Rapid, highly accurate and cost-effective open-source simultaneous complete HLA typing and phasing of class I and II alleles using nanopore sequencing

Stockton, Joanne D; Nieto, Thomas; Wroe, Elizabeth; Poles, Anthony; Inston, Nicholas; Briggs, David; Beggs, Andrew D

DOI: 10.1111/tan.13926

License: Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Stockton, JD, Nieto, T, Wroe, E, Poles, A, Inston, N, Briggs, D & Beggs, AD 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, vol. 96, no. 2, pp. 163-178. https://doi.org/10.1111/tan.13926

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
Rapid, highly accurate and cost-effective open-source simultaneous complete HLA typing and phasing of class I and II alleles using nanopore sequencing

Joanne D. Stockton¹ | Thomas Nieto¹ | Elizabeth Wroe² | Anthony Poles² | Nicholas Inston³ | David Briggs² | Andrew D. Beggs¹

¹Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK
²NHS Blood and Transplant, Filton, UK
³Queen Elizabeth Hospital Birmingham, Birmingham, UK

Correspondence
Andrew D. Beggs, Surgical Research Laboratory, Institute of Cancer and Genomic Science, University of Birmingham, Vincent Drive, Birmingham B15 2TT, UK.
Email: a.beggs@bham.ac.uk

Funding information
Cancer Research UK, Grant/Award Number: C31641/A23923; Queen Elizabeth Hospital Charity

Accurate rapid genotyping of the genes within the HLA region presents many difficulties because of the complexity of this region. Here we present the results of our proof of concept nanopore-based long read polymerase chain reaction (PCR) solution for HLA genotyping. For 15 HLA anthropology-based samples and 13 NHS Blood and Transplant derived samples 40 ng of genomic DNA underwent long-range PCR for class I and II HLA alleles. Pooled PCR products were sequenced on the Oxford Nanopore MinIoON R9.4.1 flow cell. Sequenced reads had HLA genotype assigned with HLA-LA. Called genotypes were compared with reference derived from a combination of short-read next-generation sequencing, Sanger sequence and/or single-site polymorphism (SSP) typing. For concordance, accuracy was 100%, 98.4%, 97.5% and 95.1% for the first, second, third and fourth fields, respectively, to four field accuracy where it was available, otherwise three field in 28 samples for class I calls and 17 samples for class II calls. Phasing of maternal and paternal alleles, as well as phasing based identification of runs of homozygosity, was shown successfully. Time for assay run was 8 hours and the reconstruction of HLA typing data was 15 minutes. Assay cost was £55 ($80USD)/sample. We have developed a rapid and cost-effective long-range PCR and nanopore sequencing-based assay that can genotype the genes within HLA region to up to four field accuracy, identify runs of homozygosity in HLA, reconstruct maternal and paternal haplotypes and can be scaled from multi-sample runs to a single sample.

KEYWORDS
HLA, long read, nanopore

1 INTRODUCTION

Modern organ transplantation techniques has only been made possible by the development of potent
immunosuppressive agents\textsuperscript{2} and the identification of the HLA/major histocompatibility complex\textsuperscript{3} as the determinant of recognition of a transplanted organ as ‘foreign’.

During the process of organ transplantation, HLA matching is performed in order to determine suitability for transplant. The HLA genetics system uses an international classification standard based on observed allelic variation and a common system of representation on genes that make up the HLA region contiguously within chromosome 6 (HLA-A,B,C, DQA1, DPB1, DRB1/3/4/5 and so on). The HLA region is highly polymorphic,\textsuperscript{4,6} with considerable variation observed across the entire region, as well as difficulties in resolving complex changes because of the structure of this region, including multiple paralogs of the gene (ie, HLA-DRB1/3/4/5).

Primer design around these regions using short-read technology is challenging as variation makes it difficult to design primers that span anything other than very short regions, targeting specific alleles, whereas long reads have the advantage of spanning the entire gene of interest.

The nomenclature of the HLA region is necessarily complex, in order to allow a standardised reporting system between laboratories.\textsuperscript{7,8} This nomenclature is known as the WHO Nomenclature for factors of the HLA system, which starts with the name of the locus (ie, HLA-A) followed by up to four fields indicating different levels of variation in the DNA sequence and the resulting protein. The first field defines a group of alleles that corresponds to the serologically defined specificity of HLA. The second field equates to non-synonymous base pair changes that lead to a change in the protein sequence and the third field shows synonymous base pair changes that do not cause protein changes. The fourth field represents changes in the non-coding (ie, intronic) regions. The sequence is sometimes appended with a letter code (C, N, L, Q or S) which represent expression variants (N = null). All currently known HLA alleles are recorded in the IPD-IMGT/HLA database.\textsuperscript{9}

Kidney, pancreas, heart and liver transplantation rely on at least a one field match,\textsuperscript{10} whereas the ideal with allogenic stem cell transplantation would be a two field match\textsuperscript{11} and currently the predominant technique used for this is either Sanger sequencing that provides second field resolution\textsuperscript{12} and sequence-specific polymerase chain reaction (SS-PCR)\textsuperscript{13} for first field resolution, which uses groups of primers to span specific loci in the HLA regions. Although relatively quick (2 hours) this technique is limited by poor resolution to the first or second field only and requires the use of a dedicated real time PCR instrument.

Recent advances in short-read next-generation sequencing (NGS) have allowed resolution of the region in greater detail (four field resolution) by enrichment chemistry and Illumina sequencing.\textsuperscript{14} Short reads preclude effective analysis of the haplotype and phasing of the HLA region, causing problems with accurate classifications of part of the HLA region, including regions with runs of homozygosity.\textsuperscript{15} Use of this technology remains expensive, with a large capital outlay required for the sequencing instrument as well as the use of proprietary software. Short-read technology is comparatively slow compared with SS-PCR as the library preparation and NGS steps takes greater than 24 hours, meaning that accurate four field deceased donor typing is an impossibility.

The development of long-read technology has allowed a solution to these problems. Zero-mode waveguide\textsuperscript{16} sequencing (ie, Pacific BioSciences, PacBio) allows accurate, mid-length (typically 10 kb) reads but at high-cost and no advantage of runtimes over conventional approaches, mainly because of the need to batch samples to make sequencing cost-effective. Nanopore sequencing\textsuperscript{17} (ie, Oxford Nanopore, ONT) systems allow very long read lengths (>1megabase) with rapid runtimes but at the cost of accuracy compared with short-read sequencing, although nanopore technology is continuously improving with increases in accuracy and the current ONT technology has higher raw read accuracy than the competitor PacBio technology. Long-read sequencing of the HLA region has considerable advantages as although it is more error-prone than short read, the haploblock structure is maintained as well as other genomic regions allowing accurate resolution of HLA alleles using haplotype inference\textsuperscript{18} and techniques such as population reference graphing.\textsuperscript{19}

Development of an assay that provides ‘whole gene’ sequencing of the HLA region, along with high-resolution reconstruction of the alleles (known and novel) within it, phasing into maternal and paternal haplotypes and identification of regions of homozygosity, all within a cost-effective, rapid and portable test has the potential to change the field of HLA diagnostics making this type of testing available to all.

