Check for updates

HLA

WILEY

HLA
Immune Response Genetics

ORIGINAL ARTICLE

High-Resolution HLA Class I Genotyping Using G-Seq Platform-Based Nanopore Long-Read Sequencing Integrated With the HLATyper Pipeline

Hong-Hong He[1] | Run-Yu Xu[2] | Qiu-Lei Zhong[3] | Yan Yu[2] | Ming-Yuan Wang[1] | Miao He[3] | Long-Hai Tang[1]

[1]Department of Transfusion Research Laboratory, Suzhou Blood Center, Suzhou, Jiangsu, China | [2]Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, Southwest Bio-resources R&D Key Laboratory of Sichuan Province, Laboratory for Ex Situ Conservation and Resource Utilization of Mountain Plants, College of Life Sciences, Sichuan University, Chengdu, Sichuan, China | [3]Institute of Blood Transfusion, Chinese Academy of Medical Sciences & Peking Union Medical College, Chengdu, Sichuan, China

Correspondence: Miao He (hemiao@ibt.pumc.edu.cn) | Long-Hai Tang (szbc@live.com)

Received: 3 February 2026 | **Revised:** 24 September 2026 | **Accepted:** 7 October 2026

Keywords: HLA class I genotyping | HLATyper pipeline | PCR-SBT | TGS

ABSTRACT

This study evaluated the performance and clinical applicability of a novel third-generation sequencing (TGS) strategy for high-resolution HLA class I genotyping, employing nanopore long-read sequencing combined with a dedicated analytical workflow, HLATyper. A total of 49 clinical samples were analysed in parallel using conventional PCR-based sequence-based typing (PCR-SBT) and a TGS-based approach on the G-seq500 platform with the in-house-developed HLATyper pipeline. Genotyping results for HLA-A, -B and -C loci were compared between the two methods with respect to concordance, typing resolution and the ability to resolve ambiguous allele assignments. Ambiguity rates between the two methods were evaluated using Fisher's exact test. Complete concordance was observed between PCR-SBT and the TGS method at the two-field level across all samples. Ambiguous allele assignments were detected in 63.27% (31/49) of cases using PCR-SBT, whereas the TGS strategy produced a single, unambiguous genotype for every sample ($p < 0.0001$). Furthermore, TGS enabled assignment of four-field typing through comprehensive sequence alignment and successfully resolved ambiguities and exonic coverage gaps inherent to short-read approaches. G-seq500 platform-based nanopore long-read sequencing combined with the HLATyper analytical pipeline enables high-resolution, unambiguous and comprehensive HLA class I genotyping. This integrated strategy represents a valuable complementary approach for clinical applications requiring precise HLA matching, such as transplantation and platelet transfusion support.

1 | Introduction

The HLA system, encoded within the major histocompatibility complex (MHC), exhibits the highest degree of genetic polymorphism in the human genome [1]. HLA molecules play a central role in immune recognition and are critical determinants of transplantation compatibility [2–4]. As a result, high-resolution HLA genotyping is critical for minimising graft rejection and improving the efficacy of platelet transfusion, as well as for a wide range of applications in immunogenetic research [5, 6].

The majority of known HLA polymorphisms are concentrated within Exons 2–4. Conventional polymerase chain reaction-based Sanger sequencing typing (PCR-SBT) was historically

Hong-Hong He and Run-Yu Xu contributed equally to this work.

© 2026 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.
All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

employed as a reference method for HLA genotyping and has been widely implemented in blood centres and transplant laboratories [7–9]. While PCR-SBT provides reliable results in many clinical scenarios, it is constrained by inherent technical limitations, including restricted amplicon lengths, low throughput and challenges in resolving phase ambiguities at polymorphic sites, which may lead to ambiguous allele assignments [10, 11].

With the advancement of sequencing technologies, next-generation sequencing (NGS) has become the current preferred method for HLA genotyping in transplant centres [12, 13]. However, the short read length of NGS (typically 100–300 bp) limits its ability to span extended polymorphic regions within HLA genes, complicating the resolution of allelic phasing; as a result, relatively high levels of typing ambiguity persist [1, 14].

