



# Targeted RNA-Based Oxford Nanopore Sequencing for Typing 12 Classical HLA Genes

Tiira Johansson<sup>1,2\*</sup>, Satu Koskela<sup>2†</sup>, Dawit A. Yohannes<sup>1†</sup>, Jukka Partanen<sup>2</sup> and Päivi Saavalainen<sup>1,2</sup>

<sup>1</sup> Translational Immunology Research Program and Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland, <sup>2</sup> Finnish Red Cross Blood Service, Helsinki, Finland

## OPEN ACCESS

### Edited by:

Youri I. Pavlov,  
University of Nebraska Medical  
Center, United States

### Reviewed by:

Scott Kennedy,  
University of Washington,  
United States  
Jiannis (Ioannis) Ragoussis,  
McGill University, Canada

### \*Correspondence:

Tiira Johansson  
tiira.johansson@helsinki.fi

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Genomic Assay Technology,  
a section of the journal  
Frontiers in Genetics

Received: 30 November 2020

Accepted: 11 February 2021

Published: 04 March 2021

### Citation:

Johansson T, Koskela S,  
Yohannes DA, Partanen J and  
Saavalainen P (2021) Targeted  
RNA-Based Oxford Nanopore  
Sequencing for Typing 12 Classical  
HLA Genes.  
Front. Genet. 12:635601.  
doi: 10.3389/fgene.2021.635601

Identification of human leukocyte antigen (HLA) alleles from next-generation sequencing (NGS) data is challenging because of the high polymorphism and mosaic nature of HLA genes. Owing to the complex nature of HLA genes and consequent challenges in allele assignment, Oxford Nanopore Technologies' (ONT) single-molecule sequencing technology has been of great interest due to its fitness for sequencing long reads. In addition to the read length, ONT's advantages are its portability and possibility for a rapid real-time sequencing, which enables a simultaneous data analysis. Here, we describe a targeted RNA-based method for HLA typing using ONT sequencing and SeqNext-HLA SeqPilot software (JSI Medical Systems GmbH). Twelve classical HLA genes were enriched from cDNA of 50 individuals, barcoded, pooled, and sequenced in 10 MinION R9.4 SpotON flow cell runs producing over 30,000 reads per sample. Using barcoded 2D reads, SeqPilot assigned HLA alleles to two-field typing resolution or higher with the average read depth of 1750x. Sequence analysis resulted in 99–100% accuracy at low-resolution level (one-field) and in 74–100% accuracy at high-resolution level (two-field) with the expected alleles. There are still some limitations with ONT RNA sequencing, such as noisy reads, homopolymer errors, and the lack of robust algorithms, which interfere with confident allele assignment. These issues need to be inspected carefully in the future to improve the allele call rates. Nevertheless, here we show that sequencing of multiplexed cDNA amplicon libraries on ONT MinION can produce accurate high-resolution typing results of 12 classical HLA loci. For HLA research, ONT RNA sequencing is a promising method due to its capability to sequence full-length HLA transcripts. In addition to HLA genotyping, the technique could also be applied for simultaneous expression analysis.

**Keywords:** human leukocyte antigen, HLA genotyping, nanopore sequencing, RNA sequencing, MinION

## INTRODUCTION

The human leukocyte antigen (HLA) complex on the short arm of chromosome 6 (6p21.3) is the most polymorphic region in the human genome with over 26,000 known alleles reported by the IPD IMGT/HLA database (Release 3.41.2<sup>1</sup>). It is extensively studied, and associated with various infectious and autoimmune diseases (Liu et al., 2016; Karnes et al., 2017; Hirata et al., 2019), and transplantation outcomes (Petersdorf et al., 2015). The HLA genes divided into two classes referred

<sup>1</sup><https://www.ebi.ac.uk/ipd/imgt/hla/>

to as class I (*HLA-A*, *-B*, and *-C*) and class II (*HLA-DR*, *-DQB*, and *-DP*) play a critical role in immune responses (The MHC sequencing consortium, 1999; Horton et al., 2004).

The major challenge in HLA typing is the high genetic heterogeneity, which makes allele-calling complex. Over the past years, next-generation sequencing (NGS) has taken over and replaced the more conventional HLA typing methodologies. Several different sequencing methods and data analysis tools have emerged to provide high-throughput typing of these polymorphic genes (Erlich et al., 2011; Lank et al., 2012; Shiina et al., 2012; Warren et al., 2012; Mayor et al., 2015; Buchkovich et al., 2017; Orenbuch et al., 2020). Many of the NGS technologies for HLA typing are based on the Illumina sequencing platform, which produces short reads with high sequencing accuracy and outstanding throughput. However, the limited read length has been challenging in HLA typing because of the high sequence similarity and numerous polymorphisms and mosaic structure between alleles. For that reason, sequencing platforms producing long reads, such as Oxford Nanopore Technologies (ONT) and Pacific Bio (PacBio), have raised interest. Longer reads might ease the alignment against HLA alleles and less frequently lead to ambiguous typing results. Both of these long-read sequencing platforms have already proved themselves as an option in HLA research. PacBio's Single Molecule Real-Time DNA sequencing yielded highly accurate typing results in cell lines (Turner et al., 2018) and improved the survival after hematopoietic stem cell transplantation (Mayor et al., 2019). ONT produced accurate typing results in studies using genomic amplicon sequencing (Liu et al., 2018; Ton et al., 2018; Stockton et al., 2020). ONT was also the first reported NGS HLA typing method tested for deceased donor allocation (De Santis et al., 2020). In addition to genomic sequencing data, ONT was used to genotype HLA class I loci and to determine their gene-level expression from RNA sequencing (RNA-Seq) reads (Montgomery et al., 2020).

Oxford Nanopore Technologies' MinION is a small portable sequencer connected to a computer through a USB 3.0 cable (Quick et al., 2014). It utilizes nanopore sequencing technology and has the ability to sequence both DNA and RNA with the possibility for a simultaneous real-time data analysis. MinION flow cell contains 2,048 membrane-embedded nanopores, which enable the measuring of fluctuations in the ion flow passing through the nanopores. In the past years, ONT's sequencing chemistries have developed rapidly leading to faster sequencing speeds and higher data yields.

To evaluate the potential of ONT RNA-Seq for HLA typing, we sequenced 12 HLA genes of 50 individuals in 10 multiplexed MinION sequencing runs. We developed a method, which applied a template switching oligo (TSO) in the reverse transcription (RT) of RNA molecules into cDNA. We then amplified the full-length cDNA and enriched HLA amplicons using primers specific for template switching oligo and HLA genes. We pooled the enriched amplicons, indexed them with sample-specific barcodes, and multiplexed several samples and loci together. Finally, we prepared ONT 2D sequencing libraries, sequenced them on MinION R9.4 SpotON flow cells, and performed the sequence analysis by using SeqPilot software. By

comparing the resolved HLA genotypes to alleles obtained by Luminex SSO-PCR, we assessed the accuracy of our method.

