLA subunits to produce different heterodimers that are encoded in *trans*. However, there are some constraints on expression of *trans*-encoded proteins because some combinations are not stable. Bottom. Phenotypes and genotypes are shown for the four proteins that are encoded by the chromosomes shown at the top.

Another important aspect of HLA variability is differences in levels of HLA expression which have been observed for particular haplotypes, allele families, or individual alleles [37,38]. Importantly, differences in HLA expression have been implicated in health and disease. Examples include reports of a relationship between levels of HLA-C expression and HIV control and relationships between levels of HLA-C and HLA-DP expression with transplant outcomes [39–41].

3. The HLA phenotyping path

The first HLA typing method utilized antibodies isolated from humans after transfusion or pregnancy to detect differences in HLA proteins. The first hints of HLA diversity were provided by studies showing that these antibodies reacted with cells from a subset of the population and that patterns of reactivity segregated in families. In 1965 researchers from many labs participated in a collaborative workshop to characterize these antibodies and develop cell panels that could be used to define HLA differences [42]. This work led to the first HLA Nomenclature report in 1968 which described eight HLA antigens [43]. Investigation during the next 20 years led to development of a serological HLA typing system which provided the foundation for naming HLA alleles into allele families [27].

The 1965 international Workshop which played a key role in the initial development of HLA typing systems was the first of many workshops that have brought experts together to conduct research that address challenges in the field that would benefit from large-scale research efforts. Accomplishments of the most recent Workshop (17th International Histocompatibility Workshop) include development of tools for collecting and analyzing HLA data, advancing knowledge of full-length sequences of HLA genes, and mapping of serologic epitopes [44]. Work for the 18th International Histocompatibility Workshop is underway and is expected to make considerable progress in addressing current challenges.

HLA phenotyping using serological methods detects differences in HLA proteins and is dependent upon the epitopes of the antigens recognized by the antibodies present in the typing reagents. Interpretation of primary typing data is not straightforward in part because a single HLA phenotype encompasses many different proteins. When serological reagents became available to detect some of these differences, subtypes were created. The naming of the subtypes reflects chronology of naming rather than similarity of proteins. For example, A9 phenotype was divided into A23 and A24. As a result, A9 is considered "matched" with A23 but A23 and A24 which are subtypes of A9 are considered to be different from each other. Additional complexity is caused by differences in typing reagents and data interpretation. For example, several different phenotypes were reported to the National Marrow Donor Program (NMDP) for a single genotype, B*15:02, including B75 (62%), B62 (22%) or B15 (7%). B75 is a subtype of B15, but B75 and B62 are different phenotypes [45].

A common perception is that alleles that have the same phenotype are always more similar to each other than those with a different phenotype. While this is often the case, there are exceptions. For example, using the current reference database the A*01 allele family encodes 377 proteins and after excluding those with a frameshift difference, there are some proteins encoded by A*01 alleles that differ by up to 8 amino acids. However, the alleles encoding most common protein in the A*01 family, *A*01:01*, encode a protein that has only three amino acid differences from protein encoded by the most frequent alleles in the A36 allele family, *A*36:01*.

Serological typing methods have inherent limitations. Serological typing is dependent upon availability of suitable antibody reagents [46]. Extensive screening of alloantisera is required to identify reagents that reflect the reactivity of the defined phenotypes. Many sera cannot be used for typing because they contain HLA antibodies that have reactivity that is different from the defined types. Typing sera obtained from sensitized individuals are often polyclonal and can lead to complex reaction patterns. Monoclonal antibodies have also been developed, but it has not been possible to develop monoclonal antibodies against every phenotype, in part because some phenotypes reflect a combination of epitopes in the HLA protein. Other challenges include variability caused by differences in technologist scoring, inability to detect HLA proteins that are expressed at a low level, influence of cell viability, and subjective interpretation of complex patterns of reactivity.

When DNA-based typing methods became available, it became evident that serological typing is often inaccurate. An NMDP study comparing HLA-A and HLA-B phenotypes and genotypes determined for 42,160 volunteer donors concluded that 24% of the types were discordant for one or both loci [47]. Nearly all of these discrepancies were attributed to incorrect phenotypes. HLA-C error rates are higher because many HLA-C alleles encode proteins that are not detectable using serological methods, in part because no serological specificities were assigned for HLA alleles that are named as C12-C18 allele families. For example, a study of 604 Caucasians reported a 37% error rate in HLA-C phenotypes [48]. A study of 7000 HLA-DR typings reported that 25% of the phenotypes were incorrect [49]. Given these high error rates, scientific publications that utilized serological typing data should be scrutinized to determine if inaccurate typing could affect the conclusions. For contemporary research, including subjects with such inaccurate typing data is highly questionable.

4. The HLA genotyping path

In the 1980′s, advances in molecular biology made it possible to determine the nucleotide sequences of genes encoding different phenotypes. One of the surprises was that when HLA genes from individuals with same HLA phenotype were sequenced, the nucleotide sequences were not always identical and could encode different HLA proteins. Libraries of nucleotide sequences were created and used to design molecular biology reagents that could be used to assign HLA genotypes that corresponded with HLA phenotypes and to detect differences that could not be identified using serological typing methods. HLA typing using serological or DNA-based methods that approximate serological typing have been described as low resolution or antigen level typing. DNA-based methods that detect difference within phenotypes that have been referred to as intermediate or high-resolution typing, depending upon the number and frequency of alleles that are included within the assigned type.

