COLLEGE of AMERICAN PATHOLOGISTS

ARCHIVES
of Pathology & Laboratory Medicine

EARLY ONLINE RELEASE

Note: This article was posted on the *Archives* Web site as an Early Online Release. Early Online Release articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the *Archives*. Early Online Release articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2016-0537-RA

The final published version of this manuscript will replace the Early Online Release version at the above DOI once it is available.

Targeted Next-Generation Sequencing for Human Leukocyte Antigen Typing in a Clinical Laboratory

Metrics of Relevance and Considerations for Its Successful Implementation

Manish J. Gandhi, MD; Deborah Ferriola, BS; Yanping Huang, MD, PhD; Jamie L. Duke, PhD; Dimitri Monos, PhD

• ***Context.***—**Numerous feasibility studies to type human leukocyte antigens (HLAs) by next-generation sequencing (NGS) have led to the development of vendor-supported kits for HLA typing by NGS. Some clinical laboratories have introduced HLA-NGS, and many are investigating the introduction. Standards from accrediting agencies form the regulatory framework for introducing this test into clinical laboratories.**

Objectives.—**To provide an assessment of metrics and considerations relevant to the successful implementation of clinical HLA-NGS typing, and to provide as a reference a validated HLA-NGS protocol used clinically since December 2013 at the Children's Hospital of Philadelphia (Philadelphia, Pennsylvania).**

Data Sources.—**The HLA-NGS has been performed on 2532 samples. The initial 1046 and all homozygous samples were also typed by an alternate method. The HLA-NGS demonstrated 99.7% concordance with the alternate method. Ambiguous results were most common at the DPB1 locus because of a lack of phasing between exons 2 and 3 or the unsequenced exon 1 (533 of 2954 alleles; 18.04%) and the DRB1 locus because of not sequencing exon 1 (75 of 3972 alleles; 1.89%). No ambiguities were detected among the other loci. Except for 2 false homozygous samples, all homozygous samples (1891) demonstrated concordance with the alternate method. The article is organized to address the critical elements in the preanalytic, analytic, and postanalytic phases of introducing this assay into the clinical laboratory.**

Conclusions.—**The results demonstrate that HLA typing by NGS is a highly accurate, reproducible, efficient method that provides more-complete sequencing information for the length of the HLA gene and can be the single methodology for HLA typing in clinical immunogenetics laboratories.**

(*Arch Pathol Lab Med*. doi: 10.5858/arpa.2016-0537-RA)

Advances in technology have resulted in the evolution of human leukocyte antigen (HLA) typing from serologic-based methods to molecular methods. This has enabled the characterization of HLA antigens and alleles.[1] The HLA antigens are highly immunogenic and mostly polymorphic,[2] and matching HLA alleles are associated with better outcomes in transplantation.[3] In addition, HLA alleles are relevant in pharmacogenomics,[4] disease association,[5] and human population genetics studies.[6]

Because of the lower throughput of the available technologies and to reduce the number of polymorphic sites to interrogate, early molecular strategies for clinical HLA typing focused on characterization of the antigen recognition site only (exons 2 and 3 for class I; exon 2 for class II). Initial technologies to interrogate those exons, such as the sequence-specific oligonucleotide (SSO) and sequence-specific primer (SSP) methods, were able to produce low-resolution (first field) or antigen-level typing, with subsequent improvements allowing some level of higher resolution (second field) or allele-level typing. Currently, high-resolution typing by Sanger sequencing or sequence-based typing (SBT) is considered the gold standard. However, SBT only interrogates a few exons of the HLA gene and is, thus, unable to resolve all ambiguities. In addition, SBT does not provide phase information, which often results in ambiguous typing because of phase ambiguity among heterozygous polymorphic positions (within and between exons).[7]

Next-generation sequencing, by virtue of massively parallel sequencing and clonal DNA amplification, is able to provide allele-level typing with minimal ambiguity. Numerous studies in the past few years have demonstrated the feasibility of HLA typing by NGS and have led to the development and availability of commercially available products on different platforms.[8–11]

However, NGS has its limitations: turnaround time is in the range of days by NGS, whereas SSP, quantitative

Accepted for publication December 6, 2016.

From the Division of Transfusion Medicine, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota (Dr Gandhi); the Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Ms Ferriola and Drs Huang, Duke, and Monos); and the Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Dr Monos).

Drs Duke and Monos and Ms Ferriola receive royalties from the distribution of the Holotype HLA assay (Omixon, Budapest, Hungary), and Dr Monos is a consultant to, and owns stock in, Omixon. Drs Gandhi and Huang have no relevant financial interest in the products or companies described in this article.

