Published in final edited form as:

Tissue Antigens. 2013 March ; 81(3): 150–160. doi:10.1111/tan.12072.

A One-Step DNA Sequencing Strategy To HLA Type Hematopoietic Stem Cell Donors at Recruitment —Rethinking Typing Strategies

Bin Tu^a, Nuri Cha^a, Ruyan Yang^a, Jennifer Ng^a, and Carolyn Katovich Hurley^b

^aDepartment of Pediatrics, CW Bill Young Marrow Donor Recruitment and Research Program, Georgetown University, Washington DC USA

^bDepartment of Oncology, CW Bill Young Marrow Donor Recruitment and Research Program, Georgetown University, Washington DC USA

Abstract

In order to reduce the time required to identify a match for unrelated donor hematopoietic stem cell transplantation, a one-step DNA sequencing strategy was employed at the time of recruitment. The impact of this strategy on HLA typing resolution and the effect of current registry requirements on resolution and coding of assignments were evaluated. Sanger-based DNA sequencing was used to obtain diploid exons 2 and 3 HLA-A, -B, -C assignments of 2,747 unrelated African American and 1,822 European American volunteers at recruitment. The results demonstrate the high resolution of the approach and challenge several aspects of the current registry typing strategy. Of the 46% of African American and 74% of European American individuals whose HLA typing resulted in alternative genotypes, the majority (93%) were predicted to have only a single “common” genotype among the alternatives. The common practice of adding secondary assays to resolve alternative genotype assignments that include more than two antigen groups was also evaluated. While the percentage of assignments with greater than two antigen groups reached as high as 21% (HLA-A in European Americans), only 1.8% of individuals at most carried two common genotypes encompassing three antigen groups. The assignment of NMDP-designated allele codes to the one-pass results reduced the resolution substantially and introduced genotypes that were not included in the laboratory’s assignments. We suggest the alternative strategy of using the exons 2–3 diploid nucleotide sequence as the assignment submitted to the registry with the added benefit of immortalizing the assignment in time regardless of the introduction of novel alleles. To keep pace with current donor selection criteria and with the increasing number of new alleles, it is time to rethink our recruitment typing strategies.

Corresponding author: Carolyn Katovich Hurley, Ph.D., E404 Research Building, Georgetown University Medical Center, 3970 Reservoir Road, NW, Washington DC 20057 USA; Fax: 202-687-6440; Phone: 202-687-2157; hurleyc@georgetown.edu.

Conflict of Interest

The authors have no conflicts of interest.

Introduction

Unrelated hematopoietic stem cell donors are selected based on high resolution matching (1) of a minimum of HLA-A, -B, -C, and -DRB1 loci with a searching patient (2). Volunteers joining an unrelated donor registry are usually typed for HLA-A, -B, -DRB1 at recruitment; however, the assignments are usually at intermediate resolution to limit typing costs so that many volunteers can be added to the registry (3). As a consequence, searches identify potentially matched donors who must undergo further HLA typing at a higher resolution and for additional loci (HLA-C, -DQB1, -DPB1) to identify an optimal match for a specific patient. To assist in donor selection, some registries have incorporated sophisticated algorithms into their search routines in order to predict the likelihood of a high resolution match based on allele and haplotype frequencies for each volunteer donor listed on a search report (HapLogic®, OptiMatch®) (4,5). These algorithms have increased the efficiency of the search process but the resolution of typing still remains a barrier to reducing the time from search to transplantation.

One challenge for unrelated donor registries is the complexity of the HLA system. The continuing discovery of novel alleles has resulted in loci with hundreds to thousands of alleles, for example, HLA-B with over 2000 alleles. The sharing of short sequence motifs among alleles at a locus has made it almost impossible to identify an allele by detection of a single sequence motif. For these reasons, the DNA-based typing results obtained at recruitment of registry volunteers usually include many alternative genotypes (often referred to as “genotypic ambiguity” or “ambiguous genotypes” or “typing ambiguity”) . In order to encompass those alternatives in the limited space of a search report, alternative genotypes are collapsed into a single genotype using a letter code to specify alternative alleles (6). This strategy has its limitations; one is that re-expansion of the coded HLA assignment at the transplant center produces additional genotypes that do not exist within the donor typing (7).

For many years, DNA-based HLA typing has focused on the exons encoding the functional region of the HLA molecule, the antigen recognition domain (8,9). This site is defined as the region of the HLA molecule that binds peptides and interacts with antigen receptors on T lymphocytes and with natural killer cell immunoglobulin-like receptors (KIR) and is where most polymorphic residues are found. An added complexity for typing is that more than one pair of alleles share a diploid DNA sequence for these exons. These pairs of alleles differ in the phase of the polymorphisms i.e., which of the alternative polymorphic nucleotides are located on a specific homologue of chromosome 6 (Table 1). As novel alleles are identified, the number of pairs of alleles sharing a diploid sequence increases and new genotypic ambiguities are identified (Figure 1)(10). For example, in 2006, a specific diploid sequence would be assigned as genotype A*01:01:01G + A*01:13 (Table 1). Discovery of novel allele A*01:28 in 2007, together with A*01:17, created a genotype sharing the same diploid sequence as A*01:01:01G + A*01:13 (i.e., now an “ambiguous diploid sequence”). Thus, two alternative genotypes appeared in the IMGT/HLA ambiguous typing list (<http://www.ebi.ac.uk/imgt/hla/ambig.html>) in October 2007 (Table 1) (11). The number of positions that vary between two alleles forming a genotype predicts ultimately how many alternative genotypes will share the same diploid sequence. For example, the alleles A*01:01:01G and A*01:13 differ for three nucleotide polymorphisms (i.e. 282 C/G, 299

C/T, 301 A/G) predicting that the two genotypes currently known may become a maximum of four genotypes in the future. The diploid sequence for A*02:01:01G + A*26:01:01 presents an extreme example where 29 polymorphic residues theoretically could include over 268 million alternative genotypes.

The impact of the increasing number of HLA alleles on HLA typing and HLA assignments has been the subject of much discussion (7,10,12,13). The frequency of alleles and haplotypes varies among world populations (14) and it is possible that many newly described alleles in populations already extensively studied may have very limited distributions (15). One complexity in understanding the frequency of alleles and haplotypes has been our inability to type at and capture high resolution information in a format not impacted by the discovery of new alleles (7,13,16).

While much of DNA-based registry typing is performed by sequence-specific oligonucleotide hybridization, a few registries have been employing DNA sequencing strategies at the time of donor recruitment (17,18). The higher resolution benefits the registry in two ways. First, it allows the registry to understand allele and haplotype frequencies of its volunteer donors, allowing evaluation of the size and composition of the registry and the likelihood of finding matched donors (19,20). Secondly, it provides high resolution typed donors during the search, facilitating rapid donor selection. The speed at which a suitable donor is identified can significantly impact patient survival (21).