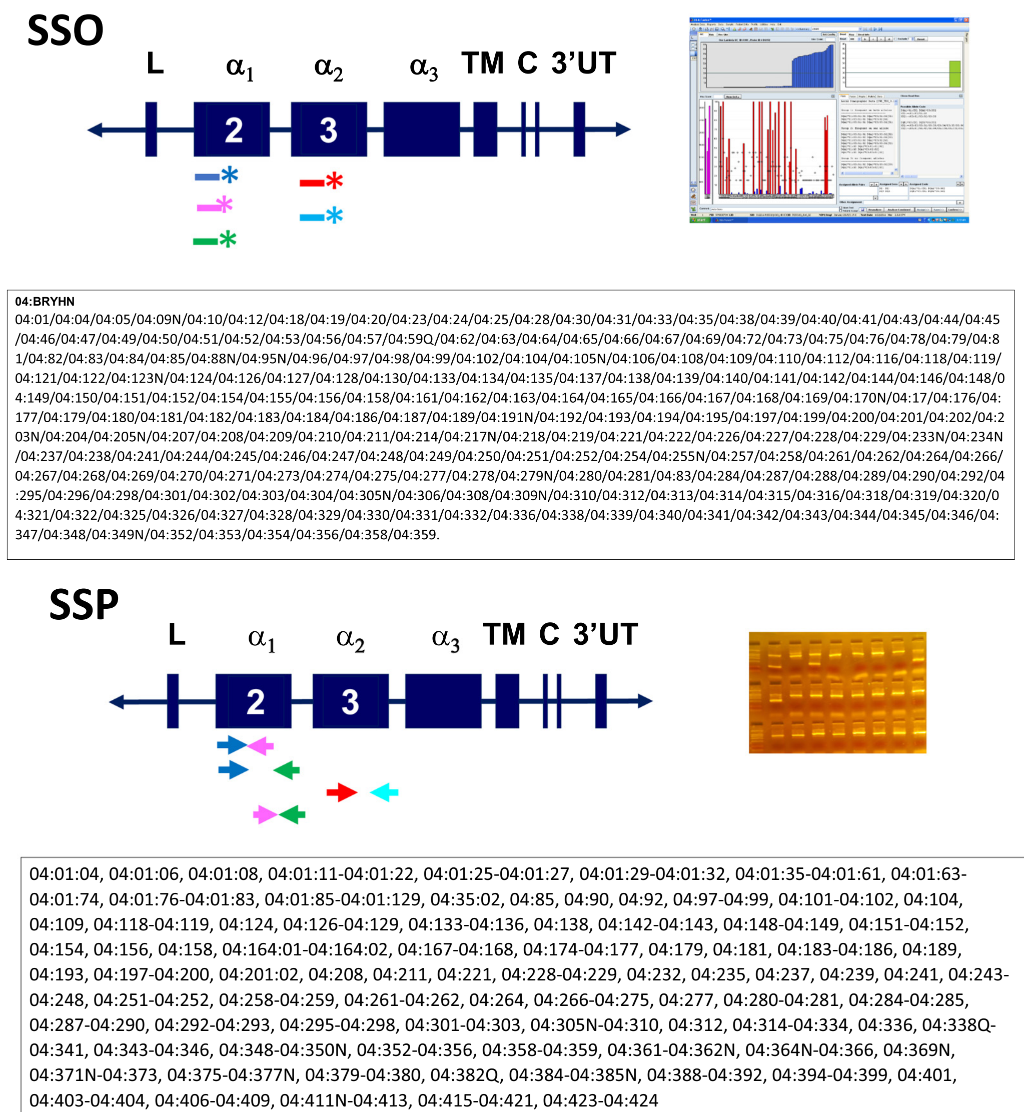

Fig. 4. SSO and SSP Typing Methods. The exon structure of a Class I HLA gene is shown in dark blue. A. SSO uses labeled probes that can be hybridized to amplified DNA to detect key sequences for assigning HLA types (motifs). Probes are frequently designed to detect several alternative motifs at a particular location of the gene. The right side shows an example of the data viewed in software used to assign HLA types based upon positive and negative hybridization of each probe. The NMDP code for this list of ambiguous types is BRYHN. B. SSP uses PCR primers that are specific for key sequences used for HLA typing. Primers for a particular HLA motif(s) and primers for an internal control are incubated with genomic DNA. After PCR there is an amplicon from the positive control. When the sequence of both HLA primers in a reaction mixture is present the genomic DNA, an amplicon of the expected size is generated. The size and quantity of amplicons can be detected by visualizing the amplicons on an agarose gel (right) or by measuring PCR products using real-time PCR (not shown). An example of the list of possible types is shown at the bottom.

The earliest DNA-based typing methods took advantage of the polymerase chain reaction (PCR) to selectively amplify segments of HLA genes and hybridization probes to detect key sequence motifs in amplified DNA (Fig. 4). This method is often described using acronyms for sequence-specific oligonucleotide probe hybridization such as SSO, SSOP, and SSOPH. Another early method took advantage PCR primers that will generate a PCR product only if the primer's sequence is present in an HLA allele. This method is often described as SSP which is an acronym for sequence-specific priming. A key feature of SSP and SSO methods is that they detect a few key sequence motifs and use the presence or absence of these motifs to assign HLA genotypes. In some respects, this is similar to serological typing methods that detect the presence of HLA proteins that have the epitopes of antibodies in typing sera. However, SSO and SSP are more accurate and precise than serological typing methods. As a result, these DNA typing methods quickly became the new gold standard. Although SSO and SSP methods remain in widespread use today, we now know that routine SSO and SSP typing methods have significant limitations in their ability to identify HLA differences that are functionally important [3,31]. One of the constraints of these methods is that it is not practical to use reagents that would be capable of detecting all known polymorphic nucleotides and their phasing. Additionally, novel alleles distinguished by previously unreported nucleotide sequences or phasing of established motifs may not be detectable.

In the 1990′s, the Sanger method for nucleotide sequencing along with automated sequencing technology provided tools for sequencing-based HLA typing, often described as SBT (Fig. 5). Routine typing was usually limited to segments of HLA genes because sequencing of full-length genes was cost and time prohibitive. Most SBT methods determined sequences of exons encoding the antigen recognition domain. SBT quickly became the gold standard because the sequencing revealed that SSP and SSO typing methods did not detect some HLA differences that are functionally important.

A major limitation of routine SSO, SSP, and SBT methods is that interpretation of the primary data using a specific version of IPD-IMGT/HLA database often matches multiple alternative genotypes which are referred to as HLA typing ambiguities (Figs. 4–7). SSO and SSP which use the presence or absence of certain sequence motifs to assign HLA types are at highest risk for ambiguities, but any method that does not determine the entire sequence of single alleles can have typing ambiguities. One cause of ambiguities is that there are alleles that differ outside the sequence that is interrogated by the typing method. Another frequent cause of ambiguities is differences in phasing of sequences that define the alleles (Fig. 7).

In some situations, data from several methods can be combined to resolve an ambiguous typing result. However, this approach is expensive, requires expertise to select appropriate methods and reagents, and can increase turn-around-time. Alternatively, there are several options for reporting ambiguous HLA genotypes (Figs. 4 and 5).

- Report the entire list of possibilities. This is not a desirable option when the list of possible types is long or when there are several common types in the list.
- Infer the most likely HLA genotype from the list of possibilities. A major problem with this approach is that the genotype that is assigned may be inaccurate and users are not alerted to limitations in the typing.
- Report the genotype using allele codes that were developed by the NMDP [27]. In this system, a code defines a list of possible alleles. The code can be affected by the version of the IPD-IMGT/HLA database that is used for interpreting the primary data. As the database expands, the list of possible alleles grows and each of these lists has a different code.
- Use P and G options which were introduced by the HLA Nomenclature Committee in 2010 [50]. A "P" is assigned to alleles that encode identical ARD (i.e., exons 2 and 3 for Class I genes and