A unique technical feature of nanopore sequencing is its scalability: from rapid, one sample, single-gene sequencing through a single flow cell to high volume, whole-genome sequencing. The method is remarkably cost-effective even for a single sample, which means not having to resort to sequencing in large batches. Thus, for full gene HLA sequencing this could mean a fast turn-around for individual patients or recipient/donor pairs, including in a near-patient setting, to multiplex testing of large cohorts, and anything in between. The single-molecule sequencing reads full-length genes in real time
so it includes any DNA variations (in-phase) that, for instance, correspond to expression level or other phenotypes. Although nanopore sequencing is not constrained by the technically demanding nature of other NGS systems this in no way bypasses the requirement for HLA expertise in the authorisation and clinical and scientific use of the information generated by these sequencers. The extreme complexity of HLA requires a corresponding level of understanding to verify, apply and exploit this information. But the agility of our method also means that, where necessary, this can be achieved by the movement of data rather than the movement of DNA or blood samples. This could occur by devolving HLA typing equipment and consumables to the local laboratory level (rather than a central reference) and data uploaded to central reference laboratories for computational analysis and specialist scientific input for resolving allelic ambiguities.

Here we present proof of concept of an assay using long-range PCR to generate full gene length amplicons of the three class I and five class II HLA genes followed by sequencing using a nanopore system and graph-based reconstruction in order to meet these aims and resolve the HLA region to four field accuracy.

2 | METHODS

2.1 | Patient samples

Anonymised patient samples from organ donors were received from NHS Blood and Transplant (NHSBT) under ethical approval (05/Q1605/66). Samples consisted of whole blood taken for routine HLA typing.

A further set of samples (the Frederick Hutchinson HLA anthropology panel) was also chosen that represents 15 samples from different regions of the world allowing us to understand the applicability of the assay to non-CEPH samples and resolve unusual alleles. Unfortunately, some samples from the Fred Hutchinson database had either only two or three field reference data was available so the data from the data from Creary et al21 and Turner et al22 were used instead.

2.2 | DNA extraction

DNA extraction was performed using the Qiagen DNEasy kit using the standard manufacturers’ protocol. DNA was quantified on the Qubit broad range v3 DNA assay (for quantity) and Agilent Tapestation and Nanodrop (for DNA quality). DNA from the Frederick Hutchinson Centre was supplied pre-extracted but was quantified prior to use using the same methodology.

2.3 | HLA Reference typing

Donor DNA was typed initially PCR-single-site polymorphism (SSP) (LinkSeq, supplied by One Lambda) and/or sequence specific oligonucleotide (Life codes, supplied by Imucor) as part of standard patient care. For Illumina-based NGS typing, pre-amplification fluorometric DNA quantitation was performed using the Qubit Broad Range kits (Thermo Fisher, UK). Prior to amplification genomic DNA was diluted to a concentration of 25 ng/μL. HLA loci were amplified using the AllType (One Lambda, California) 11 locus kit, amplifying HLA-A, -B, -C, DRB1, -DRB345, -DQA1, -DQB1, -DPA1 and -DPB1 in a multiplex PCR. Post-amplification, products were purified using AMPure XP (Agencourt) beads and fluorometric quantitation was repeated using the Qubit (Invitrogen) High Sensitivity kit (dsDNA HS assay).

Amplicons were normalised, then enzymatically fragmented. Barcode ligation was followed by size selection (AMPure XP beads), resulting in products of optimal size (300 bp-1000 bp). A secondary amplification was performed prior to subsequent purification (AMPure XP beads), quantification (Qubit dsDNA HS assay) and final equimolar pooling. The pooled library was denatured with NaOH (20%) and loaded onto an Illumina micro flow cell onto the MiSeq platform (Illumina, California). HLA types were analysed using the Type Stream Visual version 1.2 (One Lambda, California) software.

2.4 | HLA class I

Primers sequences are shown in Tables 1 and 2. Amplicons for class I HLA targets (whole gene including exon, intron and untranslated region (UTR)s of HLA A, B, C, E, F and G) were generated in a multiplex reaction using the following conditions: 25 μL PCR reactions were performed using 60 ng DNA, 100 μM primer mix, 1x GoTaq Long (Promega, UK). HLA-E to G were not used in downstream analysis as no reference sample data existed for these genes. The cycling conditions were as follows: 95 C for 2 minutes followed by 30 cycles of 94 C for 30 seconds and 65 for 4 minutes, with a final extension of 10 minutes at 72 C.

2.5 | HLA class II

Primer sequences are shown in Table 2. Amplicons for class II (whole gene including exon, intron and UTRs of
DRB1, DQB1, DQA1, DPA1 and DPB1) were generated with primers mixes as shown in Table 2 using the following conditions: 25 μL PCR reactions were performed using 60 ng DNA, 20 μM primer mix, 1× GoTaq Long (Promega, UK). The cycling conditions were as follows: 95 C for 2 minutes followed by 30 cycles of 94 C for 30 seconds and 65 for 5/7/9/10 minutes, with a final extension of 10 at 72 C. Amplicons were then quantified by Qubit (Thermo Fisher Scientific, UK) according to the manufactures’ instructions and pooled in equimolar amounts for sequencing. All primer sequences were taken from the existing literature or redesigned from this in order to ensure complete coverage of all known HLA alleles by spanning the entirety of the HLA coding regions.

Custom primer design was also carried out for risk alleles in APOL1 that predispose to focal segmental glomerulosclerosis (FSGS) in African patients. The risk alleles were rs73885319 (GRCh38 Chr22:36265860) and rs60910145 (GRCh38 Chr22: 36265988). The PCR primers for this region were spiked into the HLA region as proof of concept.