## MATERIALS AND METHODS

### Samples and RNA Extraction

Peripheral blood mononuclear cells (PBMC) of 50 healthy blood donors were isolated using Ficoll-Paque™ Plus (GE Healthcare), Dulbecco's phosphate buffered saline (DPBS) CTSTM (Gibco Life Technologies), fetal bovine serum (FBS, Sigma), and SepMate™-50 tubes according to the manufacturer's protocol (Stemcell Technologies). Isolated cell counts were determined using NucleoCounter® NC-100™ (chemometec). The use of anonymized PBMCs from blood donors is in accordance with the rules of the Finnish Supervisory Authority for Welfare and Health (Valvira). All RNA was extracted from 1 to 10 × 10<sup>6</sup> cells using RNeasy Mini kit and Rnase-Free DNase Set (both Qiagen), and the quantity was measured using The Qubit™ RNA High Sensitivity Assay Kit in Qubit® 2.0 fluorometer (ThermoScientific). The quality of RNA was assessed with an RNA 6000 Pico Kit (Agilent Genomics) in a 2100 Bioanalyzer (Agilent Genomics). Samples with RNA integrity number (RIN) value over 8.0 were used in the library preparation.

### Reverse Transcription and Amplicon Enrichment

Reverse transcription was adapted from Islam et al. (2012); 25 ng of RNA, 1% Triton™ X-100 (Sigma), 20 μM of STRT-V3-T30-VN oligo, 100 μM of DTT (Invitrogen, Life Technologies, Thermo Fisher), 10 mM dNTP (Bioline), 4 U of Recombinant RNase Inhibitor (Takara Clontech), and 1:1,000 The Ambion® ERCC RNA Spike-In Control Mix (Life Technologies, Thermo Fisher) in a total volume of 3 μl were first incubated 3 min at 72°C. 3.7 μl of RT mix with 5x SuperScript first strand buffer (Invitrogen by Thermo Fisher Scientific), 1 M MgCl<sub>2</sub> (Sigma), 5 M Betaine solution (Sigma), 134 U of SuperScript® II Reverse Transcriptase (Invitrogen by Thermo Fisher Scientific), 40 μM RNA-TSO 10 bp UMI, and 5.6 U of Recombinant RNase Inhibitor was added and the plate was incubated 90 min at 42°C and 10 min at 72°C. The double-stranded cDNA was further amplified in a 50 μl PCR reaction containing 2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems) and 10 μM ImSTRT-TSO-PCR with the following thermocycling parameters: 1 cycle of 3 min at 95°C, 20 cycles of 20 s at 95°C, 15 s 55°C, 30 s at 72°C, and 1 cycle of final elongation of 1 min at 72°C. A restriction reaction using *Sall*-HH (New England Biolabs) was used to release the 3' fragments of the cDNA. DNA was quantified using Qubit™ dsDNA High Sensitivity Assay Kit and the size distribution was checked with High Sensitivity DNA Kit (Agilent Genomics). For HLA amplicon enrichment, 3 μl of template cDNA, 10x Advantage 2 PCR buffer, 50x Advantage® 2 Polymerase Mix (Takara, Clontech), 10 mM dNTP (Bioline), 10 μM TSO forward primer, and one of the seven HLA gene-specific reverse primers (*HLA-A/-B/-C*, *-DRA*, *-DRB1/-DRB3/-DRB4/-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1*) were incubated 30 s at 98°C following 3 cycles of 10 s at 98°C, 30 s at 55°C, 30 s at 72°C and 27 cycles of 10 s at

98°C, 30 s at 71°C, 30 s at 72°C, and final elongation of 5 min at 72°C in a total volume of 15 µl. Two shared primers were used, one to amplify *HLA-A*, *-B*, and *-C* and the other for *HLA-DRB1*, *-DRB3*, *-DRB4*, and *-DRB5*. **Supplementary Figure 1** shows the binding sites of HLA gene-specific primers. All oligos were from Integrated DNA Technologies (**Supplementary Table 1**). To confirm the amplicon lengths and non-specific amplification, four samples were selected from each locus and run on a 2% agarose gel (Bioline) with 10x BlueJuice™ loading dye (Invitrogen by Thermo Fisher Scientific) in 0.5X TBE (Thermo Fisher Scientific) with the GelGreen™ (Biotium). Gels were visualized using the Quick-Load 1kb DNA Ladder (New England Biolabs). PCR amplicons were quantified using the Qubit™ dsDNA High Sensitivity Assay Kit and the fragment sizes were checked with Agilent's High Sensitivity DNA Kit. For sequencing library preparation, we divided HLA amplicons into two gene pools per sample using a 5 µl of PCR product from each locus. Gene pool 1 contained genes *HLA-A*, *-B*, *-C*, *-DRB1*, *-DRB3*, *-DRB4*, *-DRB5*, and *-DPB1*, and gene pool 2 *HLA-DRA*, *-DPA1*, *-DQA1*, and *-DQB1*. Gene pools were cleaned with 0.7X pool volume Agencourt AMPure XP beads (Beckman Coulter), eluted in 15 µl of nuclease-free water, and quantified with the Qubit™ dsDNA High Sensitivity Assay Kit. Finally, the average fragment size distribution of gene pools 1 and 2 was analyzed using Agilent's High Sensitivity DNA Kit from 10 samples of both pools. Using this information, the molarity of each pool was calculated using the DNA concentration (ng/µl) and the average fragment length (bp) of the gene pool.

## ONT Library Preparation and Sequencing

ONT's sequencing compatible barcoded fragments were prepared in a PCR reaction 0.5 nM of DNA from gene pools, 2 µl of PCR barcode from the 96 PCR Barcoding Kit (ONT), 50 µl of LongAmp Taq 2x Mix (New England Biolabs), and Nuclease-Free water in a final volume of 100 µl where ONT's universal tails were used as a template for barcode introducing primers. The PCR was performed in the following conditions: initial denaturation of 3 min at 95°C, following 15 cycles of 15 s at 95°C, 15 s at 62°C, 30 s at 65°C, and a final extension step 3 min at 65°C. A second DNA purification and size selection was done in a 1X beads:DNA ratio by using the Agencourt AMPure XP beads according to the manufacturer's instructions and eluted in 20 µl of nuclease-free water. After the purification, DNA was quantified with the Qubit™ dsDNA High Sensitivity Assay Kit and barcoded PCR amplicons were pooled with equal molarities in 10 library pools in a total volume of 50 µl each consisting of 10 individuals and either eight loci (gene pool 1) or four loci (gene pool 2). 1 µg of pooled barcoded PCR products was treated with the NEBNext Ultra II End-repair/da-tailing Module (New England Biolabs) according to a Ligation Sequencing Kit 2D (SQK-LSK208) protocol (ONT) using a DNA CS 3.6kb (ONT) as a positive control. A third DNA purification was performed using 1X beads:DNA ratio by using the Agencourt AMPure XP beads following the Ligation Sequencing Kit 2D protocol. ONT sequencing adapters were ligated using NEB Blunt/TA Ligase Master Mix (New England Biolabs) and Adapter Mix and HP Adaptor provided by ONT following a purification step

using MyOne C1 Streptavidin beads (Invitrogen by Thermo Fisher Scientific) according to the Ligation Sequencing Kit 2D protocol to capture HP adaptor containing molecules. The libraries were eluted in 25 µl of elution buffer and mixed with running buffer and library loading beads (ONT) prior to sequencing. All 10 libraries were sequenced for 48 h on R9.4 SpotON flow cells (FLO-MIN106) on MinION Mk 1b device using the MinKNOW software (versions 1.1.21, 1.3.24, 1.3.25, and 1.1.30).

