The Journal of Molecular Diagnostics, Vol. 22, No. 7, July 2020

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Long-Read Nanopore Sequencing Validated for Human Leukocyte Antigen Class I Typing in Routine Diagnostics

Benedict M. Matern, Timo I. Olieslagers, Mathijs Groeneweg, Burcu Duygu, Lotte Wieten, Marcel G.J. Tilanus, and Christina E.M. Voorter

From the Transplantation Immunology, Tissue Typing Laboratory, Maastricht University Medical Center, Maastricht, the Netherlands

Accepted for publication April 2, 2020.

Address correspondence to Christina E.M. Voorter, Ph.D., Transplantation Immunology, Tissue Typing Laboratory, Maastricht University Medical Center, P.O. Box 5800, Maastricht 6202 AZ, the Netherlands. E-mail: c.voorter@mumc.nl.

Matching of human leukocyte antigen (HLA) gene polymorphisms by high-resolution DNA sequence analysis is the gold standard for determining compatibility between patient and donor for hematopoietic stem cell transplantation. Single-molecule sequencing (PacBio or MinION) is a newest (third) generation sequencing approach. MinION is a nanopore sequencing platform, which provides long targeted DNA sequences. The long reads provide unambiguous phasing, but the initial high error profile prevented its use in high-impact applications, such as HLA typing for HLA matching of donor and recipient in the transplantation setting. Ongoing developments on instrumentation and basecalling software have improved the per-base accuracy of $1D^2$ nanopore reads tremendously. In the current study, two validation panels of samples covering 70 of the 71 known HLA class I allele groups were used to compare third field sequences obtained by MinION, with Sanger sequence—based typing showing a 100% concordance between both data sets. In addition, the first validation panel was used to set the acceptance criteria for the use of MinION in a routine setting. The acceptance criteria were subsequently confirmed with the second validation panel. In summary, the present study describes validation and implementation of nanopore sequencing HLA class I typing method and illustrates that nanopore sequencing technology has advanced to a point where it can be used in routine diagnostics with high accuracy. *(J Mol Diagn 2020, 22: 912—919; https://doi.org/10.1016/j.jmoldx.2020.04.001)*

Human leukocyte antigen (HLA) is the human major histocompatibility complex, a group of genes comprising the most polymorphic loci in the human genome.[1] The HLA genes are encoded within the short arm of the human chromosome 6, and they are grouped by both function and morphology into two general classes, HLA classes I and II. The hyperpolymorphism of HLA is demonstrated in the Immuno-Polymorphism Database—ImMunoGeneTics (IPD-IMGT)/HLA database,[2] which currently lists 18,691 class I and 7065 class II HLA alleles (release 3.38.0). The nucleotide polymorphism is reflected in the protein polymorphism, which allows the HLA class I or class II molecules to present a wide variety of intracellular and extracellular antigens, respectively.

HLA polymorphism enables the immune system to respond to a large variety of pathogens and diseases, but generates challenges in performing solid organ and stem cell transplantation requiring HLA typing. Both solid organ and stem cell transplantation involve the introduction of nonself tissue to the body, and have the inherent risk of adverse immune response. In solid organ transplantation, mismatched donor HLA can induce the production of donor-specific anti-HLA antibodies, which can bind to the HLA on the transplanted tissue,[3] triggering antibody-mediated organ rejection. Although matching of patient and donor HLA alleles may not be possible because of low organ availability, high-resolution typing of HLA alleles identifies the amino acid sequence and consequent epitope structure of the HLA molecule, providing insight in

B.M.M. and T.I.O. contributed equally to this work.
Disclosures: None declared.

Copyright © 2020 Association for Molecular Pathology and American Society for Investigative Pathology. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jmoldx.2020.04.001

the targets of the anti-HLA antibodies.[4,5] In stem cell transplantation, T cells in the donor allograft may recognize the HLA-peptide complexes expressed on recipient tissue as nonself, triggering global immune activation and leading to graft-*versus*-host disease. High-resolution HLA typing for stem cell transplantation is especially critical, because allele mismatches are linked with sometimes fatal adverse effects.[6,7]

High-resolution HLA typing has further applications in the area of drug hypersensitivity,[8] like abacavir and carbamazepine. *HLA-B*57:01* has been correlated with hypersensitivity to abacavir,[9] a drug commonly used to treat HIV, whereas *HLA-B*15:02* has been correlated with hypersensitivity to carbamazepine,[10] a drug used in epilepsy treatment. Typing for the *HLA-B* alleles informs the caregiver's choice of treatment methods and prevents adverse drug interactions. Furthermore, the HLA region is correlated with the highest numbers of human diseases in the human genome,[11] and high-resolution sequencing allows refinement of understanding of the function of HLA and its relationship to the causality of diseases.