Reprints: Manish J. Gandhi, MD, Division of Transfusion Medicine, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55902 (email: gandhi.manish@mayo.edu).

polymerase chain reaction, or even reverse SSO can generate data in a matter of hours; instrumentation costs are high; and it is difficult to determine the clinical significance of newly discovered HLA alleles because of polymorphisms outside the antigen-recognition site. In addition, NGS generates a plethora of sequencing data that needs to be reliably and reproducibly analyzed by software programs for genotyping.

Human leukocyte antigen typing by NGS includes steps with which an HLA laboratory may be unfamiliar, and HLA typing is distinct from other tests using NGS. The idiosyncrasy of the HLA system, therefore, has resulted in the development of standards by the American Society of Histocompatibility and Immunogenetics.[12] Those standards provide the regulatory framework that allows the proper introduction of this technology into clinical diagnostic laboratories. The early adopters of NGS for HLA typing include registries or laboratories that primarily support bone marrow transplantation because this technology offers accurate HLA typing that costs less than Sanger sequencing does.[13–15] HLA typing by NGS is also being used by laboratories serving solid organ transplant programs because high-resolution typing of the donor and recipient helps in the interpretation of anti-HLA antibody specificities.[16]

We describe a systematic presentation of the different elements that constitute the preanalytic, analytic, and postanalytic systems necessary for the practice of HLA typing by NGS in clinical diagnostic laboratories. This will be illustrated with a validated, robust, reproducible method for HLA typing by NGS developed at the Children's Hospital in Philadelphia (CHOP; Philadelphia, Pennsylvania)[9,17] and used clinically since December 2013. This assay was successfully tested by 6 laboratories that had varying experience with NGS, and the assay is now available as a commercial kit.

MATERIALS AND METHODS

Clinical HLA typing by NGS was performed with an assay developed at CHOP,[9] which is now licensed to Omixon (Budapest, Hungary) on the MiSeq platform (Illumina, San Diego, California). Samples (n = 2532) were tested with a commercial kit (Holotype, Omixon) for 8 or more loci (full length HLA-A, HLA-B, HLA-C, DPA1, DQA1, and DQB1; introns 1 to 4 of DRB1; and intron 1 to the 3′ untranslated region of DPB1 (details of which were previously published[9]).

Details of the 2532 samples sequenced between December 2013 and October 2016 for cases of bone marrow transplantation, solid organ transplantation, proficiency samples, outreach, disease association studies, and research are summarized in the Table. Initial typing included 5 loci (HLA-A, HLA-B, HLA-C, DRB1, and DQB1). Sequencing of DQA1, DPA1, and DPB1 was introduced later. A variable number of samples were typed at each loci due to the time point at which genes were introduced into the assay, in addition to the requests by clinicians and researchers as to the loci typed.

The initial 1046 samples (41.3%) and subsequent samples demonstrating homozygosity were characterized in parallel with a second method (SBT, SSO, or SSP) to confirm the genotyping results by NGS and to identify possible causes of discrepancies. Two analysis programs were used to genotype the NGS data, Target (version 1.7/1.8, Omixon) and NGSengine (version 1.3/1.6/1.9/2.1, GenDX, Utrecht, the Netherlands).[16] The 2 genotyping programs used for analysis to safeguard against systematic analysis errors agreed in 94.8% (26 508 of 27 488) of the best-match allele calls up to the third field.

RESULTS

For the initial 1046 samples collected between December 2013 and June 2015, concordance was 99.7% between NGS and the secondary typing method (SBT, SSP, or SSO). Discordance was due to incorrect or incomplete (ambiguous) SBT or SSP typing, allele imbalances, NGS genotyping software errors, incorrect reporting of results, and new alleles. In cases with ambiguous typing results (eg, >1 pair of alleles per locus), there was concordance between NGS and the alternate method; however, NGS was able to further narrow the alternative ambiguous choices to 2 or 3 per case.

Ambiguous results are most common at the DPB1 and DRB1 loci. The DPB1 ambiguities are due to a lack of phasing between exons 2 and 3 (large intron 2, about 4 kilobases [kb]) or the nonsequenced exon 1 (533 alleles).

DRB1 ambiguities are due to not sequencing exon 1 (75 cases or 1.89% of total DRB1 alleles). These ambiguities exclusively involved HLA-DRB1*12:01 versus HLA-DRB1*12:10, which have identical sequences for exons 2, 3, 4, 5 and 6, thus, result in the same mature molecule at the cell surface. No ambiguities were detected among the other loci up to the third field of typing.

During the study period, CHOP identified 59 new alleles (see the Table). All of the novel alleles were due to polymorphisms identified in the exonic regions. No attempt was made to compute and characterize the novelties that existed in the introns, untranslated regions, or uncharacterized exons compared to the IMGT/HLA database version in use at the time of analysis.