## HLA Typing

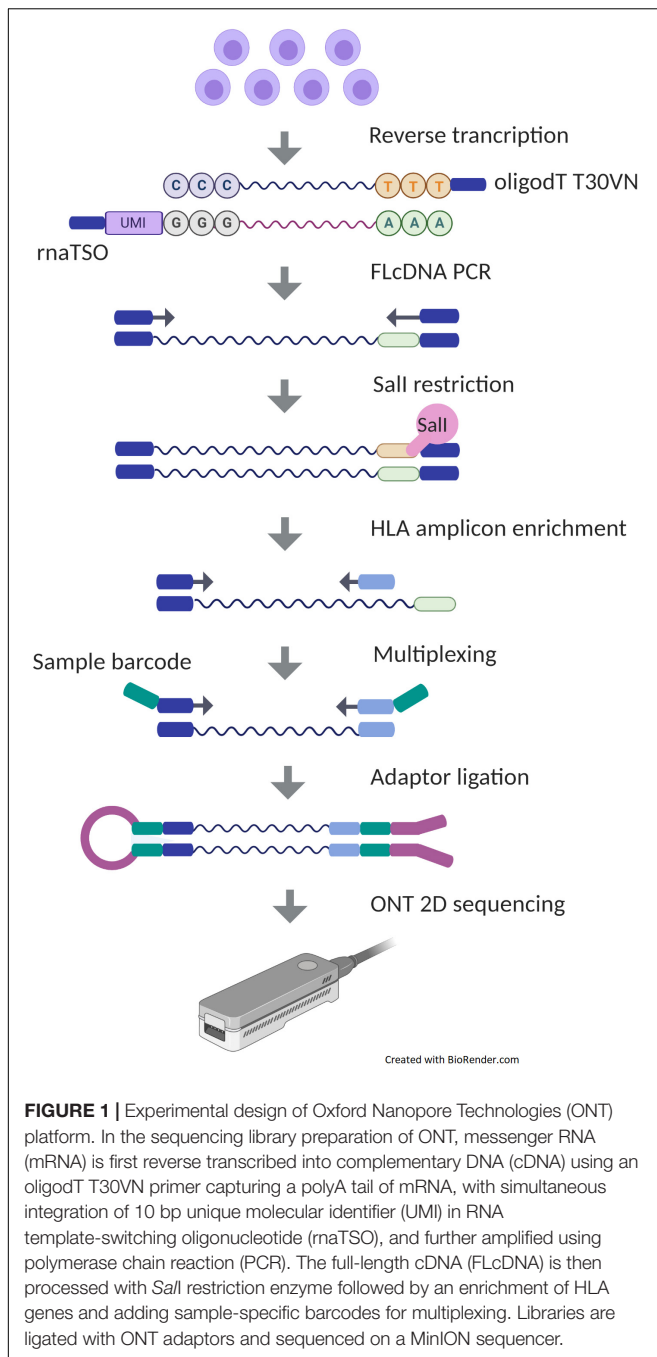
ONT's reads were processed using the 2D Basecalling plus barcoding for FLO-MIN106 250 bps workflow (version v1.125) on the cloud-based Metrichor platform (v2.45.5, v2.44.1, ONT) generating 1D template, 1D complement, and 2D reads. The fastq files were extracted from the native FAST5 files using NanoOK (Leggett et al., 2015). ONT 2D reads were used in genotyping of *HLA-A*, *-B*, *-C*, *-DRA*, *-DRB1*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1* alleles at high-resolution level with SeqPilot software (v.4.3.1, JSI Medical Systems). The software version used IPD-IMGT/HLA Database. 3.27.0. ONT 2D reads were mapped against the regions of interest (ROIs), which are the target HLA genes selected prior to the sequence analysis. For both gene pools, only those genes were selected, which were present in the gene pool. After the sequence analysis, SeqPilot software reported the genotyping result for each gene of the gene pool. Additionally, SeqPilot reported the number of assigned reads, which were the reads mapping to the selected ROIs, and the number of aligned reads, which were the reads passing the quality filter and added to the coverage. For validation of HLA typing, genomic DNA was extracted from PBMC samples by QiaSymphony SP automat (Qiagen) by following the manufacturer's protocol. The concentration and purity of DNA were measured by NanoDrop ND-1000 (Thermo Fisher Scientific). The reverse SSOP-Luminex technology (Labtype, One Lambda) was used for HLA typing of *HLA-A*, *-B*, *-C*, *-DRB1*, *-DRB3*, *-DRB4*, *-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1*. The results were analyzed with the HLA-Fusion software (v.3.2.0-HF1, One Lambda). The concordance rate of the genotyping results between ONT RNA-Seq and Luminex SSO-PCR data was calculated at one-field (allele group) and two-field (allele) resolution level. Luminex SSO-PCR reports the HLA type at one-field resolution with a list of possible alleles as NMDP codes by the National Marrow Donor Program<sup>2</sup>. For two-field resolution level comparison, we assigned the putative HLA alleles from the allele codes based on the alleles' prevalence in the Finnish population using a Finnish cohort study (Haimila et al., 2013). The HLA alleles present in our dataset of 50 individuals are shown in **Supplementary Table 2**.

## RESULTS

### ONT Sequencing Statistics

We assessed the suitability of ONT RNA-Seq for HLA typing. To do that, we used a targeted 5' RNA-Seq method for ONT platform

<sup>2</sup><https://bioinformatics.bethematchclinical.org>



(Figure 1). Simultaneously with the ONT sequencing, cloud-based Metrichor basecalled the raw data and demultiplexed reads per sample-specific barcodes. Metrichor provided both the sequencing statistics and the basecalling results. In the 48-h sequencing runs, 46 h was the maximum time that flow cells were still producing data. However, the majority of the sequence data were generated during the first 8 h. Sequencing of 100 gene pools on 10 ONT MinION SpotON flow cells generated between 97,595 and 855,489 successfully demultiplexed barcoded reads and yielded between 76.33 and 521.45 Mbases per run (Figures 2A,B).