One goal of this study was to explore, at the time of recruitment, the benefits of DNA sequencing focused on class I exons 2 and 3 encoding the antigen recognition domain of HLA. In order to reduce the time and effort required for high volume typing, to minimize errors, and to provide a more robust protocol for the capture of primary data, our strategy focused on sequencing the two alleles at a locus simultaneously without separation of the alleles and without any other secondary assays. A second goal for our study was to evaluate the impact of current registry criteria for the definition of “acceptable” HLA assignments on our results. These criteria include the requirement to use letter codes for alternative alleles to condense genotypes and the requirement to use a typing resolution generating only alternative genotypes that encode no more than two antigen families (e.g., A*01+A*02 or A*36,*02 would not be acceptable).

Materials and Methods

HLA-A, -B, and -C loci were sequenced from unrelated individuals volunteering for an unrelated donor hematopoietic stem cell registry: 2,747 who self-identified as African American and 1,822, European American. Locus specific primers, annealing in the introns 5' of exon 2 and 3' of exon 3, were used to amplify both alleles at each locus. Sanger sequencing with a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA) used primers flanking each exon. No further assays were routinely performed to resolve alternative genotypes; HLA “G” groups (22) were also not resolved. Data were analyzed using Assign™ version 3.5.1 (Conexio Genomics, Applecross, Western Australia) and interpreted using IMGT/HLA Database version 2.28.0 (January 2010) (11). Supplemental Table 1 lists the amplification and sequencing primers.

Allele frequencies for European Americans and African Americans were based on Maiers et al. (24) (<http://bioinformatics.nmdp.org/>) supplemented by information on frequencies of alleles differing in the third and fourth fields of their nomenclature (S. J. Mack, personal communication). In a small percent of samples that carried genotypes either not included in frequency tables or with more than one common genotype, additional typing was performed to resolve genotypic ambiguity using additional sequencing primers or group-specific amplification to determine phase (23). This typing was performed solely to clarify the frequency of alternative genotypes in these samples but was not routine in the one-step sequencing strategy.

Results

Only a small fraction (<9%) of the diploid HLA class I sequences interpreted as genotype ambiguities were observed in this study of 2,747 African Americans and 1,822 European Americans (Table 2, Supplemental Table 2). Of the 2,246 ambiguous diploid HLA-A sequences listed in IMGT/HLA Database version 2.28.0, only 5.0% (112 in African Americans) and 4.0% (89 in European Americans) were observed. The typing results for HLA-B and HLA-C were similar; only 2.9% (African Americans) to 3.0% (European Americans) of the HLA-B and 8.8% (African Americans) to 8.1% (European Americans) of the HLA-C ambiguous diploid sequences were found.

As expected based on allele frequencies, some ambiguous sequences were unique to each population although many were shared. For example, African Americans and European Americans shared 69 HLA-A ambiguous diploid sequences (e.g., a sequence with one interpretation being A*02:01:01G+A*03:01:01G), 43 were found only in African Americans (e.g., A*23:01:01G+A*68:02:01G), and 20 only in European Americans (e.g., A*01:01:01G+A*25:01:01G). At the HLA-B locus, African Americans and European Americans shared 123 ambiguous diploid sequences (e.g., B*07:02:01G+B*35:01:01G), 116 were found only in African Americans (e.g., B*35:01:01G+B*53:01:01G), and 121 only in European Americans (B*15:01:01G+B*44:02:01G). At the HLA-C locus, African Americans and European Americans shared 77 (e.g., C*04:01:01G+C*07:01:01), 28 were found only in African Americans (e.g., C*02:05+C*16:07), and 20 only in European Americans (C*01:02:01G+C*03:04:01G). However, despite the diversity in ambiguous diploid sequences observed, they represent only a small fraction of potential genotypic ambiguity.

In order to evaluate the resolution of typing, the number of alternative genotypes assigned to each individual by the one-step DNA sequencing was evaluated. Table 3 shows the resolution for both African American and European American populations for each locus. The HLA assignments of the majority of samples were narrowed down to three or less genotypes using IMGT/HLA Database 2.28. For example, 92.9% of the 2,732 African Americans typed for HLA-A carried one (55.5%), two (24.3%) or three (13.1%) alternative genotypes; 87.2% carried 3 HLA-B genotypes and 88.8% carried 3 HLA-C genotypes. The percentages for all three loci (~73%/locus) were lower for European Americans. The maximum number of HLA-A genotypes associated with a single diploid sequence was 11 (e.g., A*02:01:01G+A*24:02:01G) and was observed in 18 African Americans (0.7%) and

91 European Americans (5.0%). The maximum for HLA-B was 12 alternative genotypes (B*15:01:01G+B*51:01:01G) observed in 11 European Americans (0.6%) and one African American (<0.1%) and for HLA-C was 7 alternative genotypes (C*03:03:01G+C*07:01:01G) observed in 10 African Americans (0.4%) and 41 European Americans (2.3%). While use of later IMGT/HLA Databases for interpretation of sequence results will increase the number of alternative genotypes in many individuals, overall the one-step typing produces results with a limited number of genotypes.

For samples with alternative genotypes, knowledge of population-specific allele frequencies allows the prediction of genotypes most likely to be present when higher resolution testing is performed. For both populations, the likelihood of each alternative genotype for a single diploid sequence was evaluated based on a definition as “common” as any allele observed in a study of 3,136 African American or of 12,770 European Americans in the U.S. registry (24) (http://bioinformatics.nmdp.org/HLA/Haplotype_Frequencies). For African Americans, the majority of the ambiguous diploid sequences observed in our study (96.4% HLA-A (i.e., 108 of the 112 sequences), 93.3% HLA-B, 97.1% HLA-C) would likely be resolved in favor of a single genotype in this population (i.e., one genotype carried two alleles common in African Americans based on Maier et al. while the other genotypes carried only one or no common alleles) (Table 4). Similar results were obtained for European Americans: the majority of the ambiguous diploid sequences (95.5% HLA-A (i.e., 85/89), 93.8% HLA-B, 92.8% HLA-C) would be likely to be resolved in favor of a single genotype in this population. In summary, for both populations, although many recruitment typings included 2 or more genotypes, information on allele frequencies for that population predicts a specific genotype for the majority (>93%) of individuals.