Comparison of homozygous samples by NGS and an alternate method demonstrated a high degree of concordance and confirmation of homozygosity. In addition, when evaluating 1893 samples for homozygosity, CHOP found 2 cases of false homozygous typing by NGS when compared with an alternate method, one each for HLA-DRB1 and DQB1. Both of those cases were associated with poor quality/quantity of DNA. Additional analysis revealed the presence of 2 alleles; however, the minor allele represented less than 10% of all reads and was filtered out by the software. To remedy that problem, the software programs were adjusted to detect lower levels of a minor allele.

DISCUSSION

Preanalytic Systems

The preanalytic system includes sample selection; DNA extraction; DNA quality and quantity; targeted polymerase chain reaction method (PCR) amplification; quantification and normalization of the amplicons; fragmentation, end repair, adaptor ligation; and size selection and sequencing.

Sample Selection and DNA Extraction.—Current American Society of Histocompatibility and Immunogenetics guidelines defer to the laboratory to select the appropriate number of samples used to validate the NGS method. Key considerations include samples collected from multiple sources and samples that include allele specificities that make up a large percentage of each race/ethnic group. In its original validation, CHOP used 253 samples that were typed by Sanger sequencing.[9] Another laboratory used 290 samples (including 79 samples in which DNA was extracted from buccal swabs) for its validation.[18] The usual source of DNA is blood; not infrequently in pediatric centers or blood registries, however, the source is buccal cells, and for large clinical or research studies, saliva is often preferred. In

Human Leukocyte Antigen (HLA) Typing Summary of 2532 Samples at Children's Hospital of Philadelphia (Philadelphia, Pennsylvania)								
Metric	HLA-A	HLA-B	HLA-C	HLA-DPA1	HLA-DPB1	HLA-DQA1	HLA-DQB1	HLA-DRB1
Total samples typed, No.	1965	2102	1989	750	1477	1436	2039	1986
Total alleles typed, No.	3930	4204	3978	1500	2954	2872	4078	3972
Homozygous samples, No. (%) (n = 1891)	237 (12.06)	133 (6.33)	216 (10.86)	412 (54.93)	285 (19.30)	173 (12.05)	277 (13.59)	158 (7.96)
Ambiguous alleles at third field, No. (%)	0 (0)	0 (0)	0 (0)	0 (0)	533 (18.04)	0 (0)	0 (0)	75 (1.89)
Novel alleles, No. (%)	9 (0.23)	2 (0.05)	3 (0.08)	17 (1.13)	17 (0.58)	3 (0.1)	6 (0.15)	2 (0.05)

archeology or forensics, a number of other various tissues, such as bones, may be the available source of DNA.

Genomic DNA is extracted using commercially available kits, either manually or on automated platforms. At CHOP, this was achieved by using the EZ1 DNA extraction instrument (Qiagen, Hilden, Germany) with the EZ1 Blood 350-μL kit (Qiagen) for peripheral blood or the EZ1 DNA Tissue kit (Qiagen) for buccal cells. The quality of the DNA obtained may limit the method that can be used for the next steps in sequencing. For example, methods using long-range PCR require intact DNA at least several kilobases long because the amplicons for delineating the HLA genes can range from 2 to 12 kb. We measured quantity by fluorescent-based methods (PicoGreen, Qubit, Thermo Fisher Scientific; or SpectraMax, Molecular Devices, Sunnyvale, California) and measured quality with the 260/230 indications of a spectrophotometer (NanoDrop, Thermo Fisher Scientific). The different protocols for HLA typing available on the market require amounts of DNA ranging from 5 to 30 ng/μL. Both DNA quantity and quality must be within the suggested range for optimal PCR performance, which is critical for the subsequent steps.

Template Preparation.—This step involves the isolation of the HLA-specific region of the genomic DNA prior to library preparation. For the purpose of clinical HLA typing by NGS, 2 major approaches have been used: (1) targeting the full-length gene using long-range PCR, or (2) multiplex PCR targeting exons encoding for the antigen-recognition site and the surrounding genomic regions. Other approaches have been used for nonclinical purposes, including sequence capture with oligo-based hybridization, reverse transcription of RNA into complementary DNA, and whole genome sequencing.[19]

An important consideration for the laboratory in choosing from the various HLA typing by NGS protocols available is the ability to provide a library that satisfies many metrics postsequencing for accurate HLA typing by the software analysis program. Issues of cost, numbers of samples processed, and simplicity are all relevant and important, but the satisfaction of certain metrics is the top priority because of the ever-evolving complexity of the HLA system. These key metrics, including sample, locus, and allele balance as well as uniformity of coverage, are described in the Analytic Systems section below.