### 2.6 | Library preparation and sequencing

Barcoded libraries were generated using the native barcoding (EXP-NBD104, EXP-NBD114) and sequencing by ligations kits (SQK-LSK109) from ONT. Briefly 1.3 μg of amplicon pools were end-repaired and a tagged using NEBNext Ultra II module E7546 (3.5 μL End Repair Buffer, 2 μL FFPE repair mix, 3.5 μL Ultra II end-prep reaction buffer and 3 μL of Ultra II end-prep enzyme mix to 1.3 μg DNA in a total of reaction volume of 60 μL). This was incubated at 20 C for 5 minutes followed by 65 C for 5 minutes. Clean up was performed using AMPure XP beads (Beckman Coulter) in a 1× ratio. Quantification was performed using fluorimetry (Qubit) and 500 ng taken through to barcode ligation.

Native barcodes were ligated to 500 ng end-repaired/tailed DNA using NEB blunt/TA ligase M0367 (2.5 μL Native barcode, 25 μL Blunt/TA Ligase Master mIx to 500 ng DNA in a total volume of 50 μL). Following a 10 minutes incubation at room temperature the barcode ligated DNA was cleaned using AMPure XP beads (Beckman Coulter) in a 1× ratio. DNA quantification was performed using fluorimetry (Qubit) and a pool of all samples created with an overall concentration of 700 ng. To reduce the volume a further clean-up was performed using 2.5× AMPure beads and eluting into 65 μL.

Adaptors were ligated by adding 20 μL barcode adaptor mix (ONT) 20 μL quick ligation buffer and 10 μL T4 ligase (NEB Module E6056). Following a 10 minutes incubation at room temperature the adaptor ligated DNA was cleaned using AMPure XP beads in a 0.4× ratio and washed using long fragment buffer (ONT) before eluting in 15 μL of elution buffer (ONT). Final quantification by fluorometry (Qubit) was performed and 30 fmol DNA prepared for sequencing according to the manufacturers’ instructions (ONT).

Sequencing was performed on a MinION R9.4.1 flow cell (MIN-106), a MinION R10 flow cell and a MinION R9.4.1 Flongle flow cell and run for 8 hours using live base calling, files were outputted in Fast5 and Fastq format.

### 2.7 | Bioinformatics analysis

All data analysis was carried out on an Ubuntu 18.04LTS server (with 16 cores and 256 GB memory) and the University of Birmingham BEAR high-performance computing (BEAR-HPC) facility. The jobs
| Target                | Sense primer name | Sequence                        | Anti-sense primer name | Sequence                        | Amplicon size (bp) | Ext time (min) | Reference |
|-----------------------|-------------------|---------------------------------|------------------------|---------------------------------|--------------------|----------------|-----------|
| HLA-DRB1 Mix 1        | DRB1_PE2-F1       | CTGCTGCTCCTTGAGGACATCCACA       | DRB1_PE2-R1            | CTTCTGGCGTGTCCAGTACTCGGCA       | 6146-11478         | 10             | 37        |
|                       | DRB1_PE2-F2       | CTGCTACTCCTTGAGGACATCCACA       | DRB1_PE2-R2            | CTTCTGGCTGTTCCAGAAGACTCGGGA     | 37                 |                | 37        |
|                       | DRB1_PE2-F3       | CTGCTGCTGCTCCTGAGGACATCCACA     | DRB1_PE2-R3            | CTTCTGGCGTGTCCAGTACTCGGCA       | 37                 |                | 37        |
|                       |                   |                                 | DRB1_PE2-R4            | CTTCTGGCTGTTCCAGAAGACTCGGGA     | 37                 |                | 37        |
|                       |                   |                                 | DRB1_PE2-R5            | CTTCTGGCTGTTCCAGTACTCGGGA       | 37                 |                | 37        |
|                       |                   |                                 | DRB1_PE2-R6            | CTTCTGGCTGTTCCAGTACTCGGGA       | 37                 |                | 37        |
| HLA-DRB1-mix2         | DRB1-E2-1.1-F     | GCACGTTTCTTTGTGGCAGCTAATTG      | DRB1-E2-1.1-R          | ATGCACGGGAGGCCATACGTT           | 5106-6218          | 10             | 37        |
|                       | DRB1-E2-3.568-F   | GCACGTTTCTTTGTGGCAGCTAATTG      | DRB1-E2-3.568-R        | ATGCACGGGAGGCCATACGTT           | 37                 |                | 37        |
|                       | DRB1-E2-4-F       | TTTCCCTGTGGCAGCTAAGAGG          | DRB1-E2-4-R            | ATGCACGGGAGGCCATACGTT           | 37                 |                | 37        |
|                       | DRB1-E2-7-4-F     | CACACGACGTTTCTTGTGGGAGG         | DRB1-E2-7-4-R          | CAGATGCATGGGAGGCCAGGGAAGGG      | 37                 |                | 37        |
|                       | DRB1-E2-9-R       | CACACGACGTTTCTTGTGGGAGG         | DRB1-E2-9-R            | CAGATGCATGGGAGGCCAGGGAAGGG      | 37                 |                | 37        |
|                       |                   |                                 | DRB1-E2-10-F           | CACACGACGTTTCTTGTGGGAGG         | 37                 |                | 37        |
|                       |                   |                                 | DRB1-E2-10-F           | CACACGACGTTTCTTGTGGGAGG         | 37                 |                | 37        |
| HLA-DQA1              | HLA-DQA1_F2       | GCCAGGGAGGGAGAAAAACTAATC        | HLA-DQA1_R2            | ATCCAGTTGAGGAGACACGCAC          | 7488               | 7              | 37        |
| HLA-DQB1              | DQB1-F3.1         | AAAGAAACAAACTGCCCTTACACC        | DQB1-R3.1              | TAGATGTCCCTAGTGTGCAAG            | 9093               | 9              | 37        |
|                       | DQB1-F3.2         | AAAGAAACAAACTGCCCTTATACACC      | DQB1-R3.2              | TAGATGTCCCTAGTGTGCAAG            | 37                 |                | 37        |
|                       |                   |                                 | DQB1-R3.3              | TAGATGTCCCTAGTGTGCAAG            | 37                 |                | 37        |
| HLA-DPA1              | HLA-DPA1_F1       | CTCTCTGTACCACGCTGTTCAA          | HLA-DPA1_R1            | TTGGCTCTTGGCTTACCTTCTTT         | 9709               | 9              | 37        |
| HLA-DPB1              | DPB1_pro-F2       | CTCCTCTGACCCCTGATGACCTT         | DPB1_pro-R2            | CCATCTGCCCTCAAGACCTCAA          | 5940               | 5              | 37        |
|                       | HLA-DPB1_F2       | CTCAGTGCTGCCCCTCCCTAGTGA        | HLA-DPB1_R2            | CTCAGTGCTGCCCCTCCCTAGTGA        | 7272               | 5              | 37        |
submitted to the BEAR-HPC facility used 32 cores and 256 GB of system memory with a wall time of 30 minutes per sample. Raw data underwent run management with MinKnow v19.05.0 and basecalling using the Guppy 3.1.5 + 781ed57 basecaller using standard parameters. Quality control plots were generated with NanoPlot 1.26.3. Basecalled FASTQ files were demultiplexed using Guppy barcoder 3.1.5 + 781ed57 (parameters: -t 32 --trim_barcodes --require_barcodes_both_ends -q 0 --compress_fastq). Canu 2.0.0 was then used to correct the raw FASTQ files before alignment (parameters: canu_correct -nanopore-raw genome size = 3.1 g).