A small number of diploid sequences (i.e., two HLA-A, six HLA-B, one HLA-C sequences for African Americans and six HLA-B and two HLA-C for European Americans) did not include any “common” alleles and these samples were typed at a higher resolution to update the “common” allele list (Table 5). All but one were observed in single individuals; three individuals carrying the same uncommon ambiguity were typed as B*35:01:01G+B*39:05:01 and not B*35:06+B*39:07.

Table 6 includes a list of those diploid sequences that included more than one common genotype. The percentage of African American individuals in this category was 0.4% for HLA-A (10/2732), 2.3%, HLA-B, and 1.0%, HLA-C (Table 4). For European Americans, the percent of individuals was 2.1% HLA-A (39/1822 total), 1.9% HLA-B and 5.2% HLA-C. Table 6 also provides frequency estimates to predict the likelihoods of the “common” alternatives. For example, A*02:01:01G+A*31:01:02G is almost 9×10^5 times more common than its alternative, A*02:20:01+A*31:02 in European Americans. In contrast, the diploid sequence specifying B*35:01:01G+B*49:01:01 or B*50:01:01+B*53:01:01 has two equally likely alternatives in African Americans. Of the 29 diploid sequences listed in Table 6, only 8 have alternative genotypes with frequency ratios of at most 32 to 1. Three diploid sequences were observed in both African American and European American populations and, as expected, the likelihood of the alternative genotypes varies with the population. For example, the two alternative genotypes, B*35:01:01G+B*49:01:01 versus B*50:01:01+B*53:01:01, exhibit a ratio of 1:1 in African Americans but 29:1 in European

Americans. Overall, however, for this category of ambiguities, frequency estimates predict that the majority will carry a single specific genotype out of two or more alternatives.

Higher resolution testing was used to determine the genotypes present for all diploid sequences in Table 6 that showed a ratio of <1000:1. Sixteen diploid sequences were evaluated and all individuals carrying those sequences were tested. Three sequences had two genotypes with ratio of 1:1. Of the 12 African American individuals sharing one of these diploid sequences, eight carried B*35:01:01G+B*49:01:01 and four, B*50:01:01+B*53:01:01. The same diploid sequence has a ratio of 29:1 in European Americans and was resolved in five individuals to B*35:01:01G+B*49:01:01 and in one to B*50:01:01+B*53:01:01. The other two diploid sequences with 1:1 ratios were resolved to only one genotype: four individuals to B*39:06:01+B*44:03:02 (European American) and one to B*44:02:01G+B*51:09:01 (European American). One diploid sequence with a genotype ratio of 225:1 was carried by 19 individuals, 18 with the most frequent genotype C*05:01:01G+C*16:01:01 and one with the alternative, C*08:02:01+C*16:02:01. The typing of the remainder of the diploid sequences examined (11/16) identified the more frequent genotype. Overall, the higher resolution analysis supported the predicted frequency data.

While most assignments using the one-step sequencing strategy included only alleles that shared the first digits of their name suggesting that they encoded at most two antigen groups, a percentage of the ambiguous diploid sequences included genotypes specifying a third or fourth antigen group. In African Americans, 12.5% (14/112) of HLA-A diploid sequences included >2 antigens (e.g., A*02:01:01G+A*32:01:01G or A*02:01:01G+A*74:06); 23.4% HLA-B, and 12.3% HLA-C impacting 3.8% (105/2732) HLA-A, 10.3% HLA-B, and 4.4% HLA-C of the total individuals typed (Table 4). In European Americans, for HLA-A, 16 of the 89 (18.0%) ambiguous diploid sequences included genotypes that carried three antigen groups. These 16 genotypes were found in 21.0% of the individuals tested (383/1822). Lower percentages of typed individuals were impacted for HLA-B (14.3%) and HLA-C (6.9%). Thus, up to 21% of individuals typed included three or more antigen groups which may present problems for registries with algorithms that sort potential donors based on antigen assignments.

When the alternative genotypes bearing a third or fourth antigen were evaluated for their frequencies, many appeared unlikely to occur. For African Americans, only five HLA-B and two HLA-C diploid sequences were assigned to two “common” genotypes that included more than two antigens (Tables 4,6). For European Americans, four HLA-B and three HLA-C diploid sequences had these characteristics (24). In general, at most, 5% of individuals (African Americans—0% HLA-A, 1.8% HLA-B, 1.0% HLA-C; European Americans—0% HLA-A, 0.8% HLA-B, 5.0% HLA-C) carried alternative genotypes that “broke” the less than three antigen rule and that had a frequency that would make them “common” (Table 6). These assignments were more frequent in European Americans.

Since a list of alternative genotypes can not currently be reported to a registry, each result obtained from the one-step sequencing was condensed into a single genotype using letter codes to specify alternative alleles for submission. This condensation lost considerable

information obtained through the one-step typing. Table 7 demonstrates the loss of information and the introduction of new genotypes that are not present in the sample.

Discussion

In this study, we compared the typing results from two populations, African Americans and European Americans, using a one step DNA sequencing strategy focused on the exons encoding the antigen recognition domain. While the majority of volunteers for the unrelated donor registry have European ancestry, we chose to also type African Americans because their extensive HLA diversity (24) is not adequately captured by lower resolution typing. The extensive diversity would also challenge our typing strategy and provide a wider range of HLA genotypes to evaluate.

Recruitment typing is a screening-level typing meant to narrow down the HLA possibilities to facilitate the search for an allele-matched donor. It is not necessary to identify the single genotype carried by a volunteer but it is important to narrow down the possibilities to facilitate rapid donor selection. Volunteers selected as a potential match are typed at a higher resolution and that typing may include other HLA loci like DPB1 known to impact survival (25). Our results demonstrate that the one-step sequencing approach to obtain a diploid sequence for exons 2 and 3 at the time of initial HLA typing of newly recruited volunteers for a registry provides very high resolution results. The number of potential genotypes for >87% of African Americans and 73% of European Americans is three or less. The majority of samples are predicted to have a single common HLA genotype for each of the three HLA class I loci facilitating the selection of volunteers for higher resolution typing. Of course, each donor is selected based on the potential to match at several HLA loci. Therefore, the presence of a limited number of alternative genotypes at each locus in the context of matching at all key loci will limit the number of donors for consideration. A recent informatics analysis also supports our observation that a single step DNA sequencing strategy is a powerful tool for recruitment typing and provides the least ambiguity compared to other registry typing methods such as oligonucleotide typing and serology (26).

It should be noted that, for newly recruited volunteers with alternative genotypes, we do not advocate listing only single predicted genotypes in the registry. It is critical that the typing result reflects the ambiguity in the assignment. As the HLA diversity within the population increases due to admixture, we are likely to see more and more patients searching with uncommon genotypes. For this reason, it is critical that the HLA assignments be provided in a format that will serve these patients as well as those with common genotypes.