Metrics of Relevance for Library Preparation.—The 4 metric steps relevant to the NGS sequencing of HLA genes are as follows:

PCR Amplification of the Targeted HLA Genes.—Laboratories have different requirements in terms of targeting the number of HLA genes, based on their clinical programs' requirements. Because laboratories are increasingly making HLA typing by NGS the primary method of HLA typing, protocols that can cover all the clinically relevant HLA genes (such as protocols covering up to 11 HLA genes and possibly more) is optimal.

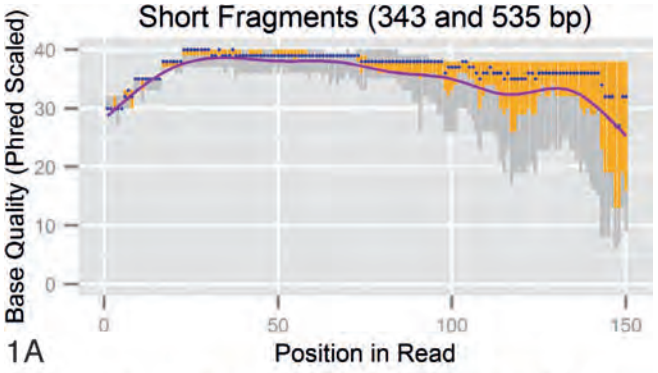
Having robust and reproducible amplification of every targeted locus and of every sample is important. There can be no credible HLA genotyping unless there is good amplification across each target with balanced representation of alleles. Some commercial protocols multiplex primers for the combined amplification of multiple genes because that simplifies the protocol by cutting down on the total number of PCR reactions. However, considering the complexity of the HLA system and the behavior of those multiple primers working together, to amplify the frequently closely related genes, multiplexing can have unpredictable consequences. The intelligence of the software programs analyzing the HLA data is very relevant. Extensive validation studies need to be performed to properly assess the efficiency and success of a multiplexed set of primers.

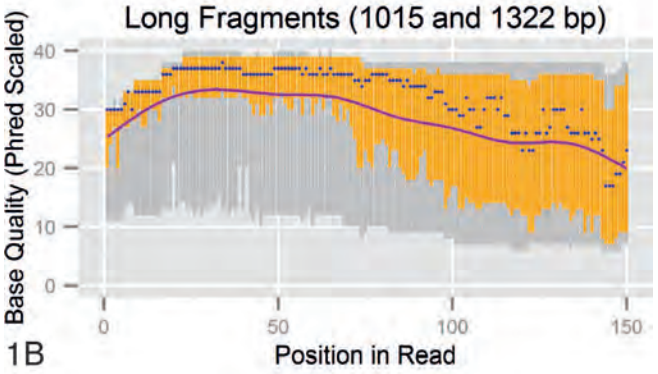
Confirmation/Quantitation of the Amplicons.—One approach to confirming and quantifying amplicons is to use gel electrophoresis to confirm amplification, followed by quantification and normalization. Because NGS includes parallel sequencing of many clones with multiple loci in a single run, all genes/amplicons in each sample need to be represented somewhat equally in the final library. When trying to optimize the capacity of a single flow cell for sequencing, no amplicons should be underrepresented or overrepresented because that risks insufficient sequencing reads for credible genotyping or unnecessary representation of reads from another amplicon/gene.

Normalization of the amount of double-stranded amplicons for library preparation has also been achieved with proprietary, paramagnetic beads and a proprietary DNA-capture method for the collection of comparable amounts of DNA bound to beads.[20] However, the use of beads for normalization does not negate the need for quantitation. Normalizing DNA amounts independent of quantitation has limitations because failures or weak amplifications may go undetected and not be rectified or adjusted early in the process.

Multiplexing the amplifications of different loci poses a challenge to the quantitation of each individual locus. If amplicons for each locus are different sizes, gel electrophoresis can be useful in detecting amplification failure.

Fragmentation of Amplicon DNA into Smaller Pieces.—Fragmentation of long amplicons into smaller pieces is required for the MiSeq (Illumina) and Ion Torrent (Thermo Fisher Scientific) platforms because they cannot sequence fragments larger than about 1200 bases and 600 to 700 bases, respectively. Fragmentation is not required with the PacBio (Pacific Biosciences, Menlo Park, California) or MinION (Oxford Nanopore Technologies, Oxford, United Kingdom) platforms.