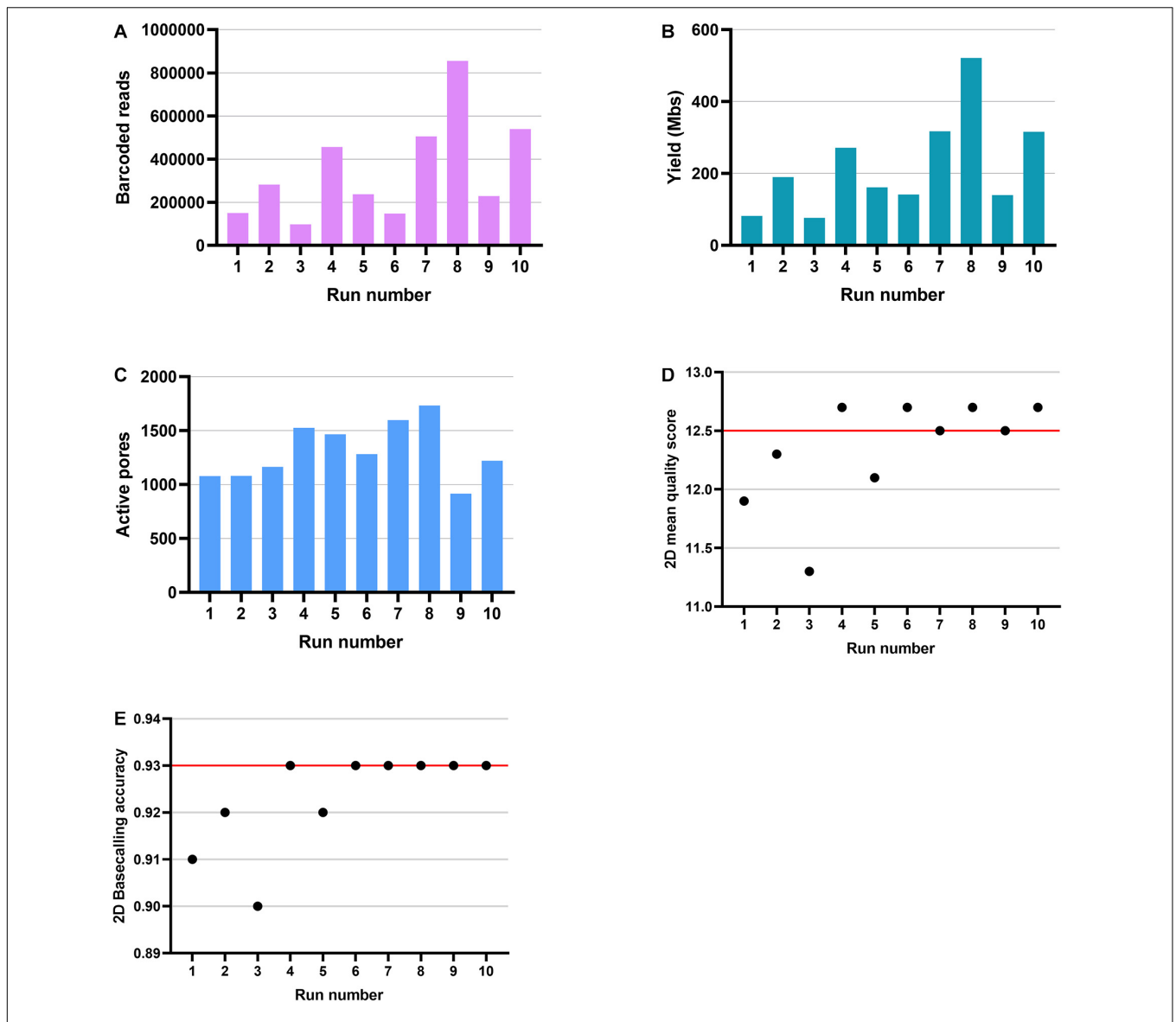
An average of 3.4% of reads remained unmultiplexed. All R9.4 flow cells exceeded the guaranteed number of 800 active pores (Figure 2C). Runs produced similar distributions of 2D mean quality scores with a mean of 12.3 (Figure 2D). To assess the quality of basecalling, we used a calibration strand provided in ONT's Ligation Sequencing Kit 2D. Based on the reads derived from this Lambda genome, the average of 2D basecalling accuracy in 10 runs was 0.92 (Figure 2E). Due to the variation in read lengths, the average read lengths between 520 and 620 bp reported by Metrichor were shorter than the expected amplicon lengths. However, in the sequence length distribution, there were peaks close to the expected amplicon sizes at 800–900 bp for gene pool 1 and at 750–850 bp for gene pool 2.

### Accuracy of ONT RNA-Seq

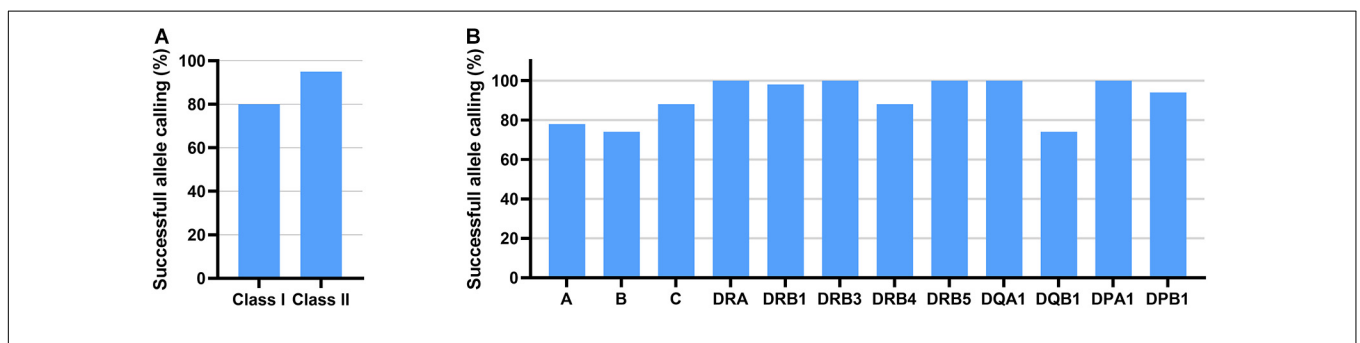
Oxford Nanopore Technologies' 2D reads underwent a fastq conversion from FAST5 files with NanoOK tool (Leggett et al., 2015). Extracted fastq files were uploaded to SeqPilot software and aligned against HLA genes selected prior to data analysis. Out of 3,610,894 reads with uniquely identified barcodes, 1,354,704 reads were mapped to ROIs in the SeqPilot software. However, most of the reads uploaded to the software, 62.5%, did not map to the ROIs and remained unmapped. Based on the information received from JSI Medical Systems, this was the first time SeqPilot was used to analyze ONT RNA-Seq data. At the time of the sequence analysis, the software was designed for the analysis of genomic data. This could have possibly contributed to the low mappability of RNA sequencing reads. Using ONT RNA-Seq data, SeqPilot was able to assign 80% of HLA class I alleles and 95% of HLA class II alleles at two-field resolution or higher (Figure 3A and Table 1). HLA typing calls were obtained in 78, 74, and 88% of the cases for *HLA-A*, *HLA-B*, and *HLA-C* and in 100, 98, 100, 88, 100, 100, 74, 100, and 94% of the cases, for *HLA-DRA*, *-DRB1*, *-DRB3*, *-DRB4*, *-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1* (Figure 3B and Table 1).

Sequence data analysis in SeqPilot interphase showed an uneven distribution of sequencing coverage for class I genes (Supplementary Figure 2). The coverage graph displayed a lower coverage in exon 1 and 2, and sometimes also in exon 3 area compared to the other exons. For class II, the read coverage was even in all exons. Most of the polymorphisms between HLA alleles lie in exons 2 and 3. Therefore, the low coverage could affect the allele assignment and explain the lower call rates of SeqPilot for class I.