Our analysis was based on an IMGT/HLA Database from January 2010. We chose this database since it allowed us to combine results obtained over the time span of our project. The increase in the length of the genotype lists and the appearance of new ambiguous genotypes since that time are likely to not impact our general observations since newly added genotypes are unlikely to occur in these populations. For example, in IMGT/HLA Database version 2.28.0 (January 2010), there were six other genotypes with the same diploid sequence as A*02:01:01G+A*03:01:01G; all six uncommon. In IMGT/HLA Database version 3.7.0 (January 2012), there were four more genotypes added to the list:

A*02:01:52+A*03:01:03, A*02:35:01+A*03:108, A*02:86+A*03:123, and A*02:237+A*03:05. Alleles found in these genotypes have not been detected in the reference populations used as comparison (24). The potential to free ourselves from the continued increase in ambiguity over time by focusing on the primary sequence data is a solution described below.

We focused on the discrimination of alternative genotypes that include greater than two antigen groups because many registries use antigen level assignments to perform an initial sort of the donor pool during a search (27). Thus, registries often require resolution of these assignments regardless of their frequencies. The advantages of eliminating secondary assays routinely used to obtain results limited to less than two antigens include the reduced cost of reagents, reduced technical effort, elimination of errors introduced by additional sample handling, and the elimination of less robust secondary assays. Since the primers and strategies (i.e., addition of another sequencing primer or a second PCR reaction followed by DNA sequencing) to obtain a hemizygous sequence differ depending on the genotypes that must be resolved, elimination of secondary assays eliminates the need to review sequence assignments and determine the type of assay needed. Thus, there are cost, accuracy and efficiency arguments in favor of a one-step in comparison to a multi-step DNA sequencing strategy. It remains, however, to be determined whether the elimination of secondary assays at recruitment will increase the number of extended typings required to identify a matching donor, adding to the cost.

Even more important, however, is the impact of secondary assays on the robust nature of the HLA assignment. One limitation is that primary data and test reagents used in secondary assays are not readily incorporated into the initial result and are not captured by the registry. This is particularly true if the secondary assay uses a different testing technology than the initial assay (e.g., DNA sequencing followed by sequence specific priming). In these cases, laboratory software is unlikely to capture and merge primary data from both results making it difficult for the registry to collect this information. A second limitation is that the reagents used are selected based on the current alternative genotypes and do not take into account new alternatives that will appear over time. For example, in 2000, a diploid sequence might have three alternative genotypes, B*07:02:01G+B*35:01:01G or B*07:09+B*35:34 or B*07:18:01+B*35:05, and a strategy for resolution might involve the addition of a sequencing primer in exon 3 to capture the phase of polymorphisms. Thus, the assignment might be narrowed down to a single genotype B*07:02:01G+B*35:01:01G. However, 12 years later, that same diploid sequence now has 14 alternative genotypes; some differing in the phase of polymorphisms in exon 2. Thus, the single genotype assignment of the volunteer now does not capture the information that these newer genotypes have not been eliminated. Thus, searches for individuals with these newer genotypes will not identify the donor typed in 2000 as a potential match. Thus, the use of secondary assays to resolve alternative genotypes at the time of recruitment should be reconsidered.

One advantage to the next generation sequencing methods is the ability to isolate individual alleles for sequencing which will greatly facilitate high resolution typing. Drawbacks to use of this technology currently are lack of robust methods, analysis of complex and extensive data, and the inability to obtain overlapping “reads” especially linking class I exons 2 and 3.

Challenges remain in the application of DNA sequencing to registry typing in terms of cost compared to other typing methods (e.g. sequence specific oligonucleotide probe typing) and high throughput. Next generation sequencing strategies have the potential to reduce the cost but these new technologies are still in the early stages of addressing HLA data analysis and accuracy (28).

While we have focused on class I, the same strategy can be applied to class II loci; however, for DRB1, the choice of locus-specific amplification primers chosen to include the entire exon 2 sequence will be more of an issue because of the homology to DRB3/4/5 loci. While typing data that does not completely characterize antigen recognition domain exons can be used in cases where primer hybridization sites prevents complete sequence information from being obtained, the inclusion of partial exon 2–3 exon sequence data prevents the use of the IMGT/HLA Database of ambiguous typings (<http://www.ebi.ac.uk/imgt/hla/ambig.html>) as a reference for interpretation of results and to standardize the primary typing information received by the registry.

One barrier to use of the one-step typing strategy is the inability of the laboratory to report alternative genotypes to a registry at present. Whether typing by DNA sequencing or oligonucleotide hybridization, the common practice is to collapse alternative genotypes arising from a typing into a single genotype using letter codes to designate alternative alleles when submitting the data to a registry. The impact of consolidating typings in this way is that resolution of the typing result is lost when the genotype pair information is not retained. Thus an assignment with two genotypes will be interpreted as four genotypes, two of them artificial, when the typing is re-expanded by the end user. The situation becomes even worse as the number of alternative genotypes grows. For example, with IMGT/HLA Database 2.28.0, the diploid sequence B*15:01G+B*51:01:01G has 12 alternative genotypes associated with it. Collapsing this commonly observed assignment into a single genotype would result in 144 (12^2) genotypes upon expansion, 129 of them that do not exist in the typed individual.

The use of allele codes to condense genotypes is also being challenged by the enormous and increasing number of new alleles. Besides the loss of resolution and the creation of genotypes that do not exist in the individual described above, further loss of information occurs from the routine procedure of truncating allele names that correspond to a specific code. Registries may shorten the names of alleles submitted by the typing laboratory to include only those digits that specify the polypeptide sequence (i.e., fields 1 and 2)(see legend for Table 7 for further explanation). This truncation loses the link to a specific diploid sequence and could introduce additional artifactual assignments. Inconsistencies are also introduced in assignments based on how genotypes are condensed into codes, specifically when the two class I alleles in a genotype share the first digits of their name. Thus, the same individual could be assigned different allele code assignments depending on how the alleles in the alternative genotypes are clustered. Furthermore, the continued effort dedicated to assigning new codes to the same typing result as the interpretation changes over time has produced an unwieldy system. Further, the relationship among the assignments given to the same diploid sequence is not readily understood by the end user, i.e., the

physician reviewing the search results. So, while useful, it is becoming clear that the condensation of genotypes loses information and should be phased out.