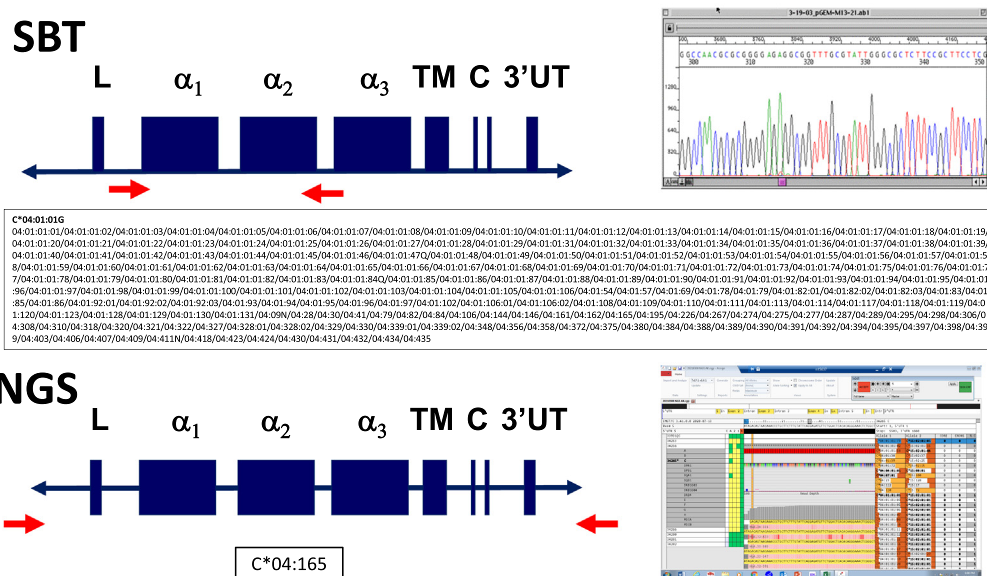

Fig. 5. SBT and NGS Typing Methods. The exon structure of Class I HLA genes is shown in dark blue. The top section shows the part of the HLA gene which is often sequenced using automated sequencing based upon the Sanger method. An example of a section of the sequencing chromatogram which is used to assign HLA types is provided at the right. The peaks show the signals for the nucleotides at each position. Most of the SBT methods use the PCR to amplify segments of the alleles for a particular HLA gene. The product is the combined sequence for a segment of the HLA gene. The combined sequence can have two nucleotides at the same position, one derived from each allele. The list of alleles shows the result for a typing with ambiguities caused by sequencing only a segment of the HLA gene. The bottom shows that most of the NGS methods determine the sequence for full-length genes or the majority of the gene. An example of the software used to interpret the sequencing data is provided at the right. An example of a relatively uncommon type that would be reported as HLA-C*04:01 by many methods is shown at the right.

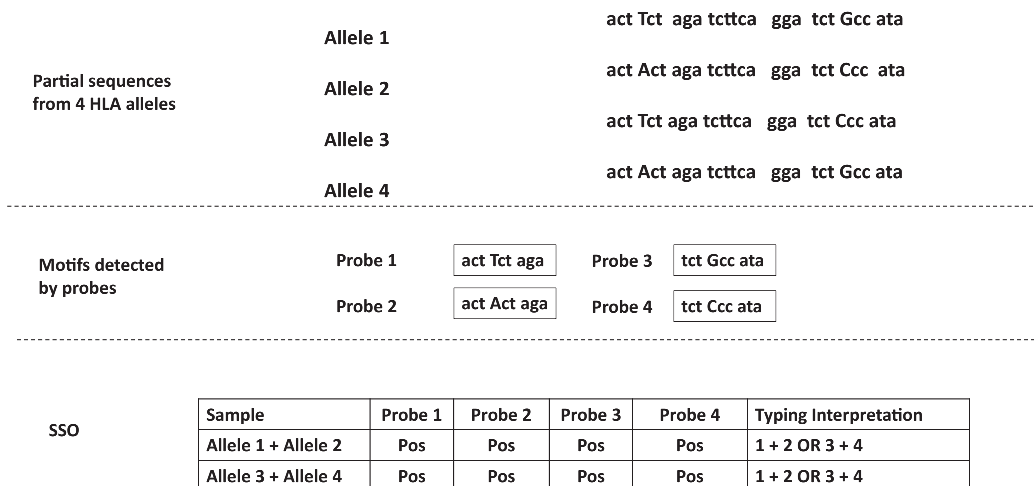

Fig. 6. Typing Ambiguity Using SSO Methods. The top panel shows the sequences of 4 alleles which are distinguished from each other at two positions shown in upper case. The middle panel shows the probes designed to detect the four motifs in these alleles. The bottom panel shows that the SSO result of a sample with alleles 1 and 2 is identical to a sample with alleles 2 and 3. These should be reported as ambiguous types.

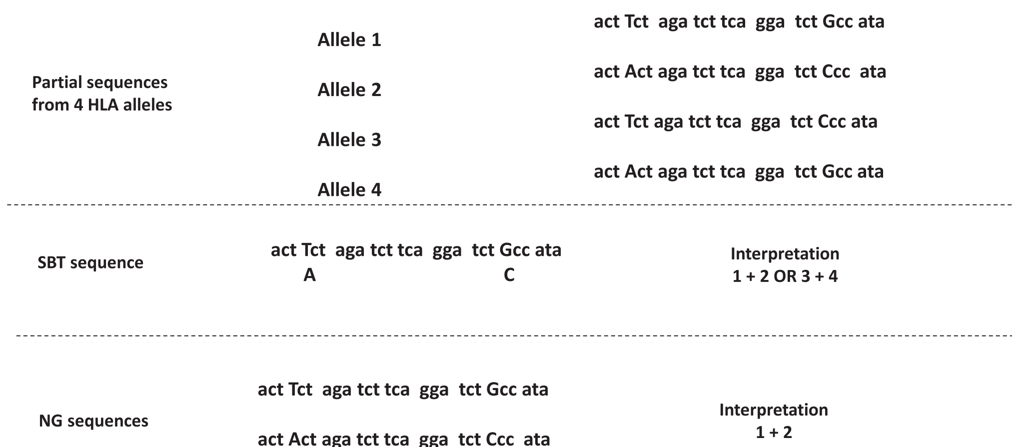

Fig. 7. Example of Typing Ambiguity Caused by Phasing. The top panel shows four sequences that differ from each other by the combination of A or T in codon 2 and C or G in codon 8. For SBT methods that determined the combined sequences of two alleles from a particular locus, this situation creates two alternative interpretations of the data. For NGS methods that determine the sequence of each allele, phasing is known.

exon 2 for Class II genes). In this system, the lowest numbered allele in the list is followed by a P (e.g., A*01:01P). Null alleles are excluded from P groups. An alternative is to assign a "G" to a group of alleles that have identical nucleotide sequences in the exons encoding the ARD (Fig. 5).

- Use a catalog of common, intermediate, and well-documented (CIWD) alleles to limit the number of alleles that are reported based upon their published frequency [51]. This catalog was created using genotypes determined by SBT or NGS methods for more than eight million individuals to classify HLA genotypes as common (≥1:10,000), intermediate (≥1:100,000), well-documented (≥5 observations), or not CIWD. Alleles in the IPD-IMGT/HLA version 3.31.0 reported in January 2018 were considered. Some genotypes were merged into P or G groups because data included in the analysis included sequencing that was limited to exons 2 and 3 for Class I and exon 2 for Class II genes. Limitations of this approach include poor representation of some populations in the CIWD dataset and the fact that non-CIWD options remain a formal possibility.