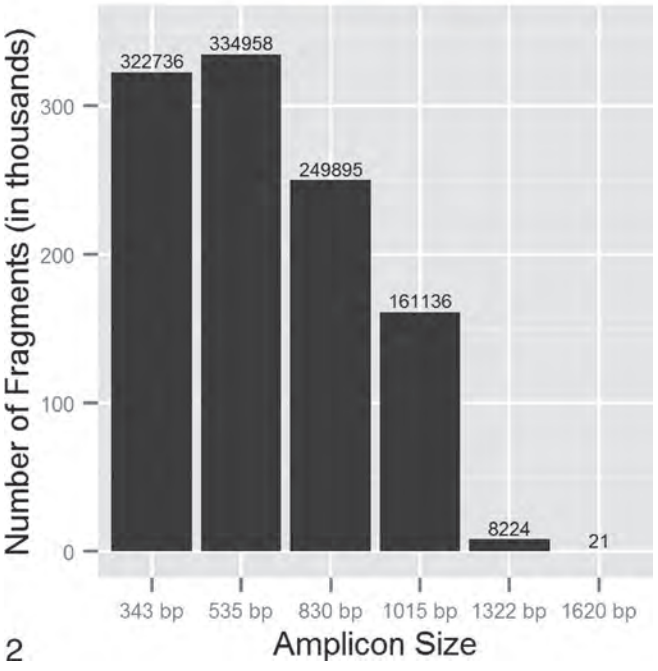


Figure 1. *Base quality along the length of sequencing reads. For each fragment size, box plots were generated showing the median (blue line), upper and lower quartiles (orange bands), and the 10% to 90% quantiles (grey bands) with a smoothed curve shown across the read length in purple. A, Short (343–535 base pair [bp]) fragments demonstrate a better quality of base calling. B, Long (1015–1322 bp) fragments demonstrate a comparatively lower quality of base calling, which drops further toward the end of the reads.*

Figure 2. *Sequencing efficiency of variable-size fragments on the MiSeq (Illumina, San Diego, California) platform: Equimolar amounts of different-size fragments (amplicons of variable sizes as shown) were sequenced together in a single run. Histogram of the number of fragments sequenced per amplicon of different sizes reflects shorter fragments being preferentially sequenced versus longer fragments. Any size fragment that is more than 1000 base pairs has a compromised sequencing efficiency.*

A key consideration for this step, with respect to the MiSeq platform, is that the DNA should be fragmented in a measured way, so there are adequate representations of both smaller and larger fragments for optimal library preparation. Smaller fragments (Figure 1, A) have higher-quality sequencing data than larger fragments do (Figure 1, B). Even though larger fragments have lower quality, they are valuable since they provide distal phase information not available from smaller fragments. Moreover, the efficiency of being able to sequence fragments of different size varies with respect to the length of the fragment—as the size of the fragment increases the sequencing efficiency decreases. Fragments of 300 to 500 bases are sequenced efficiently, and as the size of the fragment increases the sequencing efficiency decreases such that larger fragments approaching 1.3 kb are not sequenced at all (Figure 2) by the Illumina MiSeq platform. Therefore, if sequencing of large fragments, with their lower efficiency and lower-quality sequencing, is desirable, they need to be included, along with smaller sizes (no less than 300 base pairs [bp]) to secure quality genotyping results. Optimally, the very small (<150 bp) and very large fragments (>1.3 kb) have the ability to interfere with the sequencing of the optimal sized fragments and should be excluded before sequencing. Different approaches for the right selection of sizes have been adopted by different protocols involving beads such as Ampure XP beads (Beckman Coulter, Brea, California) and/or instruments for size selection, such as the Pippin Prep (Sage Science, Beverly, Massachusetts).[9] The proper selection of sizes secures optimization of the sequencing run, increases the number of samples sequenced, provides high-quality sequencing data, and maximizes possible phasing.

Attachment of Sequencing Adaptors or Indexes. The next steps involve attaching sequencing adaptors and/or indexes to the fragmented HLA amplicons to differentiate one sample from another when multiple loci of multiple samples are sequenced simultaneously. The attachments are accomplished either by direct enzymatic ligation of the adaptor sequences on the fragments or by alternative protocols involving PCR, in which the primers include the relevant adaptor sequences to generate amplicons that include both the adaptors and the HLA fragment to be sequenced. Protocols involving a ligation reaction require additional optimization of this step to successfully ligate most of the fragments.

Sequencing can be from a single direction of the DNA fragment, which is common with Ion Torrent technologies, or from both directions of the DNA fragment (paired-end), which is more common with the Illumina platforms. Paired-end sequencing is easier to interpret because it can offer haplotype phasing over longer distances.

In our protocol, libraries were prepared from individual or pooled amplicons by enzymatic fragmentation, end repair, adenylation, and ligation of indexed adaptors.