The benefits of the one-step typing strategy is best realized if the diploid sequences obtained are not interpreted into HLA assignments by the laboratory. Because interpretation changes as new alleles are described, maintaining HLA assignments long term in a registry would be best served by storing the primary sequence data (7,13). We propose a novel strategy for assigning an HLA “type” to our one-pass assignments. While the diploid exon 2-exon3 sequence itself serves as a unique identifier, it would make reporting easier to use a random identifier (e.g., a QR code) to designate the sequence. This identifier can be linked to the IMGT/HLA Database of ambiguous typings which will provide a past and current list of all genotypes that create the same diploid sequence; thus, the identifier need not be updated over time. Providing an assignment to every diploid sequence, regardless of whether it currently is interpreted as a single genotype or results in alternative genotypes will ensure that the assignment will not change as new alleles are identified. This strategy of assigning a random identifier will provide a robust and stable assignment that will not change over time and that will accommodate the continuing influx of new alleles.

The final challenge is to present this assignment on a search report so that it is understandable to the end user. One alternative is to simply list the donor as a potential match with an estimated likelihood of allele matching and not present the HLA assignment on the search report at all. Alternatively, providing a link to a list of potential genotypes with predicted frequencies would be another means to present the alternative genotype list.

In conclusion, our study has evaluated a powerful strategy for the typing of newly recruited volunteers for a registry that can produce an HLA assignment that is not impacted by the proliferation of new alleles. Despite this, this strategy would not be acceptable to some registries because of the presence of greater than two antigen groups in the assignments. Furthermore, the current requirement to assign allele codes to these high resolution data will dramatically reduce their resolution and quality of the typing. It's time to rethink registry typing and data submission strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by funding from the Office of Naval Research N00014-08-1-1078 and N00014-10-1-0199. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the U.S. government. We would like to thank the NMDP Bioinformatics Research Ginger Group for helpful discussions.

References

1. Nunes E, Heslop H, Fernandez-Vina M, Taves C, Wagenknecht DR, Eisenbrey AB, Fischer G, Poulton K, Wacker K, Hurley CK, Noreen H, Sacchi N. Definitions of histocompatibility typing terms. *Blood*. 2011; 118:e180–e183. [PubMed: 22001389]

2. Spellman SR, Eapen M, Logan BR, Mueller C, Rubinstein P, Setterholm MI, Woolfrey AE, Horowitz MM, Confer DL, Hurley CK. A perspective on the selection of unrelated donors and cord blood units for transplantation. *Blood*. 2012; 120:259–265. [PubMed: 22596257]
3. Spellman S, Setterholm M, Maiers M, Noreen H, Oudshoorn M, Fernandez-Vina M, Petersdorf E, Bray R, Hartzman RJ, Ng J, Hurley CK. Advances in the selection of HLA-compatible donors: refinements in HLA typing and matching over the first 20 years of the National Marrow Donor Program Registry. *Biol Blood Marrow Transplant*. 2008; 14:37–44. [PubMed: 18721779]
4. Hurley CK, Maiers M, Marsh SG, Oudshoorn M. Overview of registries, HLA typing and diversity, and search algorithms. *Tissue Antigens*. 2007; 69(Suppl 1):3–5. [PubMed: 17445151]
5. Eberhard HP, Feldmann U, Bochtler W, Baier D, Rutt C, Schmidt AH, Muller CR. Estimating unbiased haplotype frequencies from stem cell donor samples typed at heterogeneous resolutions: a practical study based on over 1 million German donors. *Tissue Antigens*. 2010; 76:352–361. [PubMed: 20604895]
6. Bochtler W, Maiers M, Oudshoorn M, Marsh SG, Raffoux C, Mueller C, Hurley CK. World Marrow Donor Association guidelines for use of HLA nomenclature and its validation in the data exchange among hematopoietic stem cell donor registries and cord blood banks. *Bone Marrow Transplant*. 2007; 39:737–741. [PubMed: 17438587]
7. Maiers M, Hurley CK, Perlee L, Fernandez-Vina M, Baisch J, Cook D, Fraser P, Heine U, Hsu S, Leffell MS, Mauer D, Noreen H, Tang T, Trucco M, Yang SY, Hartzman RJ, Setterholm M, Winden T, Shepherd D, Hegland J. Maintaining updated DNA-based HLA assignments in the National Marrow Donor Program Bone Marrow Registry. *Rev Immunogenet*. 2000; 2:449–460. [PubMed: 12361088]
8. Xiao Y, Lazaro AM, Masaberg C, Haagenson M, Vierra-Green C, Spellman S, Dakshanamurthy S, Ng J, Hurley CK. Evaluating the potential impact of mismatches outside the antigen recognition site in unrelated hematopoietic stem cell transplantation: HLA-DRB1*1454 and DRB1*140101. *Tissue Antigens*. 2009; 73:595–598. [PubMed: 19392807]
9. Oudshoorn M, Lie JL, Roelen DL, Verduyn W, Claas FHJ. Is an HLA-DRB1*1454/HLA-DRB1*1401 mismatch an acceptable mismatch in hematopoietic stem cell transplantation? *Tissue Antigens*. 2008; 71:315–316.
10. Voorter CE, Mulkers E, Liebelt P, Sleyster E, van den Berg-Loonen EM. Reanalysis of sequence-based HLA-A, -B and -Cw typings: how ambiguous is today's SBT typing tomorrow. *Tissue Antigens*. 2007; 70:383–389. [PubMed: 17868258]
11. Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. *Nucleic Acids Res*. 2009; 37:D1013–D1017. [PubMed: 18838392]
12. Cano P, Klitz W, Mack SJ, Maiers M, Marsh SG, Noreen H, Reed EF, Senitzer D, Setterholm M, Smith A, Fernandez-Vina M. Common and well-documented HLA alleles: report of the Ad-Hoc committee of the american society for histocompatibility and immunogenetics. *Hum Immunol*. 2007; 68:392–417. [PubMed: 17462507]
13. Hollenbach JA, Mack SJ, Gourraud PA, Single RM, Maiers M, Middleton D, Thomson G, Marsh SG, Varney MD. A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement. *Tissue Antigens*. 2011; 78:333–344. [PubMed: 21988720]
14. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol*. 2008; 69:443–464. [PubMed: 18638659]
15. Middleton D, Gonzalez F, Fernandez-Vina M, Tiercy JM, Marsh SG, Aubrey M, Bicalho MG, Canossi A, Carter V, Cate S, Guerini FR, Loiseau P, Martinetti M, Moraes ME, Morales V, Perasaari J, Setterholm M, Sprague M, Tavoularis S, Torres M, Vidal S, Witt C, Wohlwend G, Yang KL. A bioinformatics approach to ascertaining the rarity of HLA alleles. *Tissue Antigens*. 2009; 74:480–485. [PubMed: 19793314]
16. Helmberg W, Hegland J, Hurley CK, Maiers M, Marsh SG, Muller C, Rozemuller EH. Going back to the roots: effective utilisation of HLA typing information for bone marrow registries requires full knowledge of the DNA sequences of the oligonucleotide reagents used in the testing. *Tissue Antigens*. 2000; 56:99–102. [PubMed: 10958364]