In 2010 the first reports of using next generation sequencing (NGS) for HLA typing appeared [52,53]. This technology makes it practical to determine nucleotide sequences of entire HLA genes because NGS is substantially less expensive than SBT. Many typing ambiguities are eliminated by NGS because it establishes the phasing of the majority of the polymorphic sequences that define each allele. However, ambiguities remain a limitation when the phasing differences are separated by distances that exceed the length of fragments of sequence that are determined by the method. These ambiguities can be resolved by using a third- generation sequencing method that achieve substantially longer reads [54]. Although most of the current commercial reagents for NGS typing require several days to complete typing, there are recent reports of NGS methods that can determine high resolution HLA types in less than six hours [55,56]. Given these substantial improvements, NGS quickly became the gold standard for HLA typing.

The gold standard for typing has made stepwise advances with each step offering improvements in accuracy and ability to identify more HLA differences: (1) serology, (2) DNA-based typing using SSO and SSP, (3) SBT, (4) and NGS. Each step has also advanced the understanding of HLA polymorphism and its clinical significance. Knowledge generated by NGS is providing insights of the functional impact of amino acid differences throughout the entire HLA protein [57]. Sequencing of introns, 3′ non-coding regions,

and 5′ non-coding regions has revealed that differences in non-coding sequences are also important because they can affect HLA expression and regulate expression of other proteins [31,35].

Today, HLA genotyping resolution is often described using the four fields used for naming alleles. In this system, low resolution typing is described as first field or 1F typing. Typing that resolves HLA types to a single protein sequence (formerly high resolution or allele level typing) is described as second field or 2F typing. Resolution beyond this is described as third field (3F) or fourth field (4F) typing depending upon the typing resolution.

It can be confusing to use typing data generated in different eras. Historical high-resolution typing which identifies types corresponding to a single HLA protein sequence or a single common sequence was also referred to as "allele" level resolution, but rarely resolved types to the equivalent of a single allele. Today, "allele" level resolution would be described as 4F resolution. This can still be confusing because some alleles do not have 4F names. When a 2F or 3F name is sufficient to name a single nucleotide sequence (i.e., no non-coding differences discovered), the name is limited to these fields even though the entire gene sequence is known.

Many factors affect the number of fields in an HLA genotype which can cause misconceptions about the significance of the number of fields used for reporting an HLA genotype. A clarification of some of the frequent misconceptions are listed below.

- The number of fields in a genotype does not indicate gene coverage. For example, partial gene sequence can be used to report a 3F or 4F type.
- A 2F genotype does not indicate that the reference type is a partial sequence. Many HLA alleles have 2F names because there have been no reports of alleles that have differences that would require having a third or fourth field genotype.
- A 2F type does not indicate that typing data were limited to coding regions or that there are no alleles that differ in the third or fourth field.
- A 3F genotype does not indicate that the entire coding sequence has been determined.
- A 4F genotype does not mean that the entire gene was sequenced. It indicates that there is only one allele in the reference database that is a perfect fit with the available data, but these data could be limited to sequence motifs detected by SSO/SSP or partial sequence determined using SBT or NGS.
- The number of fields reported is not an indication of typing accuracy.
- HLA genotypes are not static. Assignment of HLA types is affected by the version of the reference database that is used to interpret the primary data. Systems have been created to collect primary data to allow reinterpretation as the reference data base changes, but this approach requires reinterpretation of the primary data and its use is generally limited to large registries that have stored primary data for purposes of reinterpretation [58].
- Genotypes that differ in the first field are not more different from each other than types that differ in the second field. The first field of an allele name defines a family of alleles that encode a serological type or have sequence homology to other alleles in the allele family with this serological type. However, the number of amino acid differences between two alleles and their functional consequences is dependent upon the alleles in question. The majority of first field differences have more amino acid differences than those that are identical for the first field but differ in the second field, but this is not always the case [59].

Until recently, HLA genotyping has been focused on the exons encoding the ARD because the cost and time required to characterize the entire gene was prohibitive. The rationale is that the ARD is responsible for the most important functions of the HLA protein including binding peptides, contacting T cell receptors, and contacting KIR. However, there is growing evidence that other domains of the HLA protein and the entire gene can have important functions [31,35,57]. For example, there are allelic differences that affect splicing or change the borders between the introns and exons which may not be evident from the type. This is illustrated by the allele HLA-B*44:02:01:02S which could be reported as B*44:02. This allele has a splice site mutation which causes deletion of some exons. Another example is that some HLA-DQB1 alleles have a variably expressed exon and some HLA-DQA1 alleles generate multiple transcripts due to alternative splicing and use of different polyadenylation sites in the 3′ untranslated region [60–64].

With the advent of routine NGS typing, the field is poised to advance our understanding of the clinical significance of non-coding diversity. For example, recent reports suggest that HLA matching of exons outside of the ARD, introns, and untranslated regions can significantly improve outcomes of hematopoietic stem cell transplants [57,65]. Investigation in solid organ transplantation has revealed that complexes between HLA proteins and integrins can impact rejection through their role in intracellular signaling and immune regulation [33,66]. There is evidence that cytoplasmic tails of HLA Class II proteins can influence intracellular trafficking and function [67]. The non-coding regions of HLA genes contain functional regulatory elements that control levels of HLA expression and impact the expression of many non-HLA genes [31,32].

Another development which is providing new HLA insights is RNA sequencing (RNASeq). RNASeq data can be used for HLA typing but its greatest benefit may be to provide other information such as alternative splicing of transcripts, post-transcriptional modifications, mRNA levels, and sequence binding sites in the mRNA generated by the HLA genes [68–70]. RNA-Seq can also be used to measure non-coding RNAs such as microRNAs and long non-coding RNAs (lncRNAs) which have high potential to play an important role in the relationship between HLA and disease [31,32,71].

5. Challenges created by changing the landscape

Phenotypes and genotypes which have been determined and reported over decades are fundamentally different because one typing system is based upon epitopes detected by HLA antibodies and the other is based upon gene sequence and predicted amino acid sequence. The first field of a genotype is frequently used as the equivalent of a phenotype, but there are some exceptions. The names of some allele families are linked to broad serological types rather than the subtypes. For example, the HLA-B15 allele family is associated with several phenotypes (i.e., B62, B63, B70, B71, B72, and B75). Beyond this there are many alleles (particularly HLA-C and HLA-DP) which would not be detected using serological typing methods (i.e, blank). In general, HLA genotypes with the same first field type encode HLA proteins that have amino acid sequences that are more related to each other than to those that have a different first field type, but there are exceptions [59]. This is demonstrated by an example described earlier. The HLA-A*01:01 genotype encodes only three amino acid differences from HLA-A*36:01 genotype. However, there are proteins encoded by alleles classified in the HLA-A*01 family which differ from HLA-A*01:01 genotype by eight amino acids, excluding alleles that have a frameshift.