In recent decades, increased understanding of the HLA genetics, advances in DNA sequencing technology, and a general lack of allele-specific antisera have led to a shift in HLA typing method. Serologic methods have often been replaced or supplemented by DNA sequence—based methods, allowing analysis of HLA at the nucleotide level.[12] Sanger sequencing was developed in 1977,[13] and Sanger sequence—based typing (SSBT) has been the gold standard for HLA allele assignment for many years. However, with the increase in available HLA allele sequences, the heterozygous Sanger sequencing approach became more and more cumbersome, because of the increase in ambiguous typing results, which needed additional sequencing to resolve. This problem could be circumvented by group-specific full-length amplification and separate sequencing of the alleles,[14] albeit this requires a preceding low-resolution typing to determine the allele groups. Technological advances led to next-generation sequencing (NGS), intended to be faster and cheaper than Sanger-based sequencing and with the huge advantage of separate allele sequencing. These technologies, including reversible terminator (Illumina, San Diego, CA)[15] and semiconductor (Ion Torrent Systems, Thermo Fisher Scientific, Waltham, MA)[16] methods, are commonly used for HLA typing,[17] but have the disadvantage of short sequences, impairing with correct alignment and phasing of the alleles. Third-generation sequencing is the term used to describe a new era of sequencing technologies that are focused on the analysis of single molecules (ie, long stretches of DNA without the need to fragment the DNA into smaller pieces as is the case for NGS-based techniques). Third-generation sequencing includes single-polymerase molecule (Pacific Biosystems, Menlo Park, CA)[18] and nanopore-based sequencing (Oxford Nanopore, Oxford, UK)[19] technologies.

The MinION[20] is a portable nanopore sequencing platform that generates ultralong reads and requires little initial equipment investment. MinION uses a flow cell that contains a membrane with a grid of embedded nanopores, each of which is capable of binding to a DNA molecule. An electrical potential difference between both sides of the membrane is applied, generating a current across the membrane. Single-stranded DNA is passing through the pore with help from an accompanying motor protein.[21] As nucleotides pass through the pore, disruptions in the current and resulting electrical signal are measured by the MinION integrated circuits. This signal is characteristic of the bases that are present within the pore, because of the varying size and morphology of the nucleotides. Software tools can translate the electrical signal back to the original nucleotide sequence, in a process known as basecalling.[22] The MinION is capable of natively sequencing a piece of single-stranded DNA, in a process known as 1D sequencing. A quality improvement is provided by MinION $1D^2$ protocols, where the sample preparation results in double-stranded DNA with adapter proteins attached to both ends, allowing both strands of the DNA to be individually sequenced. The MinION basecallers pair two complementary single-stranded reads *in silico*, resulting in a single, higher-accuracy read. The quality of the basecalling of a MinION read is represented by the per-base Phred quality scores,[23] which were averaged over the length of each $1D^2$ read to calculate the mean. The mean reported Phred score over all the reads was found to be 18.5, which corresponds to 98.6% read accuracy. For the hyperpolymorphic HLA genes, in which each nucleotide difference can actually account for another allele, a high level of accuracy is essential to obtain reliable results. With reaching this high level of accuracy, typing of the hyperpolymorphic HLA genes by this MinION approach came into the picture.

The potential of using the MinION as sequencing platform for the analysis of HLA has been described previously,[24–31] but up until now it has not been used in routine diagnostics. The current article describes the complete validation process of full-length HLA class I single-molecule sequencing and typing method using the Oxford Nanopore MinION and the implementation in the routine diagnostic setting.

Materials and Methods

Samples and Validation Process

To validate the MinION approach for reliable HLA typing, quality standards and validation metrics needed to be generated. A panel of samples with known HLA high-resolution typing was sequenced and typed with the standard $1D^2$ MinION protocol. The initial panel consisted of 33 samples, which cover 70 of the 71 known class I allele groups, excluding *HLA-B*83*, which was not available in our laboratory (Table 1). This panel was subjected to the