In recent years, third-generation sequencing (TGS) technologies, including Oxford Nanopore sequencing and SMRT sequencing developed by PacBio, have provided an alternative route to addressing phasing ambiguity by generating ultra-long reads (several kilobases), enabling real-time sequencing and reducing reliance on PCR amplification [15, 16]. Long reads can span full-length HLA genes and link heterozygous variants across a single molecule, thereby facilitating high-resolution typing. While early nanopore platforms were limited by high raw single-read error rates, recent advances in sequencing chemistry have dramatically improved raw read and consensus accuracy [15, 17–20]. In this study, we developed a new HLA typing strategy based on the G-seq500 nanopore sequencing platform. This platform utilises a 'sequencing-by-synthesis (SBS)' approach, which distinguishes it from traditional nanopore sequencers that directly read DNA bases [21]. By leveraging a unique biochemical system, including a specially designed nanopore, polymerase and modified nucleotides, the G-seq500 generates high-quality long-read data with extended read lengths (up to tens of kilobase pairs) and exceptional consensus accuracy exceeding Q50 at 50× coverage [21]. When combined with our in-house HLATyper tool suite, it provides a comprehensive bioinformatics solution for TGS-based HLA genotyping. HLATyper is designed to overcome the challenges posed by the relatively high error rates of long-read data by optimising sequence alignment, allele inference and phasing algorithms, thereby enabling precise and unambiguous typing of HLA class I genes (HLA-A, -B and -C). We systematically compared the performance of the G-seq500–based approach with PCR-SBT for HLA class I genotyping, focusing on improvements in resolution, typing precision and the elimination of ambiguous results, to provide more reliable technical support for transplantation matching, platelet transfusion and immunogenetic research.

2 | Materials and Methods

2.1 | Sample Collection

A total of 49 samples were obtained from routine HLA class I genotyping cases at the Suzhou Blood Center in June 2025. Each sample was analysed by PCR-SBT and, in parallel, by nanopore-based TGS using the G-seq500 platform. All samples met national donor health criteria and collection was performed in accordance with informed consent requirements.

2.2 | Instruments and Reagents

DNA extraction instrument (TGuide S32; Tiangen Biotech (Beijing) Co. Ltd., Beijing, China), PCR thermal cycler (ABI 9700; Applied Biosystems, USA), centrifuge (Eppendorf 5810R; Eppendorf AG, Germany), Sanger sequencer (ABI 3730; Applied Biosystems, USA) and third-generation sequencer (G-seq500; Geneus Technologies Co. Ltd., Chengdu, China). The commercially available TGuide Smart Blood Genomic DNA High-Efficiency Extraction Kit (A1011A; Tiangen Biotech (Beijing) Co. Ltd., Beijing, China), HLA-A, -B and -C genotyping assay kits (202411M, 202411N, 2014110; Jiangsu Weihe Biotechnology Inc., Jiangsu, China), universal sample preparation kit (20250503; Geneus Technologies Co. Ltd., Chengdu, China), universal sequencing kit (G-TR01, 20250601; Geneus Technologies Co. Ltd., Chengdu, China) and microfluidic chip (G-MK02, 20250636; Geneus Technologies Co. Ltd., Chengdu, China) were used.

2.3 | DNA Extraction and Sequencing of HLA Class I Genes by PCR-SBT and Long-Read TGS

2.3.1 | Genomic DNA Extraction

Genomic DNA was extracted from peripheral anticoagulated blood using a commercial genomic DNA extraction kit on the TGuide S32 system. DNA concentrations ranged from 20 to 40 ng/μL, with A260/A280 ratios between 1.7 and 1.9.

2.3.2 | PCR-SBT for HLA Class I Typing

According to the manufacturer's instructions for the HLA genotyping kit, sequencing of Exons 2–4 was performed for the HLA-A, -B and -C loci. The HLA-A, -B and -C loci were amplified by PCR, followed by agarose gel electrophoresis, PCR product purification and sequencing. PCR was performed with an initial denaturation at 95°C for 5 min, followed by 36 cycles of 93°C for 30 s, 63°C for 40 s and 72°C for 150 s, and a final extension at 72°C for 5 min. Sequencing reactions were run with an initial denaturation at 96°C for 1 min, followed by 25 cycles of 98°C for 25 s and 60°C for 150 s. HLA-A, -B and -C sequences and the corresponding high-resolution typing results were generated using the AccuType HLA SBT analysis software (V1.6, TBG Biotechnology Corp.).