Several tools have been developed to aid in relating HLA phenotypes and genotypes, including the HLA dictionary (https://www.ebi.ac.uk/ipd/imgt/hla/dictionary.html) and ALLAN (ALLele to Anti-

gen, (http://www.transplanttoolbox.org). The HLA dictionary contains HLA types reported for 832 alleles that were available from several sources (HLA Nomenclature Committee, the UCLA International Cell Exchange, the National Marrow Donor Program, experts, and a neural network program) [45]. ALLAN maps unambiguous and ambiguous HLA genotyping data to UNOS antigen equivalents [72]. Tools such as these can be useful to improve consistency of relating genotypes and phenotypes, but the underlying variables which influence the accuracy of typing assignments are confounding factors.

Another challenge is that genotype assignments are dependent upon the reference database used for interpreting the primary data, but the reference database (IPD-IMGT/HLA) is dynamic with quarterly updates providing names of new alleles, corrected or extended nucleotide sequences, retracted allele names, and revised names. Some donor registries have addressed this problem by storing primary typing data rather than assigned types [58]. These data can be used to dynamically update the types. Some regulatory agencies require that typing reports include the version of the IPD-IMGT database used for interpreting the primary data, however it is uncommon for users to consider the fact that the typing assignment could change if a different database were used for interpreting data.

Another factor that can affect genotype assignments is that quality of the sequencing data used to assign allele names is variable. For alleles that have high frequencies, nucleotide sequences are usually accurate and cover the entire gene. For less common alleles, quality is variable. Gene coverage ranges from partial (e.g., one exon) to full-length gene sequences. Additionally, some sequences have not been verified. There is a formal possibility that unverified sequences were generated by sequencing errors or mutations in abnormal cells (e.g., malignancies). In fact, IPD/IMGT/HLA updates frequently list names that have been withdrawn after sequencing errors are identified.

The use of a dynamic reference database with continual expansion of the number of alleles and improved sequence quality (e.g, more full-length sequences) creates challenges for clinicians and investigators because the same data can be interpreted differently depending upon the version of the database that is used as the reference. This is particularly evident when motifs detected by probes and primers or partial gene sequences are used to assign HLA genotypes, because the typing assignment ignores all uninterrogated nucleotide sequence. For example, a probe detecting a motif that is present in only one allele in the database used for interpretation could be used to assign the HLA type. However, a subsequent database could show that this motif is present in several additional alleles which would cause the original data to have several alternative interpretations. Another factor that can alter typing assignments when different databases are used is phasing of motifs. For example, a small reference database could have only one option for phasing, but a subsequent larger database could reveal alleles that have the same motifs with different phasing as illustrated in Fig. 6. As a result, if additional typing is performed to resolve the ambiguities, the updated HLA type could be different from the original. An example of this situation in transplantation is that a donor-recipient pair could be typed as HLA matched and reclassified as HLA mismatched upon retyping [73].

There are some on-line tools designed to predict HLA types based upon lower-resolution HLA typing data. Easy-HLA infers updated or missing data for HLA types that are entered into the tool (https://hla.univ-nantes.fr) [74]. Another tool, HATK generates tables of alternative names (old, standard, G-group, P-group) for HLA alleles based on the HLA nomenclature for a specific version of the database [75]. It also has a tool for converting HLA types to updated naming conventions. An EMBL-EBI site (https://www.ebi.ac.uk/ipd/imgt/hla/access.html) offers a variety of tools including one that converts allele names to current nomenclature, a tool that can be used to determine the number of alleles with a particular polymorphic motif, and comparison of search determinants used by different registries in 2007 [15]. A typing resolution score (TRS), was developed to compare HLA typings across different methods, data sets and populations [76]. This score evaluates the ambiguity or uncertainty in typing and is a component of the HaploStats web application (http://www.haplostats.org). Another application, HLA Haplotype Validator (HLAHapV) was developed to identify potential typing errors by highlighting rare or unexpected HLA types assigned by NGS typing methods [77].

Novel alleles are encountered during routine HLA typing but these may not be immediately submitted for naming by WHO Nomenclature Committee for Factors of the HLA System. One transplant center laboratory that is not a high volume lab reported discovery of 10 novel alleles during the first four months of using NGS typing methods [78]. High volume labs encounter thousands of novel alleles during their routine work [79]. Many HLA typing labs do not routinely obtain official names for novel alleles, particularly when the lab's criteria for reporting can be satisfied using an existing allele name. For example, the typing can be reported using a G or P group if the novel substitution is located outside these coding regions. A problem is that this practice can lead to incorrect assumptions that a common allele is present. Additionally, this practice causes overestimates of allele frequencies for alleles that are identical in these coding regions and underestimates of frequencies for alleles that differ in other locations of the gene.

When there is no official name, novel alleles are reported descriptively. A problem is that the same allele could have different descriptions. Variants can be described relative to an allele family or a named allele. When the allele family is used as the reference (e.g., HLA-A*01 new), the relationship to other alleles in the family remains unknown. The Human Genome Variation Society has developed a system for describing variants in relation to an accepted reference sequence [80]. However, the same allele can be reported with different names depending upon the selection of the reference allele. For example, if a variant has one nucleotide difference from a rare allele and two nucleotide differences from a common allele, some laboratories would use the common allele as the reference sequence while others would use the most similar allele as the reference. A combination of these approaches which could yield more consistent naming would be to always use the first allele of an allele family as the reference allele. Until there are well-accepted guidelines for selecting a reference sequence, this problem will persist. To encourage labs to obtain official names, a gene submission tool was developed to facilitate submission of novel sequences to the European Molecular Biology Laboratory's European Nucleotide Archive and the IPD-IMGT/HLA [81]. Another alternative to these approaches is gene feature enumeration, but this approach has not been widely used [82].

There is a substantial body of HLA typing data which has been generated over several decades. Since these HLA types were determined using different methods and/or interpreted using different databases, it is important to consider the impact of the HLA typing data on conclusions and recommendations. This problem was recently demonstrated by Senev et al who compared NGS typing with types assigned by other methods and showed how these differences could impact research conclusions [3]. A worldwide repository for HLA typing data, Allele Frequency Net Database (AFND), has addressed the challenge of variability in HLA types data by classifying types into three categories based upon their quality [23]. This website also has tools to aid in relating HLA types generated over time, including a tool that provides lower resolution alternatives to a 4F genotype. Another tool at this site provides allele frequencies using different versions of the IPD-IMGT/HLA database.

6. The path of inferred and imputed HLA genotypes

Inferred or imputed HLA genotypes are used to predict HLA types when laboratory data are not available. These terms are often used interchangeably, but using strict definitions inferred and imputed types are not the same. Inferred HLA types use available HLA typing data along with allele and haplotype frequencies from large data sets to predict the most likely HLA genotype. Imputed HLA types use linkage disequilibrium between HLA alleles and single nucleotide polymorphism (SNP) flanking HLA genes to predict the most likely HLA type.

Inferred genotypes are often used when a high-resolution genotype is desired, but available HLA typing data are phenotypes, low resolution genotypes, or high resolution types that were not determined using a full-length gene sequence. For example, inferred genotypes are useful for selecting unrelated stem cell donor candidates for HLA typing. Tools for inferring HLA genotypes include HaploStats (http://www.haplostats.org), Easy-HLA [74], and HATK [75].