Table 1 Samples Included in the Initial Validation

Sample identifier	*HLA-A**		*HLA-B**		*HLA-C**	
1	01:01:01:01	30:01:01	15:10:01	42:01:01	03:04:02	17:01:01
2	02:06:01:01	02:06:01:04	51:01:01	59:01:01	01:02:01	14:02:01
3	02:01:01:01		52:01:01	73:01	07:01:01	15:05:01
4	01:02	66:01:01:01	58:01:01	58:02:01	03:02:02:01	06:02:01
5	02:01:01	36:01	15:03:01:02	51:01:01	01:02:01	12:03:01
6	02:01:01	31:01:02:01	15:01:01:01	67:01:01	07:02:01	
7	26:01:01:01	30:02:01:01	18:01:01	40:01:02	03:04:01:01	05:01:01:01
8	24:02:01:01	32:01:01:01	14:01:01:01	18:01:01	07:01:01	08:02:01:02
9	01:01:01:01	31:01:02:01	08:01:01	40:01:02	03:04:01:01	07:01:01
10	01:01:01:01	03:01:01:01	08:01:01	45:01:01	06:02:01:03	07:01:01
11	02:06:01:01	30:02:01:01	18:01:01	39:08	05:01:01	07:02:01
12	29:02:01:01	69:01:01:01	39:06:02:01	55:01:01	01:02:01	07:02:01
13	03:01:01:05	66:01:01:01	15:03:01:02	52:01:01:01	02:10:01:01	12:02:02:01
14	43:01	74:01:01	15:03:01:02	44:03	02:10:01:01	08:04:01
15	02:01:01	34:01:01	40:02:01	56:02:01	01:02:01	15:02:01:01
16	30:01:01	33:01:01:01	53:01:01	81:01	04:01:01	08:04:01
17	02:01:01	24:02:01:01	44:02:01:01	49:03	05:01:01:02	07:01:01
18	03:01:01:05	25:01:01:01	37:01:01:01	47:01:01:03	06:02:01:01	
19	02:03:01	02:07:01	38:02:01	46:01:01	01:02:01	07:02:01
20	01:01:01:11	02:01:01	35:04:01	82:01:01:01	03:02:02:01	04:01:01
21	02:01:01		44:09	50:01:01:01	05:01:01:02	
22	24:02:01:01	26:02:01	40:06:01:01	54:01:01	01:02:01	08:01:01:01
23	24:02:01	33:03:01	15:07:01:02	15:16:01:02	03:03:01	14:02:01:02
24	26:01:01:01	74:01:01	81:01	78:01:01:02	16:01:01	18:01
25	23:01:01:01	32:01:01:01	41:02:01	44:03:01	04:01:01	17:03:01
26	11:01:01:01	24:02:01	27:06	48:01:01	01:02:01	08:01:01:01
27	01:01:38L	02:01:01	15:17:01:01	57:01:01:01	06:02:01	07:01:02
28	02:01:01	25:01:01:01	15:78:01	38:01:01:01	03:04:01:01	12:03:01:01
29	30:09	80:01:01:02	07:02:01	81:01:01	07:02:01	18:02
30	11:01:01:01	68:01:02:01	40:01:02	55:01:01	01:02:01	07:02:01
31	24:02:01:01	29:01:01:01	07:05:01	27:02:01:04	02:02:02	15:05:02
32	24:17	33:03:01	07:02:01	15:02:01:01	07:02:01:03	08:01:01:01
33	02:01:01:01	24:02:01	07:02:01	13:02:01:01	06:02:01:01	07:02:01

Samples were selected to cover 70 of the 71 known *HLA-A*, *HLA-B*, and *HLA-C* allele groups. *HLA-B*83* was not included, as it was not available in our laboratory.

MinION 1D^2 approach, with a focus on comparing results with our Sanger sequencing results (Supplemental Table S1) and determining read quality and coverage statistics. For each sample in the initial validation, read coverage was measured on each MinION run, for each demultiplexed MinION barcode, for each HLA locus within a barcode, and for both HLA alleles at a locus.

Acceptance criteria were defined in the initial validation panel, and the defined criteria were verified by sequencing and HLA typing a second panel of 67 samples (402 alleles) from our laboratory in a secondary validation phase parallel to Sanger sequencing (Supplemental Table S2). The samples were sequenced and analyzed using the combined analysis approach, with optimizations based on the initial validation. The samples were also typed using full-length allele-specific SSBT approach.[14] Typing results from MinION were compared with the SSBT results. Accuracy of HLA allele assignment of these samples was considered on the basis of criteria defined in the initial validation.

MinION Sequencing

The MinION sequencing and analysis method is outlined in Figure 1. The procedure starts with the PCR amplification of 300 ng DNA, purified and isolated from peripheral blood samples and according to the descriptions outlined in Voorter et al.[14] At a later time DNA isolated from buccal swabs was tested for MinION sequencing with comparable results. The full-length gene-specific amplification primers are located in the 5′ and 3′ untranslated regions of the *HLA-A*, *HLA-B* and *HLA-C* genes (Table 2). To sequence several samples simultaneously, specific Oxford Nanopore tag and barcode sequences were added to the primers. The tag sequence coordinates the basecalling and demultiplexing software, and the barcode sequence is used to sort