Binned reads were aligned to the Illumina Platinum GRCh38 reference genome using MiniMap v2.12 (parameter: -ax map-ont, setting a default mismatch penalty of four), sorted and indexed using Samtools 1.3.1 using htslib 1.31. The aligned BAM file was then input into the HLA-LA* v1.2 pipeline. Output at four field resolution (via the R1_bestguess.txt output) was taken as consensus output to compare to reference Illumina/Sanger/SSP calls. For FSGS risk alleles, the aligned BAM files were filtered for the region of interest (GRCh38 Chr22: 36265800-36266100) and then variant calling was performed using FreeBayes v1.0.0 outputting all sites in genome variant call file mode.

Haplotype phasing of the HLA amplicon data was carried out using WhatsHap v.0.18. Initially variant calls for the amplicon data were produced using Freebayes (parameters: -C 2 -O -q 20 -z 0.10 -E 0 -X -u -p 2 -F 0.6), then using WhatsHap to produce a phased variant call file (parameters: -o phases.vcf input. BAM). A phased haplotype gene transfer format and a haplotagged BAM file were then produced (using the Whatshap stats and Whatshap haplotag commands, respectively) for visualisation. For identification of homozygosity, visual inspection of the variant calls in integrated genome viewer (IGV) was carried out.

Concordance between reference and nanopore sequenced HLA alleles were defined at each field level as to whether there was an exact match. If there was, this was marked as correct. The numbers of correct alleles were divided by the total number of reference fields present across all the samples (Supporting Information) If there was no third or fourth field, the total number of fields was reduced by the number of samples missing the third/fourth field. A number of samples from the Cleary and Turner papers gave more than one string of alleles as potential references. As the nanopore system by default gave a ‘most likely’ string of alleles initially this was chosen as one for comparison and only these references are given.

All data is available on request.

3 | RESULTS

3.1 | Data delivery

For the NHSBT sample typing, in total 2.7 GBases of sequencing data was produced, with a median read length of 3377 bases, a read length N50 of 3606 bases and a median read quality of 9.4. For the anthropology panel sample typing a total of 3.8 GBases of sequencing data was produced, with a median read length of 3170 bases a read length N50 of 3513 bases and a median read quality of 9.9. Runtime was standardised at 8 hours for both panels. For the single Flongle sequenced sample, 43266 reads with a median read length of 1080 bases were produced with a total output of 110 megabases of sequence.

3.2 | Workflow

The timings for the total workflow are shown in Figure S1. The multiplex long-range PCR reaction took 150 minutes, followed by a modified LSK-109 protocol taking 30 minutes, followed by 120 minutes on the nanopore system and 30 minutes of assembly of the HLA calls. The yield of the flow cells over the project determined the runtime. Typically, a run of 2 hours for a single sample on the Flongle (40 mb yield) and 50 minutes for 12 multiplexed samples on the MinION (396 mb yield) allowed sufficient data for 500× coverage (Figure S2). We therefore set the runtime at 2 hours.

3.3 | Class I and class II HLA call accuracy

In a preliminary analysis, it was found that at least 500× coverage of each amplicon was required for accurate HLA calling, therefore in samples with low coverage, these were rerun. For the first set of NHSBT samples, 11 samples underwent analysis for class I alleles (Table 3). All samples were correct for the first field, NHSBT sample 1 had a reference blood transfusion service (BTS) HLA-C allele of seven, for the MiSeq call it was C*07:02:01:03 (although the C*07:123 was given as the second option in the BTS typing) and for the nanopore, it was C*07:123.

For the second set of NHSBT samples, a more challenging set of two samples were chosen. Concordance for class I and class II calls was 100% with 0% error.

For the anthropology panel, 15 samples underwent analysis for class I and class II alleles (Table 4). All samples were an exact match apart from sample IHW09376.
### TABLE 3
List of results for samples within NHSBT experiment