In addition to the coverage graph, we inspected the number of assigned reads mapping to the selected genes. We also calculated the number of aligned reads, which are the reads passing the filter and used in the allele assignment. Table 2 shows the average and range of assigned and aligned reads calculated for each locus. The highest relative ratio between aligned and assigned reads was found in *HLA-DRA*, *-DRB1*, *-DQA1*, and *-DPA1* (65–87%), and the lowest in *HLA-DQB1* (22%). In 20% of the cases for class I and in 5% for class II, SeqPilot could not make a confident call. In 76% of these cases for class I and 63% for class II, the number of aligned reads reported by SeqPilot was below the locus average. This suggests that a low sequencing coverage may interfere with



**FIGURE 2 |** Produced data and quality metrics of ONT MinION sequencing runs. Ten MinION flow cell produced between 97,595 and 855,489 barcoded reads (A) and yielded between 76.33 and 521.45 Mbases (B). The number of active pores identified in each run ranged from 915 to 1,732 (C). The distribution for 2D mean quality score was 11.3–12.7 (D) and 0.90–0.93 for 2D basecalling accuracy (E). The red lines indicate the median.



**FIGURE 3 |** HLA typing results of 50 individuals using ONT RNAseq data and SeqPilot software. Successful allele calling rate is shown for class I and class II (A) and for each gene (B).

**TABLE 1** | HLA genotyping results from ONT RNA-Seq and allele calling rate.

	Result	No result	Accurate allele calling (%)
HLA-A	39	11	78
HLA-B	37	13	74
HLA-C	44	6	88
HLA-DRA	50	0	100
HLA-DRB1	49	1	98
HLA-DRB3*	25	0	100
HLA-DRB4*	15	2	88
HLA-DRB5*	19	0	100
HLA-DQA1	50	0	100
HLA-DQB1	37	13	74
HLA-DPA1	50	0	100
HLA-DPB1	47	3	94

\*HLA-DRB3-5 genes are present only in certain haplotypes.

**TABLE 2** | Summary of assigned and aligned reads per HLA gene from the SeqPilot software.

	Average of assigned reads	Range of assigned reads	Average of aligned reads	Range of aligned reads	Aligned reads (%)
HLA-A	1,271	351–3,551	581	132–1,607	46
HLA-B	1,809	466–6,486	1,013	221–4,156	56
HLA-C	1,501	338–4,516	862	81–2,756	57
HLA-DRA	4,604	1,042–13,897	3,995	873–12,235	87
HLA-DRB1	3,677	1,098–15,607	2,394	672–12,050	65
HLA-DRB3	496	141–1,693	267	44–1,037	54
HLA-DRB4	741	79–2,961	402	32–1,872	54
HLA-DRB5	1,265	132–2,934	747	25–1,815	59
HLA-DQA1	905	179–2,690	692	85–2,263	77
HLA-DQB1	6,734	736–16,924	1,480	119–4,842	22
HLA-DPA1	4,416	803–16,547	3,769	652–14,479	85
HLA-DPB1	1,939	500–7,067	946	222–4,095	49

successful allele assignment. However, this seems to be true only in some cases since only 32 *HLA-DRB4* reads were sufficient to produce an accurate typing result at three-field resolution.

### Concordance With Luminex SSO-PCR

Human leukocyte antigen typing results obtained from ONT RNA-Seq data were compared to Luminex SSO-PCR genotyping results and the concordance was calculated both at one-field and two-field resolution. Luminex SSO-PCR reports HLA typing result at the allotype level (one-field) with a list of suggested alleles that alter the amino acid sequence (two-field). The first allele in the subtype list is the most common and well-documented one. However, the rare allele subtypes cannot be excluded from the analysis.

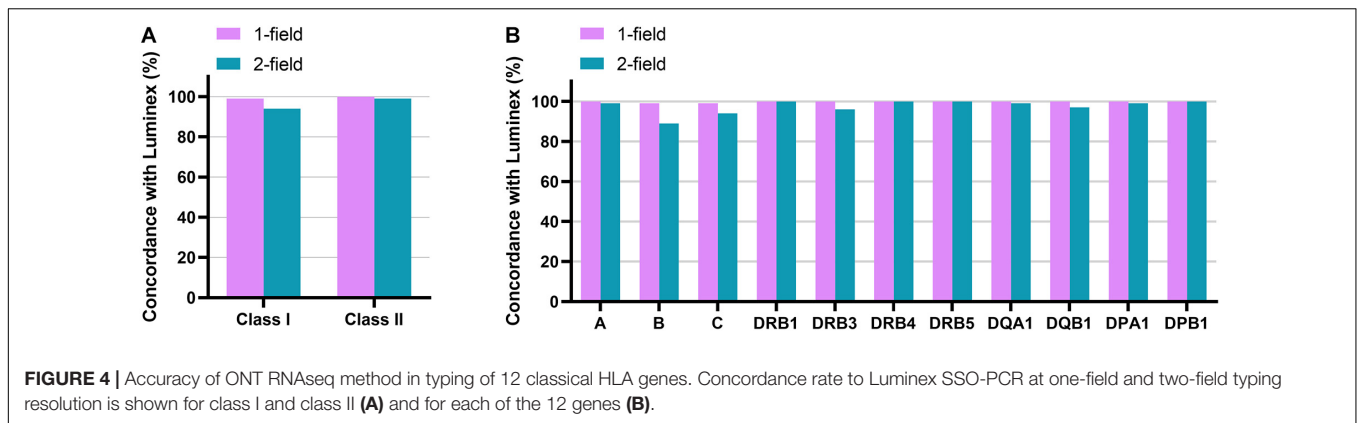
ONT RNA-Seq method had an average concordance of 99% (one-field) and 94% (two-field) for class I genes and 100% (one-field) and 99% (two-field) for class II genes with Luminex SSO-PCR results (Figure 4A and Table 3). At one-field level, the results were 100, 99, and 99% concordant for *HLA-A*, *HLA-B*, and *HLA-C* and 100% concordant for *HLA-DRB1*, *-DRB3*, *-DRB4*,

*-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1*. A missing second allele in ONT RNA-Seq data caused two one-field resolution discrepancies in one sample for *HLA-B* and *HLA-C* between the two methods. Luminex SSO-PCR reported the typing results as *B\*07:02*, *B\*35:01* and *C\*04:01*, *C\*07:02*. ONT RNA-Seq, however, could only call alleles *B\*07:02* and *C\*04:01* and reported homozygous results. The number of aligned reads was below the average for both genes. At two-field level, the concordance between ONT RNA-Seq and Luminex SSO-PCR was 99, 89, and 94% for *HLA-A*, *HLA-B*, and *HLA-C* and 100, 96, 100, 100, 99, 97, 99, and 100% for *HLA-DRB1*, *-DRB3*, *-DRB4*, *-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1* (Figure 4B and Table 3). The discrepancy at two-field resolution originated from an allele reported by SeqPilot, which was not present in the list of suggested alleles by Luminex SSO-PCR. The numbers of detected discrepancies were one for *HLA-A*, *HLA-DQA1*, and *HLA-DPA1*, two for *HLA-DQB1*, four for *HLA-C*, and seven for *HLA-B*. Out of these 16 discrepant samples, 11 (69%) had a lower number of aligned reads compared to the gene average. There were no discrepancies in *HLA-DRB1*, *-DRB4*, *-DRB5*, and *DPB1*.