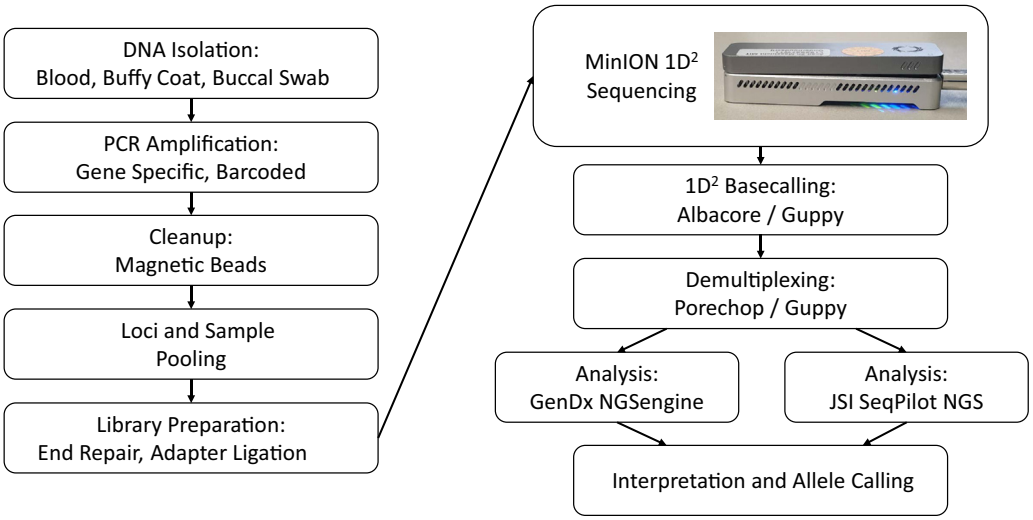

Figure 1 MinION sequencing and analysis. The first column depicts the steps of sample preparation, whereas data generation and analysis are in the second column. Data analysis is performed in a combined analysis approach, where the results from two software packages are compared for accurate allele calls. NGS, next-generation sequencing.

multiplexed samples. After amplification, presence of PCR products was confirmed by agarose gel electrophoresis, and PCR products were purified using CleanPCR magnetic beads (GC Biotech, Waddinxveen, the Netherlands). In addition, by using a bead versus DNA ratio of 1:1 during the purification, primer dimers were removed simultaneously. Subsequently, up to nine samples (27 PCR products) were equimolar pooled, with equal distribution between loci, to a total quantity of 1300 ng DNA.

Pooled samples were further prepared for MinION sequencing using the $1D^2$ sequencing kit (SQK-LSK308; Oxford Nanopore) and following manufacturer's protocol with a few minor adjustments. In short, the amplicon strands are end repaired and dA tailed using the NEBNext End Repair/dA tailing module (New England Biolabs, Ipswich, MA). These end-repaired amplicons were purified using CleanPCR beads and ligated with $1D^2$ adapters, which allows the nanopore to capture the complement strand immediately after the template. After another purification step, sequencing adapters were ligated onto the amplicons, which ensures that the DNA strands can enter the nanopore. The MinION, with an attached flow cell, was connected to a computer and quality control was performed, which checks for available and active nanopores, the R9.5 flow cell was primed, and 75 μL of the prepared $1D^2$ library was loaded into the flow cell for sequencing.

The sequencing run was performed for 16 hours and was controlled by MinKNOW software version 18.07.2 (Oxford Nanopore), which collects the 1D read data. Basecalling was performed initially using Albacore software version 2.3.1 and later updated to Guppy software version 3.2.4 (both from Oxford Nanopore). The basecaller first converts electrical signal from the 1D reads into a nucleotide sequence, and the sequences within 1D reads from complementary strands are subsequently paired and combined into higher-accuracy $1D^2$ reads. $1D^2$ reads containing the MinION tag and barcode sequences were demultiplexed initially by Porechop version 0.2.3 (*https://github.com/rrwick/Porechop*, last accessed December 1, 2019), later updated to the same Guppy software. Porechop/Guppy removes the non-HLA tag and barcode sequences from the reads and sorts the reads into FASTQ files corresponding to each barcode sequence.

Data analysis and interpretation were performed by a combined approach, using two separate software packages:

Table 2 Amplification Primers

Gene	Direction	Sequence	IMGT/HLA gDNA position
HLA-A	Forward	5′-GGATACTCACGACGCGGAC-3′	−137 to −119
HLA-A	Reverse	5′-GGGAGCACAGGTCAGCGTGGGAAG-3′	3075 to 3098
HLA-B	Forward	5′-GGCAGACAGTGTGACAAAGAGGC-3′	−420 to −398
HLA-B	Reverse	5′-CTGGGGAGGAAACACAGGTCAGCATGGGAAC-3′	3040 to 3070
HLA-C	Forward	5′-TCAGGCACACAGTGTGACAAAGAT-3′	−327 to −304
HLA-C	Reverse	5′-TCGGGGAGGGAACACAGGTCAGTGTGGGGAC-3′	3067 to 3098

This table contains the gene-specific amplification primers. Only the sequence that complements the HLA untranslated region sequence is shown; the amplification primers also include a section of sequence containing MinION tag sequences, as well as DNA barcodes, as provided by Oxford Nanopore.

gDNA, genomic DNA; HLA, human leukocyte antigen; IMGT, ImMunoGeneTics.