17. Schmidt AH, Solloch UV, Pingel J, Baier D, Bohme I, Dubicka K, Schumacher S, Rutt C, Skotnicki AB, Wachowiak J, Ehninger G. High-resolution human leukocyte antigen allele and haplotype frequencies of the Polish population based on 20,653 stem cell donors. *Hum Immunol.* 2011; 72:558–565. [PubMed: 21513754]
18. Schmidt AH, Baier D, Solloch UV, Stahr A, Cereb N, Wassmuth R, Ehninger G, Rutt C. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Hum Immunol.* 2009; 70:895–902. [PubMed: 19683023]
19. Muller CR, Ehninger G, Goldman SF. Gene and haplotype frequencies for the loci HLA-A, HLA-B, and HLA-DR based on over 13,000 German blood donors. *Human Immunology.* 2003; 64:137–151. [PubMed: 12507825]
20. Kollman C, Maiers M, Gragert L, Muller C, Setterholm M, Oudshoorn M, Hurley CK. Estimation of HLA-A, -B, -DRB1 haplotype frequencies using mixed resolution data from a National Registry with selective retyping of volunteers. *Hum Immunol.* 2007; 68:950–958. [PubMed: 18191722]
21. Lee SJ, Klein J, Haagensohn M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood.* 2007; 110:4576–4583. [PubMed: 17785583]
22. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernandez-Vina M, Geraghty DE, Holdsworth R, Hurley CK, Lau M, Lee KW, Mach B, Maiers M, Mayr WR, Muller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, Svejgaard A, Terasaki PI, Tiercy JM, Trowsdale J. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens.* 2010; 75:291–455. [PubMed: 20356336]
23. Tu B, Mack SJ, Lazaro A, Lancaster A, Thomson G, Cao K, Chen M, Ling G, Hartzman R, Ng J, Hurley CK. HLA-A, -B, -C, -DRB1 allele and haplotype frequencies in an African American population. *Tissue Antigens.* 2007; 69:73–85. [PubMed: 17212710]
24. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol.* 2007; 68:779–788. [PubMed: 17869653]
25. Fleischhauer K, Shaw BE, Gooley T, Malkki M, Bardy P, Bignon JD, Dubois V, Horowitz MM, Madrigal JA, Morishima Y, Oudshoorn M, Ringden O, Spellman S, Velardi A, Zino E, Petersdorf EW. Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study. *Lancet Oncol.* 2012
26. Paunic V, Gragert L, Madbouly A, Freeman J, Maiers M. Measuring ambiguity in HLA typing methods. *PLoS One.* 2012; 7:e43585. [PubMed: 22952712]
27. Bochtler W, Maiers M, Bakker JN, Oudshoorn M, Marsh SG, Baier D, Hurley CK, Muller CR. World Marrow Donor Association framework for the implementation of HLA matching programs in hematopoietic stem cell donor registries and cord blood banks. *Bone Marrow Transplant.* 2010
28. Holcomb CL, Hoglund B, Anderson MW, Blake LA, Bohme I, Egholm M, Ferriola D, Gabriel C, Gelber SE, Goodridge D, Hawbecker S, Klein R, Ladner M, Lind C, Monos D, Pando MJ, Proll J, Sayer DC, Schmitz-Agheguian G, Simen BB, Thiele B, Trachtenberg EA, Tyan DB, Wassmuth R, White S, Erlich HA. A multi-site study using high-resolution HLA genotyping by next generation sequencing. *Tissue Antigens.* 2011; 77:206–217. [PubMed: 21299525]
29. Cao K, Chopek M, Fernandez-Vina MA. High and intermediate resolution DNA typing systems for Class I HLA-A, -B, -C genes by hybridization with sequence specific oligonucleotide probes (SSOP). *Rev Immunogenetics.* 1999; 1:177–208. [PubMed: 11253946]
30. Cereb N, Maye P, Lee S, Kong Y, Yang SY. Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens.* 1995; 45:1–11. [PubMed: 7725305]
31. Kotsch K, Wehling J, Kohler S, Blasczyk R. Sequencing of HLA class I genes based on the conserved diversity of the noncoding regions: sequencing-based typing of the HLA-A gene. *Tissue Antigens.* 1997; 50:178–191. [PubMed: 9271828]
32. Cereb N, Yang SY. Dimorphic primers derived from intron 1 for use in the molecular typing of HLA-B alleles. *Tissue Antigens.* 1997; 50:74–76. [PubMed: 9243761]

33. Voorter CE, Swelsen WT, van den Berg-Loonen EM. Intron sequences of HLA-B*73. *Tissue Antigens*. 2001; 57:463–468. [PubMed: 11556971]

HLA-A Alternative Genotypes Over Time

IMGT/HLA first release from yrs 2009-2012

Figure 1.

The number of ambiguous genotypes increases over time. An ambiguous genotype is defined as a typing result that is interpreted as alternative genotypes. The number of unique exon 2–3 diploid sequences that can be interpreted as two or more genotypes and the number of alternative genotypes possessing the same diploid sequence is shown for the January releases of the IMGT/HLA Database from 2009 through 2012.

Table 1

Genotype interpretation of a diploid sequence based on the phase of the polymorphic nucleotides^a.

DNA Sequence ^b : CCCASTCACAGACTGACCGAGYGRACCT		
Chrom 6_a	Chrom 6_b	Genotype
C—C—A	G—T—G	A*01:01:01G+A*01:13
G—C—A	C—T—G	A*01:17+A*01:28
C—T—A	G—C—G	Potential future ^c
C—C—G	G—T—A	Potential future ^c

^aPhase is defined as the paternal or maternal chromosomal location of polymorphic nucleotides.

^bThe alternative genotypes in this example share the nucleotide sequence of exons 2 and 3 except at nucleotide positions 282 (C vs. G, S), 299 (C vs. T, Y) and 301 (A vs. G, R).

^cIn the future, the known genotypes may be joined by two others with a different phasing of the three polymorphic residues.

Table 2

Frequencies of diploid sequences yielding alternative genotypes.