For class I, there were eight ambiguities in the SeqPilot results at three-field and one at two-field typing resolution. The only ambiguous result at two-field level was in *HLA-B*. In this case, SeqPilot could not make a confident call between *B\*15:01:01G*, *\*35:01:01G* and *B\*15:20*, *\*35:43:01*. For class II, the sequence analysis resulted in 19 ambiguities at two-field resolution; one for *HLA-DRB5* and *HLA-DQB1*, seven for *HLA-DQA1*, and 10 for *HLA-DPB1*. In nine cases (47%) with ambiguous result, all alleles listed by SeqPilot software were also found in the list of alleles of Luminex SSO-PCR. However, in 10 cases (53%) with ambiguity, SeqPilot reported at least one allele, which Luminex SSO-PCR did not have in its list. In all 15 samples carrying *HLA-DRB4* gene, the possibility of a null allele could not be excluded. There were no ambiguities in *HLA-DRA*, *-DRB1*, *-DRB3*, and *-DPA1*.

## DISCUSSION

There are several potential benefits, which ONT offers to HLA research. The long-read single-molecule sequencing technology enables the sequencing of full-length HLA transcripts increasing the possibility of unambiguous phasing in the data analysis step and provides information of the full haplotype block structure. In contrast to short sequencing reads, with ONT's long reads, smaller sequencing depth may suffice to produce accurate HLA typing results. This enables the multiplexing of several samples making the method cost-effective to use. Additionally, ONT's portable desktop sequencer, MinION, can be easily adopted in basic laboratory without the need to use sequencing core facilities. Furthermore, the possibility for real-time data analysis during the sequencing may be beneficial for cases where faster HLA typing results are required. An additional benefit could be simultaneous HLA typing and expression determination. Differential expression levels of HLA have been shown to affect numerous human diseases (Apps et al., 2013; Sillé et al., 2013; Schuster et al., 2017; Ramsuran et al., 2018) and to have an impact to the outcome of hematopoietic stem cell transplantation



**TABLE 3 |** Concordance of HLA typing results between ONT RNA-Seq and Luminox SSO-PCR.

	Concordant results		Discordant results		Concordance rate %	
	One-field	Two-field	One-field	Two-field	One-field	Two-field
HLA-A	78	77	0	1	100	99
HLA-B	73	66	1	8	99	89
HLA-C	87	83	1	5	99	94
HLA-DRB1	98	98	0	0	100	100
HLA-DRB3*	25	24	0	1	100	96
HLA-DRB4*	15	15	0	0	100	100
HLA-DRB5*	19	19	0	0	100	100
HLA-DQA1	100	99	0	1	100	99
HLA-DQB1	74	72	0	2	100	97
HLA-DPA1	100	99	0	1	100	99
HLA-DPB1	94	94	0	0	100	100

\*For HLA-DRB3-5, results are presented at the sample level.

(HSCT) (Petersdorf et al., 2014, 2020). However, despite of these findings, expression analysis is not included in donor selection. With the UMIs already incorporated in our library preparation process, ONT RNA-Seq method could in the future be expanded to enable parallel HLA genotyping and accurate expression quantification using molecule counting.

The main goal of this study was to evaluate the suitability of ONT RNA-Seq for HLA typing. The method included transcription of 25 ng of RNA into cDNA using a template switching oligo, incorporation of a 5' end 10 bp UMI, followed by HLA amplicon enrichment using 5' end primer specific for template switching sequence and seven 3' end HLA gene specific primers. To make the approach cost effective, HLA amplicons were divided into two gene pools, and indexed with unique sample-specific barcode sequences. Final steps of the workflow were multiplexing of gene pools of 10 individuals, library preparation using the 2D sequencing chemistry, and sequencing of gene pools on the MinION R9.4 SpotON flow cell. The method enabled sequencing of 10 individuals and eight or four HLA genes, depending on the gene pool, together in a single sequencing run.

The sequence analysis was conducted with SeqPilot software, which has previously been utilized also in the genotyping of HLA-B alleles in genomic amplicon data (Ton et al., 2018). The

analysis consisted of two main steps. First, SeqPilot aligned 2D reads against selected HLA genes, which are called ROIs. SeqPilot calculated the number of reads aligning to these selected genes. Second, it performed a quality filtering to the aligned reads. The ones, which passed this step, were used in the allele assignment. After the sequence analysis, SeqPilot reported both the number of aligned reads and the number of assigned reads passing the quality filter. Using ONT RNA-Seq data, SeqPilot software aligned in average of 30,000 reads per individual to the ROIs of both gene pools. Of these reads, approximately 17,000 were good-quality reads, which passed the filter and took part in the allele assignment. At the time of this study, HLA typing software SeqPilot had been validated only for genomic data. Therefore, our study was the first one to use SeqPilot in the analysis of ONT RNA-Seq data. From ONT RNA-Seq reads, SeqPilot successfully called 80% of class I alleles and 95% of class II alleles at two-field resolution or higher. The gene-level allele calling rates were 78, 74, and 88% for HLA-A, HLA-B, and HLA-C. The low accuracy in these genes might have been caused by the low sequencing depth in exons 1, 2, and 3 we saw in the SeqPilot interphase (Supplementary Figure 1). We are uncertain of what could have caused the lower sequencing depth in this specific area. These are, however, the first three exons of class I genes. For this reason, one possible explanation could be an inadequate efficacy of the

RT enzyme to transcribe longer class I molecules. Since most of the allele distinguishing polymorphisms in HLA genes are located in exons 2 and 3, the coverage of these polymorphic positions might not have been sufficient in our ONT RNA-Seq data. This most likely has led to the poorer sensitivity in allele assignment in class I compared to class II genes.

In the majority of cases where SeqPilot could not make a confident call in HLA typing, the number of reads was lower than the average read count with successful calls. Other studies using ONT genomic data have also shown that enough coverage is required for accurate genotyping results (Liu et al., 2018; De Santis et al., 2020; Stockton et al., 2020). However, in some cases, very few reads were sufficient to produce accurate genotyping result. This indicated that there were also other factors affecting the accurate HLA typing such as the sequence read quality and basecalling accuracy. In the SeqPilot interphase, most of the mismatches to the reference allele were caused by a background noise, deletions, or homopolymer regions. With R9.0 MinION flow cells, the rates of miscalls, insertions, and deletions were 6.2, 3.1, and 5.7%, respectively (Jain et al., 2017). Since our study, ONT has launched two new flow cells (R10 and R10.3) and replaced the 2D sequencing chemistry with the 1D<sup>2</sup> chemistry<sup>3</sup>, which no longer uses the hairpin to physically connect the template and complement strands. However, similarly to ONT's 2D chemistry, in the 1D<sup>2</sup> chemistry, both strands of the DNA duplex are sequenced consecutively to obtain an accurate consensus sequence. According to ONT, the quality of 1D<sup>2</sup> and the number of reads produced should be higher compared to the 2D chemistry. These improvements might provide a solution to the error rate and homopolymers leading to higher allele calling rates and fewer ambiguities in ONT RNA-Seq data.