Sequencing Platform.—Sequencers that have been used to type HLA by NGS can be broadly classified into those that read and phase a moderate-length fragment between 250 and 1000 bp (MiSiq and Ion Torrent platforms) or those that are long-read instruments (>10 kb, PacBio and MinION platforms). Currently, commercial kits for HLA

typing by NGS are available on medium-read instruments. The advantages and details of those technologies have been reviewed by De Santis et al[7] and Carapito et al.[19] Our protocol was performed on the Illumina MiSeq instrument.

Analytic Systems

Assessing Technical Aspects of the Run.—Bioinformatics is important for obtaining accurate HLA typing information. The sequence data produced are usually formatted as FASTQ files, which include the sequence and its corresponding Phred quality score. The bioinformatics for NGS can be broadly divided into 2 steps: (1) an initial step to remove the adapter sequences and low-quality reads or bases, followed by alignment to a reference sequence; and (2) variant calling and annotation. For HLA typing by NGS, a single reference sequence is not feasible because of the HLA system is highly polymorphic. As such, sequencing data obtained through NGS is aligned to all known HLA alleles that are available via the international ImMunoGeneTics project–HLA database, which is part of the Immuno Polymorphism Database (European Bioinformatics Institute, Hinxton, United Kingdom) and can be accessed at https://www.ebi.ac.uk/ipd/imgt/hla/.

The software analysis tools used to perform the HLA typing provides information on a number of relevant metrics that help the analysts assess and evaluate the quality of the run and the genotyping results. Some key metrics are:

Key Metrics.—*Uniformity of Coverage*.—Information regarding the depth of coverage throughout the length of a particular HLA gene is important. Software analysis programs provide that information and may have built-in filters to define the minimum depth of coverage suitable for obtaining accurate typing. Additionally, it is important for the operator to understand how the depth of coverage impacts the accuracy of HLA typing.

For several technical reasons, different protocols generate coverage profiles that are not uniform throughout the length of the amplicon; therefore, it is important to know whether key regions (like exons) that will affect the genotyping are properly covered. Some variation in depth of coverage along the length of the gene can be tolerated. However, when the number of loci included in a particular sequencing run is increased, there is a need to optimize the uniformity of coverage because the genotyping and the consensus sequence generated are only as good as the lowest point of coverage. Some regions may have a low depth of coverage, which does not provide a confident result, whereas others may have significantly greater depth than is needed, compromising efficiency. As the number of loci and, therefore, the number of samples to be genotyped increase, coverage uniformity becomes more relevant. For example, the software, by default, may require a minimum depth of coverage of 30 reads at any point throughout the length of the gene typed for each allele; however, there are circumstances in which going below that threshold may be acceptable, specifically when the polymorphisms of the 2 alleles of a locus are phased, the typing does not have a lot of noise from sequencing errors, and the locations with a low depth of coverage occur in a region that does not affect the genotyping (eg, introns, untranslated regions), even though doing so may affect the generation of consensus sequence. It is, therefore, important that the software analysis program provides a clear and unbiased picture of the depth of coverage throughout the length of the amplified region of any HLA gene.

Balanced Representation of Reads Derived From Different Loci and Different Samples.—The balanced representation metric is relevant to the efficiency of the sequencing run. The optimization of the run for genotyping output depends on the number of reads obtained from each locus and each sample. As the system is pushed to more loci or samples, the need for balanced outputs from each locus increases. The system is forgiving when fewer loci or samples are typed because the output of the sequencing platform is such that even the least-represented locus will have enough reads for credible genotyping. Considering that some genes are longer than others during long-range amplification, it is not easy to assign a strict requirement regarding the number of reads per locus to produce a reliable typing. An estimate of about 1500 reads are necessary for class I genes that are about 3.3 kb, assuming the reads are 250 bases long (to achieve a depth of coverage of about 100X throughout the length of the gene), and proportionally more reads are required for longer class II genes. If the reads are 150 bases long, then proportionally more reads per locus are needed to obtain the same depth of coverage as a sequencing run with reads of 250 bases. The software program should allow the user to specify the number of reads analyzed to achieve the optimal depth of coverage per locus for a good genotyping result.

Number of Samples Included in the Run.—All of the metrics above aim to optimize the capacity of the sequencing run and therefore to improve the economics of the HLA typing system by NGS. Because different laboratories handle different volumes, understandably, the community can tolerate a wide spectrum of products with different efficiencies. Eventually, however, to push this exciting technology to its limits, we need to account for known parameters that gain efficiencies.

Allele Imbalance and Possible Drop Out.—Unavoidably, during the first step of amplification, inefficiencies occasionally arise, and the amplifications of the 2 alleles of a single HLA gene are not balanced. The software programs must have the ability to provide information on the relative number of reads generated from each of the 2 alleles, so users can evaluate whether the observed imbalance is acceptable. Some software programs have adopted the approach of not reporting a particular allele if it is not represented adequately in the run. However, definitions of acceptability vary among programs. Based on our experience, there are times when even 5% of the total reads derived from a particular gene can be credible and enough to confirm the presence of a minor allele.[9] Our experience leads us to believe that the software program should be flexible and should allow the operator/analyst to set the threshold at any desirable low level.