| ID | NHSBT ID | Technique | A* | B* | C* |
|----|----------|-----------|----|----|----|
| 1  | Sample 1 Reference | A*24:02:01:01 | A*31:01:02:01 | B*07:02:01:01 | B*39:06:02:01 | C*07:02:01:01 | C*07:02:01:03 |
|    | MinIon | A*24:02:01:01 | A*31:01:02:01 | B*07:02:01:01 | B*39:06:02:01 | C*07:02:01:01 | C*07:123 |
|    | BTS    | 24         | 31  | 7  | 39 | 7  | 7  | 7 |
| 2  | Sample 3 Reference | A*02:01:01:01 | A*11:01:01:01 | B*35:01:01:05 | B*40:01:02:01 | C*03:04:01:01 | C*04:01:01:05 |
|    | MinIon | A*02:01:01:01 | A*11:01:01:01 | B*35:01:01:05 | B*40:01:02:01 | C*03:04:01:01 | C*04:01:01:05 |
|    | BTS    | 2          | 11  | 35 | 40:01 | 03:02 | 4  |
| 3  | Sample 4 Reference | A*01:01:01:01 | A*02:01:01:01 | B*08:01:01:01 | B*44:02:01:01 | C*05:01:01:02 | C*07:01:01:01 |
|    | MinIon | A*01:01:01:01 | A*02:01:01:01 | B*08:01:01:01 | B*44:02:01:01 | C*05:01:01:02 | C*07:01:01:01 |
|    | BTS    | 1          | 2   | 8  | 44:02 | 5   | 7  |
| 5  | Sample 6 Reference | A*02:01:01:01 | A*03:01:01:01 | B*27:05:02:05 | B*44:02:01:01 | C*02:02:02:01 | C*05:01:01:02 |
|    | MinIon | A*02:01:01:01 | A*03:01:01:01 | B*27:05:02:05 | B*44:02:01:01 | C*02:02:02:01 | C*05:01:01:02 |
|    | BTS    | 2          | 3   | 27 | 44:02 | 2   | 5  |
| 6  | Sample 9 Reference | A*02:01:01:01 | A*03:01:01:01 | B*40:01:02:01 | B*44:02:01:01 | C*03:04:01:01 | C*05:01:01:01 |
|    | MinIon | A*02:01:01:01 | A*03:01:01:01 | B*40:01:02:01 | B*44:02:01:01 | C*03:04:01:01 | C*05:01:01:01 |
|    | BTS    | 2          | 3   | 40:01 | 44:02 | 03:02 | 5  |
| 7  | Sample 10 Reference | A*02:11:01:01 | A*11:01:01:01 | B*40:06:01:02 | B*51:01:01:12 | C*12:09 | C*15:02:01:01 |
|    | MinIon | A*02:11:01:01 | A*11:01:01:01 | B*40:06:01:02 | B*51:01:01:12 | C*12:09 | C*15:02:01:01 |
|    | BTS    | 2          | 11  | 40:06 | 51   | 12  | 15 |
| 8  | Sample 11 Reference | A*02:01:01:01 | A*33:01:01:01 | B*44:03:01:02 | B*51:01:01:05 | C*02:02:02:01 | C*15:02:01:01 |
|    | MinIon | A*02:01:01:01 | A*33:01:01:01 | B*44:03:01:02 | B*51:01:01:05 | C*02:02:02:01 | C*15:02:01:01 |
|    | BTS    | 2          | 33  | 44:03 | 51   | 02  | 15 |
| 9  | Sample 12 Reference | A*01:01:01:01 | A*68:01:02:02 | B*08:01:01:01 | B*35:08:01:01 | C*04:01:01:06 | C*07:141:02 |
|    | MinIon | A*01:01:01:01 | A*68:01:02:02 | B*08:01:01:01 | B*35:08:01:01 | C*04:01:01:06 | C*07:141:02 |
|    | BTS    | 1          | 68  | 35  | 4    | 7   |
| 10 | Sample 13 Reference | A*01:01:01:01 | A*02:01:01:01 | B*08:01:01:01 | B*15:01:01:01 | C*03:04:01:01 | C*07:01:01:16 |
|    | MinIon | A*01:01:01:01 | A*02:01:01:01 | B*08:01:01:01 | B*15:01:01:01 | C*03:04:01:01 | C*07:01:01:16 |
|    | BTS    | 1          | 2   | 8   | 15:01 | 03:02 | 7  |
| 11 | Sample 14 Reference | A*24:02:01:01 | A*25:01:01:01 | B*07:02:01:01 | B*18:01:01:01 | C*07:02:01:01 | C*12:03:01:01 |
|    | MinIon | A*24:02:01:01 | A*25:01:01:01 | B*07:02:01:01 | B*18:01:01:01 | C*07:02:01:01 | C*12:03:01:01 |
|    | BTS    | 24         | 25  | 7   | 18  | 7   | 12 |
| 12 | Sample 15 Reference | A*11:01:01:01 | A*68:01:02:01 | B*07:02:01:01 | B*44:02:01:01 | C*05:01:01:02 | C*07:02:01:03 |
|    | MinIon | A*11:01:01:01 | A*68:01:02:01 | B*07:02:01:01 | B*44:02:01:01 | C*05:01:01:02 | C*07:02:01:03 |
|    | BTS    | 11         | 68  | 7   | 44:02 | 5   | 7  |

Note: Colours represent accuracy of match—green, all fields match; yellow, second field mismatch; red, first field mismatch. RunID, internal run ID; Alternate ID, NHSBT sample ID; Technique-reference: MinIon sequencing by NHSBT; MinIon, Nanopore-based HLA typing; BTS, NHSBT serotyping derived allele. Abbreviation: NHSBT, NHS Blood and Transplant.
### Table 4  List of results for samples within anthropology panel experiment

| Sample ID | Ethnicity | Sex | Technique | A* | B* | C* |
|-----------|-----------|-----|-----------|----|----|----|
| S20       | Unknown   | Unknown | Reference | A*24:02:01:01 | B*07:02:01:01 | C*07:02:01:01 |
| S27       | Unknown   | Unknown | Reference | A*66:01:01:01 | B*14:02:01:01 | C*02:02:01:01 |
| HW093177  | Caucasian | M    | Reference | A*02:01:01:01 | B*27:09 | C*16:01:01:01 |
| HW01093   | Caucasian | M    | Reference | A*01:01:01:01 | B*44:03:01:01 | C*16:01:01:01 |
| HW09381   | Unknown   | M    | Reference | A*02:06:01:01 | B*39:08 | C*07:02:01:01 |
| HW01141   | Caucasian | M    | Reference | A*02:01:01:01 | B*39:08 | C*07:01:01:01 |
| HW09021   | South African Zulu | M    | Reference | A*01:01:01:01 | B*44:04:01:01 | C*16:01:01:01 |
| HW09107   | Japanese  | M    | Reference | A*24:02:01:01 | B*54:01:01 | C*01:02:01:01 |
| HW09388   | Caucasian | M    | Reference | A*11:01:01:01 | B*40:01:02 | C*03:04:01:02 |
| HW01175   | Caucasian | F    | Reference | A*02:01:01:01 | B*15:01:01:01 | C*03:01:01:01 |
| HW09375   | Caucasian | F    | Reference | A*33:01:01:01 | B*14:02:01:01 | C*08:02:01:01 |
| HW09376   | Hispanic? | F    | Reference | A*01:01:01:01 | B*27:03 | C*02:02:02:01 |
| HW09056   | German    | F    | Reference | A*02:01:01:01 | B*27:03 | C*02:02:02:01 |
| HW09367   | Asian     | F    | Reference | A*11:02:01:01 | B*38:02:01 | C*07:02:01:01 |
| HW09373   | Unknown   | M    | Reference | A*68:02:01:01 | B*58:01:01:01 | C*08:02:01:01 |
| HW09024   | Japanese  | F    | Reference | A*11:01:01:01 | B*35:01:01:01 | C*04:01:01:01 |
| HW09045   | French    | F    | Reference | A*03:01:01:01 | B*51:01:01:01 | C*07:04:01:01 |
| HW ID | Sex | Ethnicity | Reference | DPA1*          | DPA2*          | DQA1*          | DQA2*          | EBRI*          | DRB1*          | DRB3*          |
|-------|-----|-----------|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| S20   | Unknown | Unknown | Reference | DPA1*01:03:01:01 | DPA2*01:02:02:01 | DQA1*03:01:01:01 | DQA2*01:02:01:01 | EBRI*04:03:02:01 | DRB1*04:07:01:01 | DRB3*01:05:01:01 |
| S22   | Unknown | Unknown | Reference | DPA1*01:03:01:01 | DPA2*01:02:02:01 | DQA1*03:01:01:01 | DQA2*01:02:01:01 | EBRI*04:03:02:01 | DRB1*04:07:01:01 | DRB3*01:05:01:01 |
| M27   | Unknown | Unknown | Reference | DPA1*01:03:01:01 | DPA2*01:02:02:01 | DQA1*03:01:01:01 | DQA2*01:02:01:01 | EBRI*04:03:02:01 | DRB1*04:07:01:01 | DRB3*01:05:01:01 |
| S22   | Unknown | Unknown | Reference | DPA1*01:03:01:01 | DPA2*01:02:02:01 | DQA1*03:01:01:01 | DQA2*01:02:01:01 | EBRI*04:03:02:01 | DRB1*04:07:01:01 | DRB3*01:05:01:01 |
| S22   | Unknown | Unknown | Reference | DPA1*01:03:01:01 | DPA2*01:02:02:01 | DQA1*03:01:01:01 | DQA2*01:02:01:01 | EBRI*04:03:02:01 | DRB1*04:07:01:01 | DRB3*01:05:01:01 |