2.3.3 | TGS-Based HLA Class I Sequencing

Primers for HLA-A, -B and -C genes were designed based on previously published sequences [22]. HLA-A, -B and -C genes were amplified using the TaKaRa PrimeSTAR GXL DNA Polymerase kit (TaKaRa, Japan). The 50 μL PCR amplification reaction volume contained 150 ng of genomic DNA, 2 units of PrimeSTAR GXL DNA Polymerase, 10 μL of 5× PrimeSTAR GXL Buffer,

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 μL of dNTP Mixture (2.5 mM each) and 2.5 μL (4 pmol/μL) of each primer. The basic cycling parameters were as follows: 96°C for 2 min, followed by 35 cycles, each cycle consisting of 98°C for 10 s, 60°C for 20 s and 72°C for 6 min, and a final extension at 72°C for 10 min. The long-range PCR reactions were performed using the TGreat Expert Gradient PCR Instrument (TIANGEN, OSE-GP-03/05, China). The PCR products were purified and labelled with barcodes according to the instructions of the universal Sample Preparation Kit, then mixed to construct a DNA library, which was subsequently prepared into a sequencing complex. The prepared complexes were then diluted to the final library concentration and sequenced on the G-seq500 nanopore platform [23].

2.4 | High-Resolution Genotyping Using the HLATyper Pipeline

TGS data were analysed using HLATyper, an integrated and automated workflow developed in this study, which performs high-resolution genotyping of HLA class I genes (HLA-A, -B and -C) directly from raw sequencing files. The workflow leverages nucleotide and gene sequence information from the IPD-IMGT/HLA database [24] to extract exonic structure annotations and build the corresponding reference databases. It consists of five core steps, implemented as separate Python scripts and executed sequentially, encompassing reference sequence selection, multiple sequence alignment, sequence optimisation, allele sequence selection and genotype inference. The individual steps are described in detail below.

2.4.1 | Reference Sequence Selection

For each sample, the first N long-read sequences (default $N = 1000$) were selected as query reads and aligned against the IPD-IMGT/HLA reference database [24] using BLASTN [25] with an E-value threshold of 1e−5. Only the best hit per read was retained. FASTA files and alignment outputs were parsed using BioPython [26]. Temporary files were generated with unique identifiers to prevent file conflicts and ensure reproducibility. Candidate reference alleles were identified based on the frequency of best-hit matches.

2.4.2 | Sequence Cleaning and Optimisation

Using the reference sequences selected above, BLASTN [25] was used to align the third-generation reads to the database, and the resulting alignments were converted with MView [27] into a FASTA-formatted multiple sequence alignment file. This formatted alignment served as the input for downstream consensus analysis. To improve data quality, aligned sequences were further filtered and corrected. NumPy-enabled column-wise processing was used for efficient handling of alignment columns, and parallelisation was applied to enhance computational efficiency. Filtering criteria included: (i) removing sequences shorter than 95% of the mean length; (ii) excluding sequences whose numbers of mismatched bases relative to the reference exceeded a predefined threshold; and (iii) filling regions where the column-wise gap proportion was below 10%. The final output was a curated, high-quality FASTA file retaining the reference sequences and all sequences that met the filtering criteria.

2.4.3 | Candidate Allele Sequence Selection

A frequency-based filtering strategy for exact sequence matches was applied to identify candidate allele sequences. Curated sequences were grouped according to identical nucleotide sequences, and the occurrence frequency of each unique sequence was calculated. Low-frequency sequences were filtered using a predefined abundance threshold, whereas recurrent high-frequency sequences were retained as candidate allele sequences for subsequent HLA typing. The retained sequences were ranked in descending order of read abundance and formatted as individual allele candidates. The output FASTA file contained the reference sequence together with the retained candidate allele sequences.

2.4.4 | Allele Assignment and Variant Detection

Coding exons were extracted according to the reference annotation and concatenated for preliminary two-field allele assignment, whereas the original genomic candidate sequences were retained for subsequent genomic-level allele analysis. For preliminary typing, a k-mer–based candidate retrieval strategy was first applied to the coding-region sequences, followed by pairwise Needleman–Wunsch global alignment. Candidate references were ranked using a weighted score incorporating sequence overlap, reference-to-query length consistency, mismatches and insertions/deletions.

During k-mer-based candidate retrieval of the coding-region sequences, codon-position-specific weights were applied, with the third codon position assigned a lower weight than the first and second positions. The original, unsplit coding sequence was also included in the k-mer retrieval score to retain local sequence-context information.

Following preliminary typing, the corresponding two-field result was used only to restrict the genomic reference set to alleles belonging to the same allele family. The original genomic candidate sequence obtained from the approximately 3.5 kb HLA amplicon, rather than the exon-only sequence, was then compared with the selected genomic reference alleles using k-mer candidate retrieval followed by pairwise global alignment. The genomic comparison therefore incorporated both exonic and intronic sequence information contained within the sequenced amplicon.

For genomic allele assignment, candidate ranking incorporated sequence overlap, reference-to-query length consistency, mismatches, internal insertions/deletions and terminal reference gaps. Internal indels and terminal reference gaps were evaluated separately to reduce bias caused by incomplete or truncated reference sequences. The allele with the lowest weighted difference score was reported as the closest matching genomic reference allele.