JSI SeqPilot SeqNext (NGS) HLA module version 4.4.0 (JSI, Ettenheim, Germany) and GenDx NGSengine version 2.11.0.11444 (GenDx, Utrecht, the Netherlands). Both are configured to ignore the regions containing amplification primers. Analysis of the sorted HLA $1D^2$ read data was performed in both software packages independently, and any discrepancies between the two programs were analyzed in detail.

The estimated duration of time for the MinION sequencing procedure is as follows for the different parts: PCR amplification, 5 hours; cleanup, pooling, and library preparation, 5 hours; and MinION $1D^2$ sequencing, 16 hours. Because basecalling and analysis time is highly dependent on the power of the computer used and the experience of the user, it is not possible to make any time estimation for this part.

SSBT Data

HLA sequences obtained by MinION sequencing were compared with typing results from full-length group-specific SSBT,[14] which is considered the gold standard for HLA typing in our laboratory. In brief, DNA was isolated and amplified using group- and allele-specific primers for the HLA class I genes. The DNA product was sequenced on a Sanger 3730 analyzer (Applied Biosystems, Foster City, CA), and sequence analysis and allele calling were performed using the JSI SeqPilot SeqHLA module.

Results

Initial Validation

For the initial validation panel, consisting of 33 samples covering 70 of the 71 known HLA class I allele groups (Table 1), *HLA-A*, *HLA-B*, and *HLA-C* were successfully amplified and sequenced using the MinION sequencing method. Data were analyzed and interpreted using two separate software packages: JSI SeqPilot SeqNext (NGS) HLA module and GenDx NGSengine. Because HLA analysis software programs specifically designed for MinION data were not yet available, a combination of two different HLA analysis programs was chosen, both able to deal with data from all common NGS platforms and kits. Comparisons of the typing results with the results of Sanger full-length class I HLA sequence−based typing (SSBT) reveals that the MinION $1D^2$ sequencing protocol and redundant analysis approach was 100% concordant with the SSBT typings to third-field resolution (Supplemental Table S1).

Although most of the samples were correctly typed to three fields in both software tools immediately, some manual interpretation was necessary in a small percentage of the typing results. Because of the presence of homopolymer stretches, in 14.1% of the cases, one of the nucleotides within the homopolymer sequence was ignored in the NGSengine software, resulting in an ambiguous typing result (Supplemental Table S1). This nucleotide was, however, correctly identified with the other analysis program, resulting in a correct allele assignment and therefore a correct final HLA typing. For three alleles (1.5%), a discrepancy was observed between the allele call obtained with NGSengine and SeqPilot NGS. In these cases, SeqPilot correctly assigned the allele to two fields, but a region in the introns containing a short tandem repeat (STR) sequence was misaligned, resulting in misalignment of the exon and therefore incorrect third-field allele call. These alleles were, however, correctly typed to the third field by NGSengine, and manual inspection of the analysis details easily resolved the discrepancy. In a single sample, NGSengine was unable to assign an allele for *HLA-A* (Supplemental Table S1), whereas the same MinION sequence data in SeqPilot NGS gave the correct typing without any problems. Repeating the sample did not solve the problem, whereas other samples with the same typing did not demonstrate this problem.

Another problem observed during this initial validation was co-amplification of *HLA-Y* with *HLA-A* in samples that were positive for *HLA-A*30*, *HLA-A*31*, *HLA-A*33*, and *HLA-A*34*. Therefore, it was decided to include all pseudogenes in the program analysis to remove any off-target reads.

During initial validation, difficult positions and regions were monitored. In most cases, it concerned homopolymer or repeat sequence regions in the introns of the genes, not affecting the HLA typing result at the third field level. In these cases, the difficult region was automatically ignored in the program. In a few cases, a difficult region was present in an exon, resulting in an ambiguous typing result if this region would be ignored. However, in most cases, the second analysis program was able to analyze this region reliably and therefore the ambiguity was resolved by this second program. In some cases, the difficult region concerned a homopolymer in an exon (eg, the homopolymer C region in exon 4 of *HLA-A*, which bears seven, eight, or nine C nucleotides and is difficult to distinguish). Because this region is defining several null alleles, resolving these cases by either Sanger sequencing of exon 4 or another method that enables detection of null alleles is recommended.