(Continues)
| IHW ID | Sex   | Ethnic/Population Group | Technique | DRB3* | DRB4* | DRB5* |
|--------|-------|-------------------------|-----------|-------|-------|-------|
| S20    | Unknown | Unknown                 | Reference |       |       |       |
|        |        |                         | MinIon    | DRB4*01:03:01:01 | DRB5*01:01:01:01 |
| S27    | Unknown | Unknown                 | Reference | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
| IHWO9377 | M    | Caucasian               | Reference | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
| IHWO1093 | M    | Caucasian               | Reference | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
| IHWO9381 | M    | Unknown                 | Reference | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
| IHWO1141 | M    | Caucasian               | Reference | DRB3*01:01:02:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*01:02:01:01 | DRB4*03:01:01:01 |
| IHWO9021 | M    | South African           | Reference | DRB3*01:01:02:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*01:01:02:01 | DRB4*01:01:01:01 |
| IHWO9107 | M    | Japanese                | Reference | DRB3*01:01:02:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*01:01:02:01 | DRB4*01:01:01:01 |
| IHWO9388 | M    | Caucasian               | Reference | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
| IHWO1175 | F    | Caucasian               | Reference | DRB3*02:02:01:01 | DRB4*03:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*03:01:01:01 |
| IHWO9375 | F    | Caucasian               | Reference | DRB3*02:02:01:01 | DRB4*03:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*03:01:01:01 |
| IHWO9376 | F    | Hispanic?               | Reference | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
| IHWO9056 | F    | German                  | Reference | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:01:01:01 |
| IHWO9367 | M    | Asian                   | Reference | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
| IHWO9373 | M    | Unknown                 | Reference | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*03:01:01:01 | DRB4*01:03:01:01 |
| IHWO9024 | F    | Japanese                | Reference | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
|        |        |                         | MinIon    | DRB3*02:02:01:01 | DRB4*01:03:01:01 |
| IHWO9045 | F    | French                  | Reference | DRB3*02:02:01:01 | DRB4*01:03:01:01 |

**Note:** IHW ID, International Histocompatibility Workshop ID; Technique-reference: alleles supplied by IHW, MinIon, Nanopore-based HLA typing. Colours represent accuracy of match—green, all fields match; yellow, second field mismatch; red, first field mismatch. A = HLA class I alleles; B = HLA class II alleles DPA1, DPB1, DQA1, DQB1; C = HLA-DRB3,4,5 alleles.
For the single second field error, the reference call was HLA-B*27:05:02 and the nanopore call HLA-B*27:110. This represents a single nucleotide change (G > A) and could represent a sequencing error for either method. For the class II alleles, all samples were a match except IHW09021, where the reference for HLA-DRB1 was DRB1*03:02:01 and the MinION call was 03:03. Examination of the raw data showed that this was a sequence alignment error caused by an indel from nanopore sequencing. When manual correction was applied the allele resolved correctly.

3.4 | FSGS/APOL1 allele calling

In order to understand the use of the nanopore system for single nucleotide polymorphism (SNP) variants that may predispose to clinically relevant diseases, the G1 and G2 risk alleles for FSGS were spiked into the mix. The G1 alleles (rs73885319, Chr22:36265860, NC_000022.10:g.36661906A>G and rs60910145, Chr22:36265988, NC_000022.10:g.36662034T>G) were called in all the NHSBT samples. Of the 12 samples, all had the A reference allele. The G2 allele is a 6 bp (rs71785313, Chr22: 36266000, NC_000022.10:g.36662046_36662051delTTATAA) deletion in APOL1. Of the 12 samples, the indel was not seen. Of note, several small common SNPs within 200 bp of the region of the SNPs of the APOL1 gene were observed, for example, rs1403581130.

3.5 | R9.4.1 vs R10 pores

As part of an early access programme, the project was given to the new R10 nanopore to run HLA typing samples on (Figure 1). The R10 was called using the identical pipeline to the R9 data and displayed significantly higher single-base accuracy. In Figure 2, all three panels show IGV plots of the R10 data (top of each panel) vs R9 data (bottom of each panel) showing a greatly reduced level of single base mismatches across the three HLA genes shown —HLA-DQB1 (top), HLA-DPB1 (middle) and the highly polymorphic HLA-DRB5. Interestingly, raw average mapping quality (MAPQ) scores were similar between R10 and R9 (49 vs 44) and base MAPQ scores (16.2 vs 15.5) equivalent to base error rates of 2.4% vs 2.8%. Median alignment score (AS, where higher is a
FIGURE 2  IGV plot showing read accuracy between R9 and R10 pore. Top panel = HLA-DRB1; middle panel = HLA-DPB1; bottom panel = HLA-DRB5. IGV, integrated genome viewer.
better score) as reported by MiniMap2 was 4350 for the R10 pore vs 722 for the R9.4.1 pore (Mann-Whitney P < .0001, Figure 3). Median number of mismatches (NM, where fewer mismatches is better) as reported by MiniMap2 was 51 for the R10 pore vs 551 for the R9.4.1 pore (Mann-Whitney P < .0001, Figure 3).

3.6 | Single sample calling on the Flongle device

In order to understand whether the output of a miniaturised nanopore device—the Flongle flow cell—a single sample (NHSBT sample 27) was run on a R9.4.1 Flongle. Data output was 0.9 Gb and 100% accuracy was seen at four field level for both class I and class II fields for this sample.