Nucleotide-level differences between the candidate sequence and the selected reference allele were subsequently identified

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from the pairwise alignment and reported in a variant table. A corresponding link to the EBI IPD-IMGT/HLA database was provided for reference allele verification [24], and a separate typing and variant report was generated for each sequence (Figure 1).

2.5 | Sequencing Data Quality and Computational Performance Assessment

Sequencing performance was assessed directly from FASTQ files. For each sample, the number of reads, total bases and read length statistics (minimum, maximum and mean) were calculated. Base quality was summarised using the mean Phred quality score, the proportions of Q20 and Q30 bases, the expected error rate and expected accuracy derived from Phred scores.

Because only FASTQ files were available, alignment-dependent metrics were not directly estimated. These included strict per-base depth, detailed error decomposition (e.g., substitution versus indel errors), and homopolymer-associated indel profiling. Instead, theoretical mean coverage depth was estimated as total bases divided by the target length, providing a FASTQ-level approximation of sequencing depth.

For computational performance assessment, runtime and memory usage were evaluated using GNU time -v. Benchmarking was performed with Python 3.12.9 (Anaconda) on a server running Ubuntu 22.04.5 LTS (x86_64; Linux kernel 5.15.0–130-generic) and equipped with an Intel Xeon Gold 6148 CPU @ 2.40 GHz (80 cores) and 125 GiB RAM.

2.6 | Statistical Analysis

Fisher's exact test was used to compare the high-resolution HLA class I genotyping results obtained by PCR-SBT and TGS. $p < 0.05$ was considered indicative of statistical significance.

3 | Results

3.1 | Comparison of HLA Class I Genotyping Between PCR-SBT and TGS

High-resolution genotyping of HLA-A, -B and -C loci was performed on 49 samples, with comparisons made at the two-field level. Results from TGS and PCR-SBT were 100% concordant. Regarding typing resolution, PCR-SBT provided results at the two-field level (e.g., *HLA-A*24:02*), whereas TGS demonstrated the capability to resolve four-field assignments through full-length sequencing of the HLA-A, -B and -C loci (e.g., *HLA-A*24:02:01:01*; see Table S1).

PCR-SBT sequencing yielded ambiguous results in 31 samples (63.27%), with only 18 samples (36.73%) providing unambiguous