Read Distribution

During the initial validation, read distribution between barcodes, loci, and alleles was compared and used to identify imbalances in amplification or sequencing, as well as to define minimum coverage values. During this process, it was noticed that the same barcodes gave consistent imbalanced sequence results between different runs and this was independent of the sample. The imbalance in read coverage between the different barcodes could not be explained by differences in amplification efficiency because equimolar pooling of the amplicons was performed before library preparation. Therefore, it is assumed that this was an effect of the library preparation, most probably due to the ligation

of the adapters. In total, 24 barcode sequences were tested and two panels of barcodes were designed to optimize balanced read coverage between different samples. Because each run consists of nine samples, the balance was optimized to be between 10% and 13% for each barcode.

With the equimolar pooling of *HLA-A*, *HLA-B*, and *HLA-C* of one sample, the read coverage between the different loci from one sample was comparable and therefore no adaptation in the pooling process was needed.

Concerning the allelic distribution, the heterozygous positions were used to determine the median percentage of the second allele being the second most abundant nucleotide at this position. In the initial validation, values for the second allele were varying between 36% and 47%, irrespective of the locus. Because all typing results were correct, the criterion of allelic distribution was set at a minimum of 35% for the second allele. In samples that were homozygous for one or more of the HLA genes typed, the software program SeqPilot showed in three cases the presence of a second allele (Supplemental Table S1), but in these cases this second allele was due to recognition of one heterozygous position in an exon, with a frequency varying between 5% and 20%, significantly less than the threshold of 35%. This heterozygous position was not recognized by the GenDx program.

During this initial validation, the read coverage per allele was assessed in the NGSengine program as well. Because the barcode panel was not yet optimized during this validation, the read coverage per allele varied enormously, from 26 to 4070 reads. Even the allele with only 26 reads was correctly typed, indicating that the minimum coverage can be rather low. A minimum coverage per allele was set conservatively at 150 reads, ensuring that typing results are supported by sufficient coverage.

Second Validation

The criteria established with the initial validation, as described above, were verified with a second validation panel. This panel consisted of 67 diagnostic samples that were sequenced for *HLA-A*, *HLA-B*, and *HLA-C* with both group-specific Sanger SBT and MinION and analyzed with SeqPilot NGS and NGSengine software packages. These samples comprised 15 different *HLA-A*, 22 different *HLA-B*, and 13 different *HLA-C* allele groups (Supplemental Table S2).

In total, 198 HLA typing results (98.5%) obtained to the third field level by MinION sequencing were found to be identical to third field level typing by Sanger sequencing, taking the software differences as identified in the initial validation into account (Supplemental Table S2). In one sample that was homozygous for *HLA-C*, the NGSengine software showed the presence of a second allele, but the proportion was below the threshold of 35%. In two cases (1%), no typing result was obtained by both programs, because of lack of amplification of one of the HLA genes. In one case (0.5%), a correct typing result was obtained by one program, whereas the other program was not able to analyze the sequences because of insufficient number of reads. The criteria of 150 reads per allele and minimal allele distribution of 35% to 65% were verified in all heterozygous samples. The read coverage in the homozygous samples always exceeded 150 reads, and typing results were evaluated in detail by two different individuals to ascertain homozygosity.

Discussion

In the current study, the MinION full-length HLA sequencing approach, with a combined analysis strategy using SeqPilot NGS and NGSengine software packages, was validated in two steps: an initial validation with 33 samples, including all HLA class I allele groups, except *HLA-B*83*, and a second validation, running 67 diagnostic samples in parallel with Sanger sequencing. All sample typing results were compared with previously described group-specific Sanger sequence—based typing approach[14] and all were verified to be correct. Validation was focused on three-field analysis, because one of the programs was not able to take the intron sequences into account in the HLA typing results and because there is still a huge lack of full-length reference allele sequences. Furthermore, the presence of homopolymeric regions in the introns also interfered with fourth-field allele assignment. Overall, this study showed that the MinION nanopore approach for high-resolution HLA typing was valid for diagnostic purposes and as such is now implemented in the routine laboratory setting for high-resolution typing of HLA class I.

The current NGS approaches for HLA typing are based on Illumina, Ion Torrent, and PacBio technologies, which each have advantages and disadvantages. One of the major advantages of MinION single-molecule sequencing is the generation of long reads that span the entire HLA gene, which allows unambiguous phasing of polymorphism across the gene. Because the HLA genes are homologous, correct separation of genes and alleles can be a challenge with small DNA pieces. However, the long reads provided by MinION can be easily separated by locus and allele, allowing accurate analysis of multiple HLA genes and alleles. Furthermore, the MinION technology is based on the sequencing of each different PCR strand by directing it through a nanopore, ensuring the sequencing of each PCR strand only once, instead of multiple times like in other approaches. This procedure ensures detection of all different variants in a sample.