	African Americans	European Americans
HLA-A		
Total individuals typed	2732	1822
Individuals with single genotypes	1515 (55.5%)	517 (28.4%)
Individuals with alternative genotypes	1217 (44.5%)	1305 (71.6%)
HLA-A unique diploid sequences with 2 or more genotypes in IMGT/HLA v2.28.0	2246	2246
No. unique diploid sequences with 2 or more genotypes observed	112 (5.0%)	89 (4.0%)
HLA-B		
Total individuals typed	2747	1805
Individuals with single genotypes	1531 (55.7%)	473 (26.3%)
Individuals with alternative genotypes	1216 (44.3%)	1328 (73.7%)
HLA-B unique diploid sequences with 2 or more genotypes in IMGT/HLA v2.28.0	8108	8108
No. unique diploid sequences with 2 or more genotypes observed	239 (2.9%)	243 (3.0%)
HLA-C		
Total individuals typed	2726	1813
Individuals with single genotypes	1365 (50.1%)	444 (24.5%)
Individuals with alternative genotypes	1361 (49.9%)	1368 (75.5%)
HLA-C unique diploid sequences with 2 or more genotypes in IMGT/HLA v2.28.0	1193	1193
No. unique diploid sequences with 2 or more genotypes observed	105 (8.8%)	97 (8.1%)

Table 3

Number and percentage of individuals with one or more genotypes

	Population	Number of genotypes					
		1	2	3	4	5	>5 ^a
HLA-A	African American (n=2732)	1515 55.5%	663 24.3%	358 13.1%	84 3.1%	14 0.5%	98 3.6%
	European American (n=1822)	517 28.4%	253 13.9%	213 11.7%	330 18.1%	90 4.9%	419 23.0%
HLA-B	African American (n=2747)	1531 55.7%	617 22.5%	248 9.0%	133 4.8%	49 1.8%	169 6.1%
	European American (n=1801)	473 26.3%	348 19.3%	275 15.3%	118 6.6%	220 12.2%	367 20.4%
HLA-C	African American (n=2726)	1365 50.1%	636 23.3%	419 15.4%	187 6.9%	102 3.7%	17 0.6%
	European American (n=1812)	444 24.5%	503 27.8%	343 18.9%	196 10.8%	221 12.2%	105 5.8%

^a Maximum number of genotypes is 11 for HLA-A, 12 HLA-B, 7 HLA-C based on IMGT/HLA Database 2.28.

Table 4

Characteristics of the diploid sequences interpreted as two or more genotypes.

Categories	African Americans		European Americans	
	Unique Diploid Sequences	Individuals	Unique Diploid Sequences	Individuals
Total interpreted as more than one genotype	A: 112 B: 239 C: 105	A: 1217 B: 1216 C: 1361	A: 89 B: 243 C: 97	A: 1305 B: 1328 C: 1368
• With no “common” genotypes	A: 2 B: 6 C: 1	A: 2 B: 8 C: 1	A: 0 B: 6 C: 2	A: 0 B: 6 C: 2
• With a single “common” genotype	A: 108 B: 223 C: 102	A: 1205 B: 1144 C: 1332	A: 85 B: 228 C: 90	A: 1266 B: 1288 C: 1271
• With two or more “common” genotypes	A: 2 B: 10 C: 2	A: 10 B: 64 C: 28	A: 4 B: 9 C: 5	A: 39 B: 34 C: 95
• Including >2 antigen groups	A: 14 B: 56 C: 13	A: 105 B: 283 C: 119	A: 16 B: 58 C: 8	A: 383 B: 258 C: 126
◦ With a single common genotype	A: 14 B: 51 C: 11	A: 105 B: 233 C: 91	A: 16 B: 54 C: 5	A: 383 B: 244 C: 35
◦ With two common genotypes encompassing >2 antigen groups	A: 0 B: 5 C: 2	A: 0 B: 50 C: 28	A: 0 B: 4 C: 3	A: 0 B: 14 C: 91

Table 5

Diploid sequence with no “common” genotypes based on Maiers et al. (2007).

Population^a	Genotype Observed^b	Alternative Genotype	Alternative Genotype
AfAm	A*02:01:01G+A*68:03:01	A*02:01:02+A*68:03:02	A*02:90+A*68:05
AfAm	A*03:01:02+A*68:02:01	A*03:01:01G+A*68:28	
EurAm	B*07:02:01G+B*35:43:01G	B*07:24+B*35:86	B*07:55+B*35:99
EurAm	B*13:02:01G+B*15:39	B*13:16+B*15:05:01	
AfAm	B*15:01:01G+B*39:05:01	B*15:58+B*39:07	
AfAm	B*15:09+B*53:01:01	B*15:23+B*35:04:02	
AfAm	B*35:01:01G+B*39:05:01	B*35:06+B*39:07	
AfAm	B*35:01:01G+B*40:04	B*35:28+B*40:68	
EurAm	B*35:01:01G+B*56:10	B*35:93+B*56:02	
EurAm	B*35:06+B*51:01:01G	B*35:04:01+B*51:13:02	B*35:04:02+B*51:13:01
EurAm	B*35:43:01G+B*40:02:01G	B*35:44+B*40:03	
EurAm	B*40:02:01G+B*44:03:02	B*40:02:09+B*44:03:01G	B*40:64+B*44:54
AfAm	C*03:04:06+C*04:01:01G	C*03:04:02+C*04:01:02	
EurAm	C*03:05+C*04:01:01G	C*03:27+C*04:10	
EurAm	C*03:05+C*07:02:01G	C*03:27+C*07:39	

^a Observed in single individuals with the exception of B*35:01:01G+B*39:05:01 (3 individuals).

^b Additional testing was used to resolve the alternative genotypes. The genotype present is listed in this column.

Table 6

Alternatives with more than one common genotype observed and the frequency^a of the two alleles included in each genotype