Another key consideration is determining true homozygosity from allele dropout. Answering that question with absolute certainty is difficult, but there are different ways to address that issue: (1) use of an alternate method, such as SSO, for a second, independent analysis; (2) use of available haplotype data; (3) analysis of sample quality by assessing both the quality and quantity of DNA; and (4) awareness of the performance of the primers for a particular locus and focusing attention on the limitations and sensitivities of the amplification reactions.

Postanalytic Systems

Protocol Simplicity; Protocol Robustness.—Evaluating the system for simplicity is important because complicated

processes frequently lead to mistakes. Simpler protocols most likely translate into more-efficient, faster processes with decreased chances for error, and that translates into better turnaround times. Protocol robustness is also relevant because having reproducible results is important for clinical laboratories, in particular, because the HLA system is a polymorphic genomic system with unpredictable variation.

Failures.—Inevitably, and primarily because of DNA quality/quantity, certain PCR reactions fail, and unless the typing protocol assesses the amplification at an early stage, the failure can go undetected until the end of the analysis process. In our experience, robust PCR amplification leads to successful genotyping, with the exception of software performance.

Turnaround Times.—Even though the manual time is significantly less than SSO, SSP, or SBT, the turnaround time of this test can be as much as several days and as few as 3 days. For example, in a protocol using long-range PCR, which takes several hours, laboratories can, preferably, set them up in late afternoon of the first day. The next day, the library preparation is performed, which takes several hours, depending on the setup process (manual or automated), and the sample can be placed on the sequencing platform that same day. If the laboratory adopts the 150-bp, paired-end chemistry of the Illumina MiSeq, the sequencing process takes close to 24 hours, followed by the sequence analysis in the evening hours of the third day after the run is completed. By the next morning (fourth day), the sequencing data are available for reviewing and reporting. Automation of many of those steps can increase the efficiency of the overall process to achieve a better turnaround time, which may be further improved in laboratories with multiple shifts. Availability of new sequencing platforms with improved capabilities will provide opportunities to adjust the current protocols to achieve shorter turnaround times.

New Alleles.—The detailed characterization of the HLA genes obtained by NGS may result in the identification of new alleles. These new alleles derive from polymorphisms within previously uncharacterized exons and intronic regions. Most new alleles thus identified by NGS are due to intronic polymorphisms[21] or to nonsynonymous mutations from single-nucleotide polymorphisms.[22,23] Methods to standardize reporting of those new alleles are being developed.[24,25]

Financial Considerations.—The NGS-based HLA typing offers robust reproducible assays that can type multiple samples at a minimum of 2 fields (4 digits) and 11 loci in a single run, which results in a more-efficient and streamlined workflow, with decreased reflective typing and reduced employee time. Many laboratories are now implementing this technology not only to support stem cell transplantation but also to support solid organ transplantation. CHOP has been using NGS technology for HLA typing since December 2013, which has streamlined the workflow and significantly reduced the cost per test. Various vendor-developed protocols offer competitive prices per sample. The considerations mentioned above will help users evaluate and choose the best product or products for their needs and operations.

SUMMARY

Next-generation sequencing is a feasible option for performing all HLA typing in a clinical laboratory that serves various transplant programs. The method is reproducible, cost effective, accurate, efficient, and provides more-complete sequencing information of the HLA gene. Successful HLA typing by NGS requires the combination and molding of various elements, including the sequencing platform, the sample preparation protocol, and the bioinformatics (to analyze data quality and the ability of the software to handle the complexity of the HLA genes) to provide accurate HLA typing. This report provides a brief description of the key elements that contribute to successful HLA typing by NGS. The 4 million-bp major histocompatibility complex region in which the HLA genes are located has the greatest density of markers associated with different diseases and traits.[26] Complete characterization of the HLA genes and the major histocompatibility complex will lead to better understanding of their roles in health and disease.

We thank Patricia Vasalos, BS, for providing support and coordination for all the NGS validation manuscripts in this series; Ms Vasalos is an employee of the College of American Pathologists (Northfield, Illinois). We thank the members of College of American Pathologists Histocompatibility and Identity Testing Committee for their ongoing deliberations and discussions, which generated publications and checklist items.