Another major advantage is the low price of the equipment, making it feasible to introduce it in even small HLA laboratories with a limited number of samples. Furthermore, the small size of the MinION allows its use in virtually any laboratory without the need for enormous dedicated desk space for equipment. Turnover time,

sequencing, and allele calling using the MinION approach require a similar time frame to other NGS methods,[32] with comparable or even less hands-on time. For high-throughput laboratories, Oxford Nanopore provides the GridION and PromethION equipment, which can run 5 and 24/48 flow cells at once, respectively.

Analysis of read quality can be challenging, and it is important to choose read quality metrics that are meaningful and unambiguous. The Phred quality scores, as reported within the MinION reads, are estimates based on the confidence the neural network within the basecaller has in its interpretation of an electric signal. Because the neural network is trained on the basis of reads from known non-HLA DNA sequences, any variation between the nature of the training samples and the targeted DNA may have an influence on read quality. In addition, as with many sequencing platforms, MinION faces challenges in analyzing regions containing STRs and homopolymers, and STR/homopolymer length is regularly underestimated in the reads. The reason for this is again that the training of the neural network has been established with mostly microorganism sequences, which do not have any STR or homopolymer regions. This is different to the reason for problems with STR/homopolymers in other NGS methods, which are based on incorporation of nucleotides in the sequence. For sequencing with the MinION, adequate training of the neural network with human sequences or even HLA sequences might solve this problem and open the door to ultrahigh resolution typing of both class I and class II to the allelic resolution level.

The use of $1D^2$ MinION reads, which are generated by pairing two single-stranded reads into a single sequence, provides both advantages and challenges. Read pairing and discarding of unpaired reads reduce the amount of data by over half, which is reflected in the final read coverage. On the other hand, this $1D^2$ method enables a higher reliability in single-nucleotide polymorphism calling. Because random basecalling errors can be resolved with additional read coverage, the use of 1D sequencing seems to be feasible in the future for the described MinION approach as well. Furthermore, it is still necessary to use two separate software packages to analyze nanopore reads for obtaining an accurate HLA typing. The problem is that the current NGS programs for HLA typing analysis are either optimized for small reads (like Ion Torrent and Illumina) or optimized for PacBio single-molecule reads. The interpretation and allele calling of single-molecule nanopore reads from HLA present unique bioinformatic challenges, which are now traced and starting to be addressed by commercial companies. As soon as HLA typing analysis programs dealing specifically with the nanopore reads are developed, it will be feasible to use a single software package.

MinION is a long-read sequencing platform, and the length of a sequenced region is in fact limited by the sample preparation, and not by the sequencing platform. Read lengths of >2 MB have been reported for the MinION platform,[33] which is sufficient for sequencing the whole HLA class I or class II region. Methods like probe capturing[34] or genomic fishing might enable isolation of the major histocompatibility complex region, providing the possibility to abolish PCR amplification and even determine the arrangement of HLA haplotypes by MinION sequencing in the future.

MinION and nanopore sequencing have evolved rapidly from a low-quality sequencing platform for research purposes to the now available high standard third-generation sequencing method, which can be used to define the hyperpolymorphic HLA genes. Further development and improvement of the 1D procedure will provide higher coverage, lower hands-on time, and faster basecalling technology and therefore improve HLA typing in the near future. Overall, the present study describes the validation and implementation of the nanopore sequencing HLA class I typing method and illustrates that nanopore sequencing technology has advanced to a point where it can be used in routine diagnostics with high accuracy.

Acknowledgments

We thank Christel Meertens, Fausto Palusci, Sophie Onclin, and Stefan Molenbroeck for assistance in the validation process; and Diana van Bakel for contributions to manuscript submission.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2020.04.001*.

References

1. Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M: Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 2013, 339: 1578−1582
2. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE: The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 2015, 43:D423−D431
3. Patel R, Terasaki PI: Significance of the positive crossmatch test in kidney transplantation. N Engl J Med 1969, 280:735−739
4. Duquesnoy RJ: Human leukocyte antigen epitope antigenicity and immunogenicity. Curr Opin Organ Transplant 2014, 19:428−435
5. El-Awar N, Jucaud V, Nguyen A: HLA epitopes: the targets of monoclonal and alloantibodies defined. J Immunol Res 2017, 2017: 3406230
6. Fürst D, Müller C, Vucinic V, Bunjes D, Herr W, Gramatzki M, Schwerdtfeger R, Arnold R, Einsele H, Wulf G, Pfreundschuh M, Glass B, Schrezenmeier H, Schwarz K, Mytilineos J: High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. Blood 2013, 122:3220−3229
7. Mayor NP, Hayhurst JD, Turner TR, Szydlo RM, Shaw BE, Bultitude WP, Sayno J-R, Tavarozzi F, Latham K, Anthias C, Robinson J, Braund H, Danby R, Perry J, Wilson MC, Bloor AJ, McQuaker IG, MacKinnon S, Marks DI, Pagliuca A, Potter MN,