Inferred HLA genotypes are most useful for populations with large databases that have high quality typing data but are less reliable for multiethnic populations and populations with limited or nonexistent HLA allele frequency data [3,23,83–85]. For example, one study compared high resolution HLA types determined using SBT with those predicted from low resolution types using HLA Matchmaker and Haplostats. [83] In this study, Haplostats showed concordance in only 36% of the HLA-A, -B, -C, -DRB1, and DQB1 types [83]. For non-Caucasians, concordance was only 29%. In this study, HLAMatchmaker which is frequently used to describe potential epitopes in HLA proteins was unable to provide high resolution HLA haplotypes for 45% of Caucasians and 64% of non-Caucasians. Of those with high resolution types provided for all five loci, only 10% of Caucasians and 5% of non-Caucasian types had 10 concordant HLA types.

For many tools, the most probable high resolution multi-locus genotype is determined but the probability can be low. For these tools, a winner-take-all approach for genotype assignment can be error-prone, especially using lower resolution typing data. This is not a limitation of the algorithms or the dataset. If high resolution typing were performed for 100 individuals with same low-resolution HLA types and the same race/ethnicity, several high resolution multi-locus genotypes are likely to be observed, but a winner-take-all approach would give the same high resolution genotype as most probable for all 100. An alternative approach is to consider the probabilities of multiple options when using inferred HLA types [86].

Although inferred HLA genotypes are appealing in situations where available data are limited to low or intermediate resolution, caution is warranted because errors in inferred genotypes can significantly impact research conclusions and clinical decision-making. This is illustrated by a retrospective study of HLA matching in kidney transplantation [3]. In this study donor specific antibodies (DSA) were assigned using inferred high-resolution genotypes and then compared with those assigned using high resolution genotypes determined using NGS. The discordance rate was high; NGS typing showed that 23% of DSA assigned using inferred genotypes were incorrectly assigned. When patients were DSA positive by both methods, 79% were diagnosed with antibody mediated rejection within five years. In contrast, none of the patients that were misclassified using inferred genotypes experienced humoral rejection within five years after transplant. Ten-year graft survival was 62% when DSA were assigned using both methods and 94% when DSA were assigned using low resolution types but absent using NGS genotypes.

Imputed HLA genotypes are used when SNP data are available, but HLA typing has not been performed. Although imputed genotypes are less accurate than genotypes determined using laboratory methods, they can be useful for using large datasets to detect associations between variation in the Major Histocompatibility Complex and disease [87]. Imputed types can be converted to amino acid sequence and used to identify associations that are dependent upon a particular feature of the HLA protein. Investigation that uncovers associations between HLA and disease can be useful to guide further research that uses accuate and precise HLA typing methods to understand the relationship.

HLA imputation accuracy is highest for individuals with European ancestry, but accuracy can be much lower for other populations [87,88]. When four SNP-based HLA methods (e-HLA, HIBAG, HLA*IMP:02 and MAGPrediction) were used to impute HLA-A, -B, -C, and –DRB1 genotypes for the Human Genome Diversity Project cell panel, concordance was greater than 89% amongst all imputation methods [88]. However, for any single method, genotypes were correctly imputed for all HLA loci for less than 28% of the panel. Imputation accuracy was influenced by the locus, with highest accuracy for HLA-C (89.9%–94.6% depending upon the method). Accuracy for HLA-A was 89.7%–92.2%, HLA-B was 69%–77%, and HLA-DRB1 was lowest with 62.4%–70.1%. Attempts to improve accuracy by using confidence metrics were not successful. A recently reported tool, CookHLA, has reported improved accuracy, but the reference data were P-groups rather than individual genotypes [89]. The goals of the SNP-HLA Reference Consortium which is a component of the 18th International Histocompatibility Workshop are to gather data, enhance HLA imputation and broaden access to highly accurate imputation models for the immunogenomics community [90].

Several factors influence the accuracy of imputation. The density and proximity of the SNPs near the HLA genes are important factors. A lack of informative SNPs near some HLA genes along with structural variants can affect imputation accuracy, particularly HLA-DR [88]. The size and diversity of the reference panel are particularly important for HLA because allele and haplotype frequencies are influenced by ancestry [23]. There is bias against rare alleles because they are unlikely to be represented in reference panels. To further complicate the situation, chromosomes with similar SNPs can have different HLA alleles.

7. Summary

HLA typing plays an essential role in clinical and research settings, but HLA typing data are complex and have changed dramatically over time. To ensure that HLA genotyping data are used appropriately, it is important to understand the nature of available data. Factors to consider include the typing method, gene coverage, the database used for interpreting data, and criteria for assigning HLA types. Tools have been developed to synchronize variable data, but these have limitations which should be taken into consideration. Inferred and imputed HLA types can be inaccurate but can be useful in certain situations. NGS typing methods have created a new era of typing resolution and accuracy that are dramatically improving the quality of reference databases. Major improvements in technology and knowledge about HLA diversity and function are revolutionizing how HLA typing data are used. The significance of variation that affects HLA expression, alternative splicing, and regulatory factors are exciting new areas of investigation. In solid organ transplantation, the field is moving toward defining HLA antibodies by their HLA epitopes rather than HLA types. These developments have poised the field for breakthroughs that will

substantially advance our understanding of the role of HLA in health and disease.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2021.04.011.