Observed in Population/ No. Individuals	Genotype 1 (G1)	Frequency	Genotype 2 (G2)	Frequency	Ratio of Most Common to Least Common Genotype ^b	Genotype When Higher Resolution Testing Performed/No. Tested
Genotypes with 2 antigen assignments^c						
AFAm/8	A*02:01:01G+A*02:05:01G ^d	0.004664275	A*02:02+A*02:06:01G	1.76442E-05	264	G1/8
EurAm/10	A*02:01:01G+A*02:05:01G	0.004742561	A*02:02+A*02:06:01G	3.2702E-06	1450	NT ^e
EurAm/23	A*02:01:01G+A*31:01:02G	0.013919801	A*02:20:01+A*31:02	1.56E-08	892294	NT
AFAm/2	A*24:02:01G+A*66:02	0.00034839	A*24:07+A*66:03	3.8014E-06	91	G1/2
EurAm/5	A*24:02:01G+A*68:01:02G	0.004348212	A*24:07+A*68:03:01	7.2E-09	603918	NT
EurAm/1	A*25:01:01+A*68:01:02G	0.000963655	A*25:02+A*68:03:01	7.2E-09	133840	NT
AFAm/1	B*07:02:01G+B*15:18:01	0.000151902	B*07:09+B*15:10:01	1.27218E-05	11	G1/1
AFAm/4	B*08:01:01G+B*15:01:01G	0.00074841	B*08:12:01+B*15:07	2.604E-07	2874	NT
AFAm/1	B*14:02:01+B*35:08:01	1.80318E-05	B*14:03+B*35:01:01G	0.000323401	17	G2/1
AFAm/3	B*15:03:01G+B*78:01	0.001658672	B*15:61+B*78:02:02	8.82E-08	18805	NT
AFAm/5	B*15:10:01+B*51:01:01G	0.001319432	B*15:37+B*51:02:01	8.528E-07	1547	NT
EurAm/3	B*18:01:01G+B*39:01:01G	0.000880572	B*18:03+B*39:05:01	1.482E-07	5941	NT
EurAm/11	B*27:05:02G+B*51:01:01G	0.002957235	B*27:10+B*51:09:01	5.28E-08	56008	NT
EurAm/1	B*39:03+B*40:06:01	0.00000003	B*39:06:02+B*40:02:01G	9.94964E-05	3316	NT
EurAm/4	B*39:06:01+B*44:03:02	0.000498285	B*39:06:02+B*44:03:01G	0.000498285	1	G2/4
EurAm/1	B*44:02:01G+B*51:09:01	7.92968E-05	B*44:03:01G+B*51:08	5.65782E-05	1	G1/1
EurAm/2	C*03:04:01G+C*16:02:01	0.00042718	C*03:07+C*16:01:01	4.2468E-06	100	G1/2
EurAm/2	C*07:01:01G+C*16:02:01	0.000866216	C*07:09+C*16:01:01	4.2468E-06	203	G1/2
Genotypes with >2 antigen assignments						
EurAm/3	B*15:01:01G+B*38:01:01	0.002901144	B*15:24+B*39:05:01	1.56E-08	185970	NT
EurAm/2	B*15:01:01G+B*49:01:01	0.001784603	B*15:24+B*50:01:01	9.684E-07	1842	NT
EurAm/3	B*35:01:01G+B*38:01:01	0.002490868	B*39:05:01+B*53:01:01	8.424E-07	2956	NT
AFAm/27	B*35:01:01G+B*44:18	2.72748E-05	B*45:01+B*53:01:01	0.010124998	371	G2/27

Observed in Population/ No. Individuals	Genotype 1 (G1)	Frequency	Genotype 2 (G2)	Frequency	Ratio of Most Common to Least Common Genotype ^b	Genotype When Higher Resolution Testing Performed/No. Tested
AFAm/12	B*35:01:01G+B*49:01:01	0.003665214	B*50:01:01+B*53:01:01	0.002100566	1	G1/8, G2/4
EurAm/6	B*35:01:01G+B*49:01:01	0.001532227	B*50:01:01+B*53:01:01	5.22936E-05	29	G1/5, G2/1
AFAm/1	B*35:01:01G+B*57:02:01	0.000539002	B*35:08:01+B*57:03:01	2.77324E-05	19	G1/1
AFAm/5	B*44:18+B*50:01:01	3.9228E-06	B*45:0101G+B*49:01:01	0.002540929	647	G2/5
AFAm/5	B*44:18+B*78:02:02	8.82E-08	B*45:01:01G+B*51:01:01G	0.001961071	22234	NT
EurAm/34	C*03:04:01G+C*05:01:01G^d	0.015010448	C*03:07+C*08:02:01	0.00000465	3228	NT
AFAm/9	C*04:01:01G+C*08:04	0.004363235	C*04:04:01+C*08:02:01	3.06106E-05	142	G1/9
EurAm/51	C*05:01:01G+C*07:01:01G	0.030437498	C*07:09+C*08:02:01	0.00000465	6545	NT
EurAm/6	C*05:01:01G+C*16:01:01	0.006466461	C*08:02:01+C*16:02:01	0.0002015	32	G1/6
AFAm/19	C*05:01:01G+C*16:01:01	0.006902498	C*08:02:01+C*16:02:01	3.06106E-05	225	G1/18, G2/1

^a Population-specific frequencies from <http://bioinformatics.nmdp.org>. Genotype frequency = 2x frequency allele 1 x frequency allele 2.

^b Bold indicates the more common genotype based on calculated genotype frequency. In some cases, two genotypes appear to be present at similar frequencies and both are bolded.

^c B*15:03:01, B*15:10:01, B*15:18:01, B*15:37, B*15:61 all type serologically as B70; B*15:01:01 and B*15:07 type as B62; B*40:02:01 and B*40:06:01 type as B61 (HLA dictionary at http://www.ebi.ac.uk/cgi-bin/imgt/hla/get_dictionary.cgi).

^d Another "common" genotype with the same diploid sequence as A*02:01:01G+A*02:05:01G is A*02:14+A*02:22:01G with a frequency of 8.82E-08. Likewise, C*03:09+C*05:05 with a frequency of 2.28E-08 is another "common" genotype of C*03:04:01G+C*05:01:01G.

^e NT, not tested

Table 7

Impact of assigning an allele code to the result

Assignment of alternative genotypes to a single diploid sequence	Condensed assignment ^a	Expanded assignment ^{b,c}
A*02:01:01G+A*23:01:01G or A*02:36+A*23:04 or A*02:90+A*23:12	A*02:DRKS+A*23:FPRS	A*02:01:01G+A*23:01:01G or A*02:36+A*23:04 or A*02:90+A*23:12 or <i>A*02:01:01G+A*23:04 or</i> <i>A*02:01:01G+23:12 or</i> <i>A*02:36+A*23:01:01G or</i> <i>A*02:36+A*23:12 or</i> <i>A*02:90+A*23:01:01G or</i> <i>A*02:90+A*23:04</i>

^a Letter codes assigned by on National Marrow Donor Program (6) can be interpreted based on <http://bioinformatics.nmdp.org> under DNA type lookup tool.

^b Italics indicate genotypes created when the code is expanded without knowledge of the original alternative genotypes.

^c To simplify the example, the G coding has been used to represent alleles identical in their antigen recognition domain exons. Also not included is the fact that the letter codes do not include nonsynonymous variation so that the number of combinations is actually larger than listed here. For example, A*02:DRKS includes all alleles designated as A*02:01; not just the subset of alleles assigned by sequencing (A*02:01:01:01, A*02:01:01:03, A*02:01:08, A*02:01:11, A*02:01:14, A*02:01:15, A*02:01:21). A*02:01:01:02L is designated as A*02:01L in the NMDP code list.