Potter VT, Russell NH, Thomson KJ, Madrigal JA, Marsh SGE: Recipients receiving better HLA-matched hematopoietic cell transplantation grafts, uncovered by a novel HLA typing method, have superior survival: a retrospective study. Biol Blood Marrow Transplant 2019, 25:443−450
8. Fan WL, Shiao MS, Hui RC, Su SC, Wang CW, Chang YC, Chung WH: HLA association with drug-induced adverse reactions. J Immunol Res 2017, 2017:3186328
9. Illing PT, Purcell AW, McCluskey J: The role of HLA genes in pharmacogenomics: unravelling HLA associated adverse drug reactions. Immunogenetics 2017, 69:617−630
10. Leckband SG, Kelsoe JR, Dunnenberger HM, George AL Jr, Tran E, Berger R, Muller DJ, Whirl-Carrillo M, Caudle KE, Pirmohamed M: Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. Clin Pharmacol Ther 2013, 94:324−328
11. Dendrou CA, Petersen J, Rossjohn J, Fugger L: HLA variation and disease. Nat Rev Immunol 2018, 18:325−339
12. Erlich H: HLA DNA typing: past, present, and future. Tissue Antigens 2012, 80:1−11
13. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977, 74: 5463−5467
14. Voorter CEM, Palusci F, Tilanus MGJ: Sequence-based typing of HLA: an improved group-specific full-length gene sequencing approach. Edited by Beksaç M. In Bone Marrow and Stem Cell Transplantation. New York, NY: Springer New York, 2014. pp. 101−114
15. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al: Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008, 456:53−59
16. Merriman B, Rothberg JM: Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 2012, 33:3397−3417
17. 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, 76:883−886
18. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al: Real-time DNA sequencing from single polymerase molecules. Science 2009, 323: 133−138
19. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA: The potential and challenges of nanopore sequencing. Nat Biotechnol 2008, 26:1146−1153
20. Jain M, Olsen HE, Paten B, Akeson M: The Oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 2016, 17:239
21. Lieberman KR, Cherf GM, Doody MJ, Olasagasti F, Kolodji Y, Akeson M: Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. J Am Chem Soc 2010, 132:17961−17972
22. Wick RR, Judd LM, Holt KE: Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 2019, 20:129
23. Ewing B, Green P: Base-calling of automated sequencer traces using phred, II: error probabilities. Genome Res 1998, 8:186−194
24. Montgomery MC, Liu C, Petraroia R, Weimer ET: Using nanopore whole-transcriptome sequencing for human leukocyte antigen genotyping and correlating donor human leukocyte antigen expression with flow cytometric crossmatch results. J Mol Diagn 2020, 22: 101−110
25. Matern BM, Olieslagers TI, Voorter CEM, Groeneweg M, Tilanus MGJ: Insights into the polymorphism in HLA-DRA and its evolutionary relationship with HLA haplotypes. HLA 2020, 95: 117−127
26. Duke JL, Mosbruger TL, Ferriola D, Chitnis N, Hu T, Tairis N, Margolis DJ, Monos DS: Resolving MiSeq-generated ambiguities in HLA-DPB1 typing by using the Oxford nanopore technology. J Mol Diagn 2019, 21:852−861
27. Lang K, Surendranath V, Quenzel P, Schofl G, Schmidt AH, Lange V: Full-length HLA class I genotyping with the MinION nanopore sequencer. Methods Mol Biol 2018, 1802:155−162
28. Liu C, Xiao F, Hoisington-Lopez J, Lang K, Quenzel P, Duffy B, Mitra RD: Accurate typing of human leukocyte antigen class I genes by Oxford nanopore sequencing. J Mol Diagn 2018, 20: 428−435
29. Chua EW, Ng PY:, MinION: A novel tool for predicting drug hypersensitivity? Front Pharmacol 2016, 7:156
30. Ammar R, Paton TA, Torti D, Shlien A, Bader GD: Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes. F1000Res 2015, 4:17
31. Tilanus MGJ: The power of Oxford nanopore MinION in human leukocyte antigen immunogenetics. Ann Blood 2017, 2:12
32. Mayor NP, Robinson J, McWhinnie AJM, Ranade S, Eng K, Midwinter W, Bultitude WP, Chin C-S, Bowman B, Marks P, Braund H, Madrigal JA, Latham K, Marsh SGE: HLA typing for the next generation. PLoS One 2015, 10:e0127153
33. Payne A, Holmes N, Rakyan V, Loose M: BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics 2019, 35:2193−2198
34. Dapprich J, Ferriola D, Mackiewicz K, Clark PM, Rappaport E, D'Arcy M, Sasson A, Gai X, Schug J, Kaestner KH, Monos D: The next generation of target capture technologies: large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. BMC Genomics 2016, 17:486