References

[1] J. Dehn et al., Selection of unrelated donors and cord blood units for hematopoietic cell transplantation: guidelines from the NMDP/CIBMTR, Blood 134 (12) (2019) 924–934.
[2] W.R. Mulley et al., Tissue typing for kidney transplantation for the general nephrologist, Nephrology (Carlton) 24 (10) (2019) 997–1000.
[3] A. Senev et al., Clinical importance of extended second field high-resolution HLA genotyping for kidney transplantation, Am. J. Transplant. 20 (12) (2020) 3367–3378.
[4] A. Saris, K. Pavenski, Human leukocyte antigen alloimmunization and alloimmune platelet refractoriness, Transfus. Med. Rev. 34 (4) (2020) 250–257.
[5] Y. Ghodke et al., HLA and disease, Eur. J. Epidemiol. 20 (6) (2005) 475–488.
[6] P. Deitiker, M.Z. Atassi, MHC genes linked to autoimmune disease, Crit. Rev. Immunol. 35 (3) (2015) 203–251.
[7] C. Tersigni et al., Role of human leukocyte antigens at the feto-maternal interface in normal and pathological pregnancy: an update, Int. J. Mol. Sci. 21 (13) (2020).
[8] P.T. Illing, A.W. Purcell, J. McCluskey, The role of HLA genes in pharmacogenomics: unravelling HLA associated adverse drug reactions, Immunogenetics 69 (8–9) (2017) 617–630.
[9] A.J. Redwood et al., HLAs: Key regulators of T-cell-mediated drug hypersensitivity, HLA 91 (1) (2018) 3–16.
[10] F. Sabbatino et al., Role of human leukocyte antigen system as a predictive biomarker for checkpoint-based immunotherapy in cancer patients, Int. J. Mol. Sci. 21 (19) (2020).
[11] D. Chowell et al., Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy, Nat. Med. 25 (11) (2019) 1715–1720.
[12] S. Gwozdowicz et al., KIR specificity and avidity of standard and unusual C1, C2, Bw4, Bw6 and A3/11 amino acid motifs at entire HLA:KIR interface between NK and target cells, the functional and evolutionary classification of HLA class I molecules, Int. J. Immunogenet. 46 (4) (2019) 217–231.
[13] E. Graczyk-Pol et al., Role of donor HLA class I mismatch, KIR-ligand mismatch and HLA:KIR pairings in hematological malignancy relapse after unrelated hematopoietic stem cell transplantation, HLA 92 (Suppl 2) (2018) 42–46.
[14] S. Gras et al., A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned, Immunol. Rev. 250 (1) (2012) 61–81.
[15] J. Robinson et al., IPD-IMGT/HLA database, Nucleic Acids Res. 48 (D1) (2020) D948–D955.
[16] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, Annu. Rev. Genomics Hum. Genet. 14 (2013) 301–323.
[17] W.L. Lo, P.M. Allen, Self-peptides in TCR repertoire selection and peripheral T cell function, Curr. Top. Microbiol. Immunol. 373 (2014) 49–67.
[18] G.P. Morris, P.M. Allen, How the TCR balances sensitivity and specificity for the recognition of self and pathogens, Nat. Immunol. 13 (2) (2012) 121–128.
[19] M. Del Val et al., Endogenous TAP-independent MHC-I antigen presentation: not just the ER lumen, Curr. Opin. Immunol. 64 (2020) 9–14.
[20] S. Temme, N. Temme, N. Koch, Assembly, intracellular transport, and release of MHC class II peptide receptors, Methods Mol. Biol. 1988 (2019) 297–314.
[21] H.H. Li et al., Amplification and analysis of DNA sequences in single human sperm and diploid cells, Nature 335 (6189) (1988) 414–417.
[22] W. Klitz, P. Hedrick, E.J. Louis, New reservoirs of HLA alleles: pools of rare variants enhance immune defense, Trends Genet. 28 (10) (2012) 480–486.
[23] F.F. Gonzalez-Galarza et al., Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools, Nucleic Acids Res. 48 (D1) (2020) D783–D788.
[24] T.H. Lam et al., Population-specific recombination sites within the human MHC region, Heredity (Edinb) 111 (2) (2013) 131–138.
[25] I. Alter et al., HLA class I haplotype diversity is consistent with selection for frequent existing haplotypes, PLoS Comput. Biol. 13 (8) (2017) e1005693.
[26] K. Osoegawa et al., Tools for building, analyzing and evaluating HLA haplotypes from families, Hum. Immunol. 80 (9) (2019) 633–643.
[27] C.K. Hurley, Naming HLA diversity: a review of HLA nomenclature, Hum. Immunol. 82 (7) (2021) 457–465.
[28] M. Bauer et al., Frequencies and haplotype associations of non-expressed HLA alleles in ethnically diverse populations on the National Marrow Donor Program's Be The Match Registry, Hum. Immunol. 81 (10–11) (2020) 580–587.
[29] J.A. Hollenbach et al., A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer, Immunogenetics 64 (8) (2012) 559–569.
[30] W.W. Kwok et al., HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation, J. Immunol. 150 (6) (1993) 2263–2272.
[31] E.W. Petersdorf, C. O'HUigin, The MHC in the era of next-generation sequencing: Implications for bridging structure with function, Hum Immunol 80 (1) (2019) 67–78.
[32] M. Shieh, N. Chitnis, D. Monos, Human leukocyte antigen and disease associations: a broader perspective, Clin. Lab. Med. 38 (4) (2018) 679–693.
[33] N. Valenzuela et al., HLA Class I and Class II-induced intracellular signaling and molecular associations in primary human endothelial cells, Methods Mol. Biol. 1788 (2018) 23–41.
[34] E.W. Petersdorf et al., Role of HLA-B exon 1 in graft-versus-host disease after unrelated haemopoietic cell transplantation: a retrospective cohort study, Lancet Haematol 7 (1) (2020) e50–e60.
[35] N.S. Chitnis, M. Shieh, D. Monos, Regulatory noncoding RNAs and the major histocompatibility complex, Hum. Immunol. 82 (7) (2021) 532–540.
[36] N. Chitnis et al., An expanded role for HLA genes: HLA-B encodes a microRNA that regulates IgA and other immune response transcripts, Front. Immunol. 8 (2017) 583.
[37] B.S. Carey, K.V. Poulton, A. Poles, Factors affecting HLA expression: a review, Int. J. Immunogenet. 46 (5) (2019) 307–320.
[38] B.S. Carey, K.V. Poulton, A. Poles, HLA-C expression level in both unstimulated and stimulated human umbilical vein endothelial cells is defined by allotype, HLA 95 (6) (2020) 532–542.
[39] E.W. Petersdorf et al., HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation, Blood 124 (26) (2014) 3996–4003.
[40] S. Kulkarni et al., Differential microRNA regulation of HLA-C expression and its association with HIV control, Nature 472 (7344) (2011) 495–498.
[41] E.W. Petersdorf et al., Role of HLA-DP expression in graft-versus-host disease after unrelated donor transplantation, J. Clin. Oncol. 38 (24) (2020) 2712–2718.
[42] Report on the Workshop on Histocompatibility Testing, held in Leiden, August 1965. Vox Sang, 1966. 11(3): p. 385–7.
[43] Nomenclature for factors of the HL-a system, Bull. World Health Organ. 39 (3) (1968) 483–486.
[44] T.A. Vayntrub, S.J. Mack, M.A. Fernandez-Vina, Preface: 17th International HLA and Immunogenetics Workshop, Hum. Immunol. 81 (2–3) (2020) 52–58.
[45] R. Holdsworth et al., The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens, Tissue Antigens 73 (2) (2009) 95–170.
[46] G. Rodey, HLA Beyond Tears, De Novo Inc, Durango CO, 2000.
[47] H.J. Noreen 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 57 (3) (2001) 221–229.
[48] M. Bunce et al., High resolution HLA-C typing by PCR-SSP: identification of allelic frequencies and linkage disequilibria in 604 unrelated random UK Caucasoids and a comparison with serology, Tissue Antigens 48 (6) (1996) 680–691.
[49] G. Opelz et al., DNA typing: an important step forward? Collaborative Transplant Study, Transpl. Int. 5 (Suppl 1) (1992) S580–S582.
[50] S.G. Marsh et al., Nomenclature for factors of the HLA system, 2010, Tissue Antigens 75 (4) (2010) 291–455.
[51] C.K. Hurley et al., Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0, HLA 95 (6) (2020) 516–531.
[52] C. Lind et al., Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing, Hum. Immunol. 71 (10) (2010) 1033–1042.
[53] V. Bravo-Egana, H. Sanders, N. Chitnis, New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing, Hum. Immunol. 82 (7) (2021) 478–487.
[54] J.L. Duke et al., Resolving MiSeq-generated ambiguities in HLA-DPB1 typing by using the oxford nanopore technology, J. Mol. Diagn. 21 (5) (2019) 852–861.
[55] D. De Santis et al., Rapid high-resolution HLA genotyping by MinION Oxford nanopore sequencing for deceased donor organ allocation, HLA 96 (2) (2020) 141–162.
[56] T.L. Mosbruger et al., Utilizing nanopore sequencing technology for the rapid and comprehensive characterization of eleven HLA loci; addressing the need for deceased donor expedited HLA typing, Hum. Immunol. 81 (8) (2020) 413–422.
[57] E.W. Petersdorf et al., HLA-B leader and survivorship after HLA-mismatched unrelated donor transplantation, Blood 136 (3) (2020) 362–369.
[58] M. Maiers et al., Maintaining updated DNA-based HLA assignments in the national marrow donor program bone marrow registry, Rev Immunogenet 2 (4) (2000) 449–460.
[59] L.A. Baxter-Lowe et al., HLA-A disparities illustrate challenges for ranking the impact of HLA mismatches on bone marrow transplant outcomes in the United States, Biol Blood Marrow Transplant 15 (8) (2009) 971–981.
[60] J. Kralovicova, I. Vorechovsky, Position-dependent repression and promotion of DQB1 intron 3 splicing by GGGG motifs, J. Immunol. 176 (4) (2006) 2381–2388.