3.7 | HLA phasing and identification of homozygosity in HLA-DRB1

Identification of maternal and paternal contributions to HLA alleles is vital to identify runs of homozygosity, which may affect organ matching, as well as being difficult to detect using short-read technologies. In order to show the ability of nanopore long-read sequencing to phase HLA as well as identifying runs of homozygosity, a single sample (anthropology panel sample 1, IHW09377) was chosen for analysis. After variant calling with FreeBayes, haplogroups were generated with WhatsHap. For this sample, two haplogroups were derived for each sample, presumably the maternal and paternal contribution to the inherited HLA of the proband. This could be clearly seen in IGV for HLA-DRB1 (Figure 1) by generating a haplogroup tagged BAM files. In this figure, the separate contributions from maternal and paternal alleles can be seen in the differently coloured reads (green for haplogroup 1, blue for haplogroup 2). Each haploblock spanned the entire amplicon, reinforcing the co-dominant inheritance of the HLA system. Visual inspection of sample IHW09377 in the anthropology panel showed that HLA-DRB1 was homozygous (Figure 4).

3.8 | Speed and cost-effectiveness

The nanopore-based assay showed considerable speed-based advantages over conventional typing. DNA extraction took 1 hour, library preparation 3 hours and sequencing 4 hours to 20 hours depending on the volume of sequence data required. Bioinformatics analysis took 1 hour on a 16 core Intel Xeon server with 256 GB of system memory running Ubuntu LTS 18.04, meaning that in total the assay could be run within 8 hours which is a considerable time saving over NGS and SSP methods. In terms of cost-effectiveness, assuming the pooling of 15 samples on a MinION flow cells, we calculated that the blood DNA extraction costs £25, the long ranger PCR £3, barcoding and library preparation £10 and flow cell costs £27.36, in total giving a cost of £38 without DNA extraction. Typical commercial HLA typing costs range from £300 to £800. For users who do not wish to invest in dedicated hardware, a Ubuntu VM with 64 cores and 256 GB of RAM costs £2.87 (as of February 2020) an hour on the Microsoft Azure cloud (https://azure.microsoft.com/en-gb/pricing/calculator/), meaning a sample compute cost of £2.87/sample (as time is billed per hour).

4 | CONCLUSIONS

In this study, we have shown that full-length HLA typing using long-range PCR and sequencing on a nanopore sequencing system is highly accurate, cheaper than the nearest alternative and feasible for deployment into the field using a ‘laboratory in a suitcase’ approach. This approach uses the portability of nanopore sequencing, coupled to a laptop computer and portable PCR equipment to allow HLA typing in resource-poor conditions.

Current methodologies for typing of HLA rely on highly specific, but not broad assays such as SSP assays that can sequence individual alleles but not provide in-depth reconstruction of the entire region of interest. This means that for rarer alleles although SSP provides accuracy this is at the cost of a single assay that can be used for all patients. Long amplicons, provided by long-range
PCR have been previously performed using short-read sequencing, our method coupled with the long-read capability of the nanopore system provides a unique ability to accurately understand the HLA region.

Our use of long-range PCR has advantages in that the entire gene can be encompassed in one PCR reaction, allowing reconstruction of haplotypes and accurate resolution of complex parts of the HLA region. It also requires limited sample input (typically 50 ng of genomic DNA). The disadvantage is the time needed for the extension phase of the PCR reaction, with the longest PCR amplicon (>10 kb) requiring over 10 minutes per cycle which means that a typical long-range PCR reaction for HLA typing takes up to 3 hours. This methodology has the advantage that it can be performed in relatively resource-poor environments enabling its use in lower and middle-income countries (LMIC). When conceived, our original experimental plan for this assay was one that could be used in LMIC as an alternative to expensive and slow out of country HLA typing. An alternative to overcome the disadvantage of extensive long-range PCR amplification may be the use of loop-mediated isothermal amplification (LAMP-PCR), which has the advantage of rapidity but would require extensive primer redesign.

One potential solution to this problem would be the use Cas9 enrichment to pull out HLA regions from samples of genomic DNA, which are then subject to a ligation reaction and sequenced using the nanopore device. This has multiple advantages, the first of which is the ability to retrieve in an unbiased fashion the HLA region of interest. In addition, methylation of HLA regions could be natively called off the retrieved DNA because of the inherent ability of nanopore systems to detect methylated bases as signal change within the pore. This would offer the ability to understand the relationship between HLA expression and methylation. A disadvantage of the Cas9 method is the relatively large amounts of genomic DNA required for input (>3 μg), although blood samples will typically provide this, DNA extraction has to be performed carefully in order to maximise retrieved DNA fragment length. Spin column-based extraction can typically obtain >20 kb fragments, although for reads similar to the full-length of HLA (2 mb), more exotic methods of DNA extraction such as the Sambrook and Russell method or pulsed gel field

**FIGURE 4** IGV plot showing that HLA-DRB1 is homozygous, represented by the VCF allele call plot (panel below ideogram) is composed of mostly homozygous (red) SNPs and occasional heterozygous (blue) SNPs. IGV, integrated genome viewer
electrophoresis must be used. These methods would make it considerably less accessible to LMIC and would restrict this type of application for research environments only.

Another potential benefit to the nanopore system would be parallel sequencing of an expression assay of HLA in order to understand the effect of HLA expression on transplant outcome. It is highly likely that accurate 3 to 4 field resolution of HLA alleles along with methylation and expression data within the same assay would transform our understanding of the importance of this region in HLA typing. There is evidence that HLA typing of the class II system is of considerable importance in haemopoetic stem cell transplantation, and the HLA expression has a bearing on outcomes in this type of transplantation.

The algorithm we have used for reconstruction of the HLA region (HLA-LA) has significant advantages as it uses a population reference graph of HLA alleles to accurately reconstruct the HLA region to high accuracy. One issue with population reference graph reconstruction is that it is both computationally and memory intensive, especially with long-read nanopore data. We are working with the authors to modify the algorithm to work on nanopore-based data such that reconstruction is feasible on computers in the field. Another option is the use of a cloud-based infrastructure where nanopore sequencing data is uploaded from the field and HLA types called in real time. This has the advantage of centralised control of the algorithm and quality assurance, but the disadvantage of requiring a method of transfer of nanopore sequencing runs (typically 5-6 GB) which may be difficult in LMIC.

In conclusion, we present our methodology for using nanopore sequencing for four field resolution of all class I and class II alleles using nanopore technology. It is cost-effective, rapid and has many practical advantages over short-read sequencing and we suggest it may represent the most suitable future methodology for HLA typing.