In the comparison between ONT RNA-Seq and Luminex SSO-PCR, the average concordance at two-field typing resolution was 94% for class I genes and 99% for class II genes. Our results showed that *HLA-B* in ONT RNA-Seq was particularly difficult for SeqPilot to interpret. In all cases with discrepancy in *HLA-B* between the two methods, SeqPilot suggested a more uncommon allele option. The discrepancies were focused to allotypes *B\*07* and *B\*35*. Compared to the previous studies using MinION in the sequencing of *HLA-B* (Ton et al., 2018; De Santis et al., 2020), our ONT RNA-Seq method did reach as high concordance with the validation method. This might be due to methodological differences or differences in data analysis since both of the previous studies were based on the analysis of genomic data. Additionally, the low sequencing coverage in exons 1, 2, and 3 most likely decreased the class I typing accuracy and hence further affected the concordance rate.

While SeqPilot's performance in allele calling was promising, the software was not optimized for ONT RNA-Seq data at the time of this study. Several HLA typing algorithms for Illumina RNA-Seq already exist, such as seq2HLA (Boegel et al., 2012), HLAProfiler (Buchkovich et al., 2017), and arcasHLA (Orenbuch et al., 2020). However, the HLA typing tool set for ONT RNA-Seq

has remained very limited, and additional HLA typing software dedicated for MinION data would be beneficial. For genomic ONT amplicon data, some HLA genotyping algorithms have already been developed (Liu et al., 2018; Schöfl et al., 2018). However, they have been limited to only certain HLA genes. In addition to these academic algorithms, another commercial HLA typing software, NGSengine (GenDX), for ONT genomic data has been validated recently (van Deutekom et al., 2017). Despite the very promising results obtained with these software, the high error rate of 10–15% (Jain et al., 2017, 2018) occurring in ONT reads and the homopolymer issue (Liu, 2020) are still hampering accurate allele assignment and preventing ONT's routine applicability for clinical HLA typing. ONT is still a relatively novel technology in HLA research, and the constantly changing sequencing chemistries and high error rates might have hindered the development of comprehensive HLA typing tools. Our study was the second reported work to perform HLA typing from ONT RNA-Seq data and the first where both class I and II genes were included. With the growing interest toward ONT's long read technology and ONT's ability to perform accurate HLA typing, we expect the number of bioinformatics tools to grow in the future.

Oxford Nanopore Technologies has already provided direct sequencing from RNA without the need for any amplification steps included in the sequencing library preparation<sup>4</sup>. This is definitely an interesting option and may enable simultaneous HLA typing and HLA expression quantification. With increasing data yields, it might be possible to determine accurately HLA expression at both gene and allele levels. With no additional amplification steps in the library preparation protocol, the workflow would accelerate and eliminate possible PCR bias. While there are many advantages for ONT to be utilized in HLA research, several challenges such as the high error rate and the lack of optimized HLA typing software for ONT data need to be addressed. In addition to the updates in the sequencing chemistry, ONT's improved nanopore technology of R10.3 flow cells may further lower error rates and result in higher data yields in the future. With improved read accuracy and the homopolymer issue solved, we can also expect more accurate HLA allele calling in ONT RNA-Seq.

In summary, we developed a highly multiplexed RNA-based assay for ONT that enabled genotyping of 12 classical HLA genes of 50 individuals in 10 MinION runs. With the available and future developments in the flow cells and associated sequencing chemistries, we expect the accuracy of the ONT RNA-Seq to improve further. Our method is applicable to the typing of 12 HLA loci in various tissues. With the higher read counts, it may also be used to multiplex even more samples or genes together in a single MinION sequencing run.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

<sup>3</sup><https://nanoporetech.com/about-us/news/1d-squared-kit-available-store-boost-accuracy-simple-prep>

<sup>4</sup><https://nanoporetech.com/applications/rna-sequencing>



and accession number(s) can be found below: The European Genome-phenome Archive (EGA), EGAS00001004918.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Finnish National Supervisory Authority for Welfare and Health. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

TJ, PS, and JP designed the study and interpreted the results. TJ managed all the samples, prepared the sequencing libraries, performed the sequencing runs, and wrote the manuscript. SK managed Luminex HLA typing. DY extracted the ONT fastq data. TJ analyzed and interpreted the data. All authors read and approved the final manuscript.

## FUNDING

This work was supported by grants from Clinical Research Funding (EVO/VTR), the Academy of Finland, Business Finland (the Finnish Funding Agency for Technology and Innovation),

## REFERENCES

- Apps, R., Qi, Y., Carlson, J. M., Chen, H., Gao, X., Thomas, R., et al. (2013). Influence of HLA-C expression level on HIV control. *Science* 340, 87–91. doi: 10.1126/science.1232685
- Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boisguérin, V., et al. (2012). HLA typing from RNA-seq sequence reads. *Genome Med.* 4:102. doi: 10.1186/gm403
- Buchkovich, M. L., Brown, C. C., Robasky, K., Chai, S., Westfall, S., Vincent, B. G., et al. (2017). HLAProfiler utilizes k-mer profiles to improve HLA calling accuracy for rare and common alleles in RNA-seq data. *Genome Med.* 9, 1–15. doi: 10.1186/s13073-017-0473-6
- De Santis, D., Truong, L., Martinez, P., and D'Orsogna, L. (2020). Rapid high-resolution HLA genotyping by MinION Oxford nanopore sequencing for deceased donor organ allocation. *HLA* 96, 141–162. doi: 10.1111/tan.13901
- Erlich, R. L., Jia, X., Anderson, S., Banks, E., Gao, X., Carrington, M., et al. (2011). Next-generation sequencing for HLA typing of class I loci. *BMC Genomics* 12:42. doi: 10.1186/1471-2164-12-42
- Haimila, K., Peräsaari, J., Linjama, T., Koskela, S., Saarinen, T., Lauronen, J., et al. (2013). HLA antigen, allele and haplotype frequencies and their use in virtual panel reactive antigen calculations in the Finnish population. *Tissue Antigens* 81, 35–43. doi: 10.1111/tan.12036
- Hirata, J., Hosomichi, K., Sakaue, S., Kanai, M., Nakaoka, H., Ishigaki, K., et al. (2019). Genetic and phenotypic landscape of the major histocompatibility complex region in the Japanese population. *Nat. Genet.* 51, 470–480. doi: 10.1038/s41588-018-0336-0
- Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., et al. (2004). Gene map of the extended human MHC. *Nat. Rev. Genet.* 5, 889–899. doi: 10.1038/nrg1489
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P., et al. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protoc.* 7, 813–828. doi: 10.1038/nprot.2012.022

Sigrid Jusélius Foundation, Väre Foundation for Pediatric Cancer Research, and The Finnish Concordia Fund.