[61] J.J. Hoarau et al., A new splicing acceptor site and poly(A)+ sequence signal within DQA1*0401 and DQA1*0501 mRNA 3'UTR contribute to increase the extraordinary diversity of mRNA isoforms, Immunogenetics 57 (3–4) (2005) 182–188.
[62] J. Kralovicova et al., Branch site haplotypes that control alternative splicing, Hum. Mol. Genet. 13 (24) (2004) 3189–3202.
[63] J.J. Hoarau et al., HLA DQA1 genes generate multiple transcripts by alternative splicing and polyadenylation of the 3' untranslated region, Tissue Antigens 63 (1) (2004) 58–71.
[64] P. Briata et al., Alternative splicing of HLA-DQB transcripts and secretion of HLA-DQ beta-chain proteins: allelic polymorphism in splicing and polyadenylylation sites, Proc. Natl. Acad. Sci. U.S.A 86 (3) (1989) 1003–1007.
[65] I. Vazirabad et al., Direct HLA genetic comparisons identify highly matched unrelated donor-recipient pairs with improved transplantation outcome, Biol Blood Marrow Transplant 25 (5) (2019) 921–931.
[66] Y.P. Jin et al., HLA Class II-triggered signaling cascades cause endothelial cell proliferation and migration: relevance to antibody-mediated transplant rejection, J. Immunol. 200 (7) (2018) 2372–2390.
[67] J. Thibodeau, M.A. Moulefera, R. Balthazard, On the structure-function of MHC class II molecules and how single amino acid polymorphisms could alter intracellular trafficking, Hum. Immunol. 80 (1) (2019) 15–31.
[68] R. Orenbuch et al., arcasHLA: high-resolution HLA typing from RNAseq, Bioinformatics 36 (1) (2020) 33–40.
[69] Y.J. Jin et al., Differential alternative splicing regulation among hepatocellular carcinoma with different risk factors, BMC Med. Genomics 12 (Suppl 8) (2019) 175.
[70] C.E. Voorter et al., The role of gene polymorphism in HLA class I splicing, Int. J. Immunogenet. 43 (2) (2016) 65–78.
[71] Y. Xiao, J. Hu, W. Yin, Systematic identification of non-coding RNAs, Adv. Exp. Med. Biol. 1094 (2018) 9–18.
[72] N. Kaur et al., Mapping molecular HLA typing data to UNOS antigen equivalents, Hum. Immunol. 79 (11) (2018) 781–789.
[73] N.P. Mayor et al., Recipients receiving better HLA-matched hematopoietic cell transplantation grafts, uncovered by a novel HLA typing method, have superior survival: a retrospective study, Biol Blood Marrow Transplant 25 (3) (2019) 443–450.
[74] E. Geffard et al., Easy-HLA: a validated web application suite to reveal the full details of HLA typing, Bioinformatics 36 (7) (2020) 2157–2164.
[75] W. Choi et al., HATK: HLA analysis toolkit, Bioinformatics 37 (3) (2021) 416–418.
[76] V. Paunic et al., Charting improvements in US registry HLA typing ambiguity using a typing resolution score, Hum. Immunol. 77 (7) (2016) 542–549.
[77] K. Osoegawa et al., HLA haplotype validator for quality assessments of HLA typing, Hum. Immunol. 77 (3) (2016) 273–282.
[78] K.J. Ingram et al., New HLA alleles discovered by next generation sequencing in routine histocompatibility lab work in a medium-volume laboratory, Hum. Immunol. 80 (7) (2019) 465–467.
[79] C.J. Hernandez-Frederick et al., Identification of 2127 new HLA class I alleles in potential stem cell donors from Germany, the United States and Poland, Tissue Antigens 83 (3) (2014) 184–189.
[80] J.T. den Dunnen et al., Sequence Variant Descriptions: HGVS Nomenclature and Mutalyzer, Curr Protoc Hum Genet 90 (2016), p. 7 13 1-7 13 19.
[81] B. Schone et al., TypeLoader2: Automated submission of novel HLA and killer-cell immunoglobulin-like receptor alleles in full length, HLA 93 (4) (2019) 195–202.
[82] S.J. Mack, A gene feature enumeration approach for describing HLA allele polymorphism, Hum. Immunol. 76 (12) (2015) 975–981.
[83] R.M. Engen et al., Substituting imputation of HLA antigens for high-resolution HLA typing: Evaluation of a multiethnic population and implications for clinical decision making in transplantation, Am. J. Transplant. 21 (1) (2021) 344–352.
[84] Y. D'Souza et al., Inaccuracies in epitope repertoire estimations when using Multi-Locus Allele-Level hla genotype imputation tools, HLA (2018).
[85] D. Allan et al., Reducing ethnic disparity in access to high-quality HLA-matched cord blood units for transplantation: analysis of the Canadian Blood Services' Cord Blood Bank inventory, Transfusion 59 (7) (2019) 2382–2388.
[86] M. Kamoun et al., HLA amino acid polymorphisms and kidney allograft survival, Transplantation 101 (5) (2017) e170–e177.
[87] D. Meyer, K. Nunes, HLA imputation, what is it good for?, Hum. Immunol. 78 (3) (2017) 239–241.
[88] D.J. Pappas et al., Significant variation between SNP-based HLA imputations in diverse populations: the last mile is the hardest, Pharmacogenomics J. 18 (3) (2018) 367–376.
[89] S. Cook et al., Accurate imputation of human leukocyte antigens with CookHLA, Nat. Commun. 12 (1) (2021) 1264.
[90] N. Vince et al., SNP-HLA Reference Consortium (SHLARC): HLA and SNP data sharing for promoting MHC-centric analyses in genomics, Genet. Epidemiol. 44 (7) (2020) 733–740.