References

1. Bontadini A. HLA techniques: typing and antibody detection in the laboratory of immunogenetics. *Methods*. 2012;56(4):471–476.
2. Trowsdale J. The MHC, disease and selection. *Immunol Lett*. 2011;137(1–2):1–8.
3. Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*. 2007;110(13):4576–4583.
4. Pavlos R, Mallal S, Phillips E. HLA and pharmacogenetics of drug hypersensitivity. *Pharmacogenomics*. 2012;13(11):1285–1306.
5. Holoshitz J. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discov Med*. 2013;16(87):93–101.
6. Sanchez-Mazas A, Meyer D. The relevance of HLA sequencing in population genetics studies. *J Immunol Res*. 2014;2014:971818.
7. De Santis D, Dinauer D, Duke J, et al. 16th IHIW: review of HLA typing by NGS. *Int J Immunogenet*. 2013;40(1):72–76.
8. Monos D, Maiers MJ. Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): consequences, opportunities and challenges. *Hum Immunol*. 2015;76(12):883–886.
9. Duke JL, Lind C, Mackiewicz K, et al. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA*. 2016;87(3):141–152.
10. Profaizer T, Lázár-Molnár E, Close DW, Delgado JC, Kumánovics A. HLA genotyping in the clinical laboratory: comparison of next-generation sequencing methods. *HLA*. 2016;88(1–2):14–24.
11. Moalic-Allain V, Mercier B, Gueguen P, Ferec C. Next generation sequencing with a semi-conductor technology (Ion Torrent PGM™) for HLA typing: overall workflow performance and debate. *Ann Biol Clin (Paris)*. 2016;74(4):449–456.
12. American Society of Histocompatibility and Immunogenetics. *Standards for Accredited Laboratories*. http://c.ymcdn.com/sites/www.ashi-hla.org/resource/resmgr/Docs/Accreditation/Standards_/2014ASHIStandardsandGuidance.pdf. Accessed November 1, 2016.
13. Zhou M, Gao D, Chai X, et al. Application of high-throughput, high-resolution and cost-effective next generation sequencing-based large-scale HLA typing in donor registry. *Tissue Antigens*. 2015;85(1):20–28.
14. Lange V, Böhme I, Hofmann J, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. 2014;15:63.
15. Cereb N, Kim HR, Ryu J, Yang SY. Advances in DNA sequencing technologies for high resolution HLA typing. *Hum Immunol*. 2015;76(12):923–927.
16. Duke J, Ferriola D, Huynh A, et al. HLA typing by NGS in the clinical lab: our one and a half years experience [abstract P128]. *Hum Immunol*. 2015;76(suppl):129.
17. Lind C, Ferriola D, Mackiewicz K, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Hum Immunol*. 2010;71(10):1033–1042.
18. Weimer ET, Montgomery M, Petraroia R, Crawford J, Schmitz JL. Performance characteristics and validation of next-generation sequencing for human leucocyte antigen typing. *J Mol Diagn*. 2016;18(5):668–675.

19. Carapito R, Radosavljevic M, Bahram S. Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies [published online ahead of print April 5, 2016]. *Hum Immunol*. 2016;77(11):1016–1023. doi:10.1016/j.humimm.2016.04.002.

20. Data Sheet: TruSightR HLA v2 Sequencing Panel, San Diego, Illumina. http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/trusight-hla-v2-data-sheet-070-2016-002.pdf. Accessed November 1, 2016.

21. Clark PM, Duke JL, Ferriola D, et al. Generation of full-length class I human leukocyte antigen gene consensus sequences for novel allele characterization [published online ahead of print September 27, 2016]. *Clin Chem*. 2016; 62(12):1630–1638. doi:10.1373/clinchem.2016.260661.

22. Hernández-Frederick CJ, Cereb N, Giani AS, et al. Detection of 549 new HLA alleles in potential stem cell donors from the United States, Poland and Germany [published correction in *HLA*. 2016;88(1–2):74]. *HLA*. 2016;87(1):31–35.

23. Hernández-Frederick CJ, Giani AS, Cereb N, et al. Identification of 2127 new HLA class I alleles in potential stem cell donors from Germany, the United States and Poland. *Tissue Antigens*. 2014;83(3):184–189.

24. Mack SJ, Milius RP, Gifford BD, et al. Minimum information for reporting next generation sequence genotyping (MIRING): guidelines for reporting HLA and KIR genotyping via next generation sequencing. *Hum Immunol*. 2015; 76(12):954–962.

25. Milius RP, Heuer M, George M, et al. The GL service: Web service to exchange GL string encoded HLA & KIR genotypes with complete and accurate allele and genotype ambiguity. *Hum Immunol*. 2016;77(3):249–256.

26. Clark PM, Kunkel M, Monos DS. The dichotomy between disease phenotype databases and the implications for understanding complex diseases involving the major histocompatibility complex. *Int J Immunogenet*. 2015;42(6): 413–422.