## ACKNOWLEDGMENTS

We would like to thank Hanne Ahola and Sisko Lehmonen for their excellent technical assistance and personnel at Biomedicum Functional Genomics Unit (FuGU) and at national HLA laboratory at the Finnish Red Cross Blood Service. We would also like to thank Sabine Gronert-Sum for introduction and technical assistance with SeqPilot software, and CSC – IT Center for Science, Finland, for computational resources.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.635601/full#supplementary-material>

**Supplementary Figure 1** | Binding sites of HLA gene-specific primers.

**Supplementary Figure 2** | Sequencing coverage graph for class I and class II in the SeqPilot interphase.

**Supplementary Table 1** | List of primer sequences.

**Supplementary Table 2** | Alleles of 12 classical HLA genes of 50 individuals.

- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060
- Jain, M., Tyson, J. R., Loose, M., Ip, C. L. C., Eccles, D. A., O'Grady, J., et al. (2017). MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry. *F1000Research* 6:760. doi: 10.12688/f1000research.11354.1
- Karnes, J. H., Bastarache, L., Shaffer, C. M., Gaudieri, S., Xu, Y., Glazer, A. M., et al. (2017). Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci. Transl. Med.* 9:eaai8708. doi: 10.1126/scitranslmed.aai8708
- Lank, S. M., Golbach, B. A., Creager, H. M., Wiseman, R. W., Keskin, D. B., Reinherz, E. L., et al. (2012). Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing. *BMC Genomics* 13:378. doi: 10.1186/1471-2164-13-378
- Leggett, R. M., Heavens, D., Caccamo, M., Clark, M. D., and Davey, R. P. (2015). NanoOK: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics* 32, 142–144. doi: 10.1093/bioinformatics/btv540
- Liu, C. (2020). A long road/read to rapid high-resolution HLA typing: the nanopore perspective. *Hum. Immunol.* doi: 10.1016/j.humimm.2020.04.009
- Liu, C., Xiao, F., Hoisington-Lopez, J., Lang, K., Quenzel, P., Duffy, B., et al. (2018). Accurate typing of human leukocyte antigen class I genes by Oxford nanopore sequencing. *J. Mol. Diagn.* 20, 428–435. doi: 10.1016/j.jmoldx.2018.02.006
- Liu, J., Ye, Z., Mayer, J. G., Hoch, B. A., Green, C., Rolak, L., et al. (2016). Phenome-wide association study maps new diseases to the human major histocompatibility complex region. *J. Med. Genet.* 53, 681–689. doi: 10.1136/jmedgenet-2016-103867
- Mayor, N. P., Hayhurst, J. D., Turner, T. R., Szydlo, R. M., Shaw, B. E., Bultitude, W. P., et al. (2019). Recipients receiving better HLA-matched hematopoietic cell transplantation grafts, uncovered by a novel HLA typing method, have superior

- survival: a retrospective study. *Biol. Blood Marrow Transplant.* 25, 443–450. doi: 10.1016/j.bbmt.2018.12.768
- Mayor, N. P., Robinson, J., McWhinnie, A. J. M., Ranade, S., Eng, K., Midwinter, W., et al. (2015). HLA typing for the next generation. *PLoS One* 10:e0127153. doi: 10.1371/journal.pone.0127153
- Montgomery, M. C., Liu, C., Petrarora, R., and Weimer, E. T. (2020). 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.* 22, 101–110. doi: 10.1016/j.jmoldx.2019.09.005
- Orenbuch, R., Filip, I., Comito, D., Shaman, J., Pe'er, I., Rabadan, R., et al. (2020). ArcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* 36, 33–40. doi: 10.1093/bioinformatics/btz474
- Petersdorf, E. W., Bengtsson, M., De Santis, D., Dubois, V., Fleischhauer, K., Gooley, T., et al. (2020). Role of HLA-DP expression in graft-versus-host disease after unrelated donor transplantation. *J. Clin. Oncol.* 38, 2712–2718. doi: 10.1200/JCO.20.00265
- Petersdorf, E. W., Gooley, T. A., Malkki, M., Bacigalupo, A. P., Cesbron, A., Du Toit, E., et al. (2014). HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation. *Blood* 124, 3996–4003. doi: 10.1182/blood-2014-09
- Petersdorf, E. W., Malkki, M., O'hugin, C., Carrington, M., Gooley, T., Haagenson, M. D., et al. (2015). High HLA-DP expression and graft-versus-host disease. *N. Engl. J. Med.* 373, 599–609. doi: 10.1056/NEJMoa1500140
- Quick, J., Quinlan, A. R., and Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience* 3:22. doi: 10.1186/2047-217X-3-22
- Ramsuran, V., Naranbhai, V., Horowitz, A., Qi, Y., Martin, M. P., Yuki, Y., et al. (2018). Elevated HLA-A expression impairs HIV control through inhibition of NKG2A-expressing cells. *Science* 359, 86–90.
- Schöfl, G., Klasberg, S., Surendranath, V., Putke, K., Schmidt, H. A., and Lange, V. (2018). P0184. typing in the third generation: a HLA typing approach for nanopore sequencing data. *Hum. Immunol.* 80:116. doi: 10.1016/j.humimm.2019.07.137
- Schuster, H., Peper, J. K., Bösmüller, H. C., Röhle, K., Backert, L., Bilich, T., et al. (2017). The immunopeptidomic landscape of ovarian carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9942–E9951. doi: 10.1073/pnas.1707658114
- Shiina, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari, A., et al. (2012). Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens* 80, 305–316. doi: 10.1111/j.1399-0039.2012.01941.x
- Sillé, F., Conde, L., Zhang, J., Akers, N., Sanchez, S., Maltbaek, J., et al. (2013). Follicular lymphoma-protective HLA class II variants correlate with increased HLA-DQB1 protein expression. *Genes Immun.* 15, 133–136. doi: 10.1038/gene.2013.64
- Stockton, J. D., Nieto, T., Wroe, E., Poles, A., Inston, N., Briggs, D., et al. (2020). Rapid, highly accurate and cost-effective open-source simultaneous complete HLA typing and phasing of class I and II alleles using nanopore sequencing. *HLA* 96, 163–178. doi: 10.1111/tan.13926
- The MHC sequencing consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401, 921–923. doi: 10.1038/44853
- Ton, K. N. T., Cree, S. L., Gronert-Sum, S. J., Merriman, T. R., Stamp, L. K., and Kennedy, M. A. (2018). Multiplexed nanopore sequencing of HLA-B locus in Māori and Pacific Island samples. *Front. Genet.* 9:152. doi: 10.3389/fgene.2018.00152
- Turner, T. R., Hayhurst, J. D., Hayward, D. R., Bultitude, W. P., Barker, D. J., Robinson, J., et al. (2018). Single molecule real-time DNA sequencing of HLA genes at ultra-high resolution from 126 international HLA and immunogenetics workshop cell lines. *HLA* 91, 88–101. doi: 10.1111/tan.13184
- van Deutekom, H. W., Kooter, R., Geerligts, J., Ruzius, F.-P., Meulenber, P., Surendranath, V., et al. (2017). P177 NGS typing results using oxford nanopore sequencing: can minion data be reliably used for HLA typing? *Hum. Immunol.* 78:190. doi: 10.1016/j.humimm.2017.06.237
- Warren, R. L., Choe, G., Freeman, D. J., Castellarin, M., Munro, S., Moore, R., et al. (2012). Derivation of HLA types from shotgun sequence datasets. *Genome Med.* 4:95. doi: 10.1186/gm396

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Johansson, Koskela, Yohannes, Partanen and Saavalainen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.