**ACKNOWLEDGMENTS**
We acknowledge the efforts of Leila Luhesi, Alica Ham and James Bryer for their review of the manuscript prior to publication. This project was supported in its entirety by a grant from the Queen Elizabeth Hospital Charity. Salary support for A. D. B. is provided by a Cancer Research UK Advanced Clinician Scientist award.

**CONFLICT OF INTEREST**
A. D. B. has received travel funding from Oxford Nanopore. The other authors declare no conflict of interest.

**DATA AVAILABILITY STATEMENT**
Data available on reasonable request.

**ORCID**
Andrew D. Beggs https://orcid.org/0000-0003-0784-2967

**REFERENCES**
1. Linden PK. History of solid organ transplantation and organ donation. Crit Care Clin. 2009;25(1):165-184. ix.
2. Colaneri J. An overview of transplant immunosuppression—history, principles, and current practices in kidney transplantation. Nephrol Nurs J. 2014;41(6):549-560. quiz 61.
3. Terminology: nomenclature for factors of the HLA system, 1980. World Health Organization. Immunology. 1982;46(1):231-234.
4. Williams TM. Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. J Mol Diagn. 2001;3(3):98-104.
5. Klitz W, Hedrick P, Louis EJ. New reservoirs of HLA alleles: pools of rare variants enhance immune defense. Trends Genet. 2012;28(10):480-486.
6. Robinson J, Guethlein LA, Cereb N, et al. Distinguishing functional polymorphism from random variation in the sequences of >1,000 HLA-A, -B and -C alleles. PLoS Genet. 2017;13(6):e1006862.
7. Nunes E, Heslop H, Fernandez-Vina M, et al. Definitions of histocompatibility typing terms. Blood. 2011;118(23):e180-e183.
8. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010;75(4):291-455.
9. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA database. Nucleic Acids Res. 2020;48(D1):D948-D955.
10. Montgomery RA, Tatapudi VS, Leffell MS, Zachary AA. HLA in transplantation. Nat Rev Nephrol. 2018;14(9):558-570.
11. Tieryc JM. How to select the best available related or unrelated donor of hematopoietic stem cells? Haematologica. 2016;101(6):680-687.
12. Lazar R, Bu T, Yang R, et al. Human leukocyte antigen (HLA) typing by DNA sequencing. Methods Mol Biol. 2013;1034:161-195.
13. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue Antigens. 1992;39(5):225-235.
14. Levene MJ, Korfich J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;299(5607):682-686.
15. Shah N, Decker WK, Lapushin R, et al. HLA homozygosity and haplotype bias among patients with chronic lymphocytic leukemia: implications for disease control by physiological immune surveillance. Leukemia. 2011;25(6):1036-1039.
16. Levene MJ, Korfich J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;299(5607):682-686.
17. Stoddart D, Hereaj AJ, Mikhailova E, Maglia G, Bayle H. Single-nucleotide discrimination in immobilized DNA
oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A.* 2009;106(19):7702-7707.

18. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet.* 2013;93(4):687-696.

19. Dillthey A, Cox C, Iqbal Z, Nelson MR, McVean G. Improved genome inference in the MHC using a population reference graph. *Nat Genet.* 2015;47(6):682-688.

20. Petersdorf EW, Malkki M, O’Huigin C, et al. High HLA-DP expression and graft-versus-host disease. *N Engl J Med.* 2015;373(7):599-609.

21. Creary LE, Guerra SG, Chong W, et al. Next-generation HLA typing of 382 International Histocompatibility Working Group reference B-lymphoblastoid cell lines: report from the 17th International HLA and Immunogenetics Workshop. *Hum Immunol.* 2019;80(7):449-460.

22. Turner TR, Hayhurst JD, Hayward DR, et al. Single molecule real-time DNA sequencing of HLA genes at ultra-high resolution from 126 International HLA and Immunogenetics Workshop cell lines. *HLA.* 2018;91(2):88-101.

23. De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34(15):2666-2669.

24. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. * Genome Res.* 2017;27(5):722-736.

25. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094-3100.

26. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987-2993.

27. Li H, Handsaker B, Wyssker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.

28. Dillthey AT, Gourraud PA, Mentzer AJ, Cereb N, Iqbal Z, McVean G. High-accuracy HLA type inference from whole-genome sequencing data using population reference graphs. *PLoS Comput Biol.* 2016;12(10):e1005151.

29. Garrison E, Gabor M. Haplotype-based variant detection from short-read sequencing. 2012; arXiv:1207.3907 [q-bio.GN]. https://arxiv.org/abs/1207.3907.

30. Patterson M, Marschall T, Pisanti N, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol.* 2015;22(6):498-509.

31. Bunce M, Passey B. HLA typing by sequence-specific primers. *Methods Mol Biol.* 2013;1034:147-159.

32. Yin Y, Lan JH, Nguyen D, et al. Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: results from over 10,000 donor recruitment samples. *PLoS One.* 2016;11(10):e0165810.

33. Jia H, Guo Y, Zhao W, Wang K. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Sci Rep.* 2014;4:5737.

34. Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Pereira NF, Petzl-Erler ML, Donadi EA. Evaluation of computational methods for the reconstruction of HLA haplotypes. *Tissue Antigens.* 2010;76(6):459-466.

35. Lee PL. DNA amplification in the field: move over PCR, here comes LAMP. *Mol Ecol Resour.* 2017;17(2):138-141.

36. Gabrieli T, Sharihi H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.* 2018;46(14):e87.

37. Watson CM, Crinnion LA, Hewitt S, et al. Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Lab Invest.* 2020;100(1):135-146.

38. Liu Q, Fang L, Yu G, Wang D, Xiao CL, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford nanopore sequencing data. *Nat Commun.* 2019;10(1):2449.

39. Soneson C, Yao Y, Bratus-Neuschwander A, Patrignani A, Robinson MD, Hussain S. A comprehensive examination of nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun.* 2019;10(1):3359.

40. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* 2018;36(4):338-345.

41. Bertaina A, Andreani M. Major histocompatibility complex and hematopoietic stem cell transplantation: beyond the classical HLA polymorphism. *Int J Mol Sci.* 2018;19(2):621-622.

42. Park M, Seo JJ. Role of HLA in hematopoietic stem cell transplantation. *Bone Marrow Res.* 2012;2012:680841.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Stockton JD, Nieto T, Wroe E, et al. Rapid, highly accurate and cost-effective open-source simultaneous complete HLA typing and phasing of class I and II alleles using nanopore sequencing. *HLA.* 2020;96:163–178.  
[https://doi.org/10.1111/tan.13926](https://doi.org/10.1111/tan.13926)