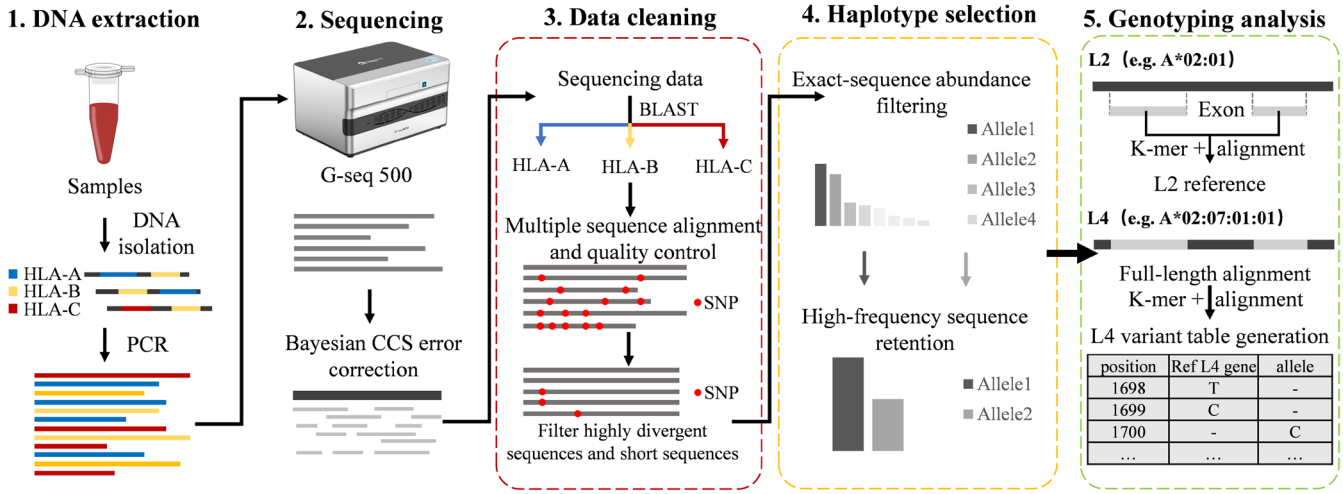


FIGURE 1 | Workflow of HLA genotyping. (1) Genomic DNA is extracted and purified from peripheral blood samples. The obtained DNA is subjected to PCR amplification of the HLA-A, HLA-B and HLA-C genes. After purification, the amplicons are labelled, libraries are constructed and sequencing complexes are prepared using a universal sample preparation kit and a universal sequencing kit based on nanopore sequencing technology. (2) The prepared sequencing complexes are diluted according to the final library concentration and sequenced on the G-seq500 nanopore sequencing platform. The G-seq500 instrument performs basecalling and polishing of the raw reads to generate relatively accurate circular consensus sequences (CCS). (3) The sequencing data are aligned against three BLAST databases corresponding to HLA-A, HLA-B and HLA-C using BLASTN, retaining only the unique best match for each sequence. After batch BLASTN alignment, MView is used to convert BLAST outputs into multiple sequence alignment formats, followed by removal of low-quality sequences with excessive SNPs or insufficient length. (4) High-frequency haplotypes (putative alleles) are selected by exact-sequence abundance filtering, whereby identical sequences are grouped and recurrent high-frequency sequences are retained as candidate allele sequences. (5) Exons are trimmed and concatenated, and the resulting sequences are compared against the reference database using a k-mer-based candidate retrieval strategy followed by pairwise global alignment to identify the best-matching reference alleles for preliminary typing. For each sample, a sample-specific reference subset is then constructed, and full-length alignments are performed to obtain high-resolution genotyping results.

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

genotypes. In contrast, TGS did not produce any ambiguous results. The HLA-A, -B and -C high-resolution genotyping results obtained by TGS exhibited significantly higher resolution compared with PCR-SBT, with the difference being statistically significant ($p < 0.0001$; Table 1).

3.2 | Resolution of Ambiguous HLA Class I Assignments by TGS

As described above, the long-read sequencing advantage of TGS allowed for unique genotyping results across all samples, while PCR-SBT detected multiple allele combinations. The number of ambiguous combinations at the HLA-A, -B and -C loci was 14, 10 and 11, respectively, with the highest number of combinations being 8 (at the HLA-B locus) and the lowest being 2. Upon analysing the causes of ambiguity, we found that 26 cases were due to phase uncertainty between adjacent SNPs, resulting from the limited read length of PCR-SBT (Figure 2). Additionally, six cases arose because primers failed to cover other exons, making it impossible to distinguish alleles, and two cases were caused by both factors. The specific genotyping results are presented in Tables 2–4. For consistency in data presentation and comparison, TGS results in this section were summarised at the third-field level.

TABLE 1 | Comparative analysis of high-resolution genotyping results of HLA-A, -B and -C in 49 specimens.

Sequencing method	Ambiguous cases, *n* (%)	Unique results, *n* (%)	*p*
PCR-SBT	31 (63.27%)	18 (36.73%)	< 0.0001
TGS	0 (0%)	49 (100%)	

3.3 | Sequencing Data Quality and Computational Performance Assessment

Representative FASTQ-level assessment showed consistently high sequencing quality across samples. In three representative samples (IDs 137, 78 and 35), the number of reads ranged from 7801 to 9771, corresponding to 37,789,111–47,920,149 bases per sample. Mean read length ranged from 4,797.76 to 4,904.32 bp, indicating stable long-read performance across samples.

The mean Phred quality scores ranged from 31.32 to 31.55. The proportion of Q20 bases ranged from 98.18% to 98.64%, whereas the proportion of Q30 bases ranged from 70.16% to 71.67%. The expected base-level accuracy derived from FASTQ quality values ranged from 99.6908% to 99.7656%, supporting the overall high quality and consistency of the sequencing data.

The resulting theoretical mean coverage depth ranged from 2540.10× to 3221.09× in the representative samples, indicating that sequencing depth was sufficient for downstream HLA typing. Although strict alignment-based per-base depth could not be calculated from FASTQ-only data, these results demonstrate that the sequencing yield and quality were adequate for robust analysis.

To assess computational performance, six FASTQ files (42, 110, 220, 120, 175 and 172 MB; approximately 839 MB total) were processed in a representative benchmark for parallel HLA-A, -B and -C typing. The observed wall-clock runtime was 1 h 47 min 6 s. GNU time -v reported 71,322.38 s of user CPU time and 4,463.42 s of system CPU time, corresponding to a total of 75,785.80 s (approximately 21.1 CPU-hours). The average CPU utilisation was 1179%, equivalent to using an average of approximately 11.8 CPU cores.

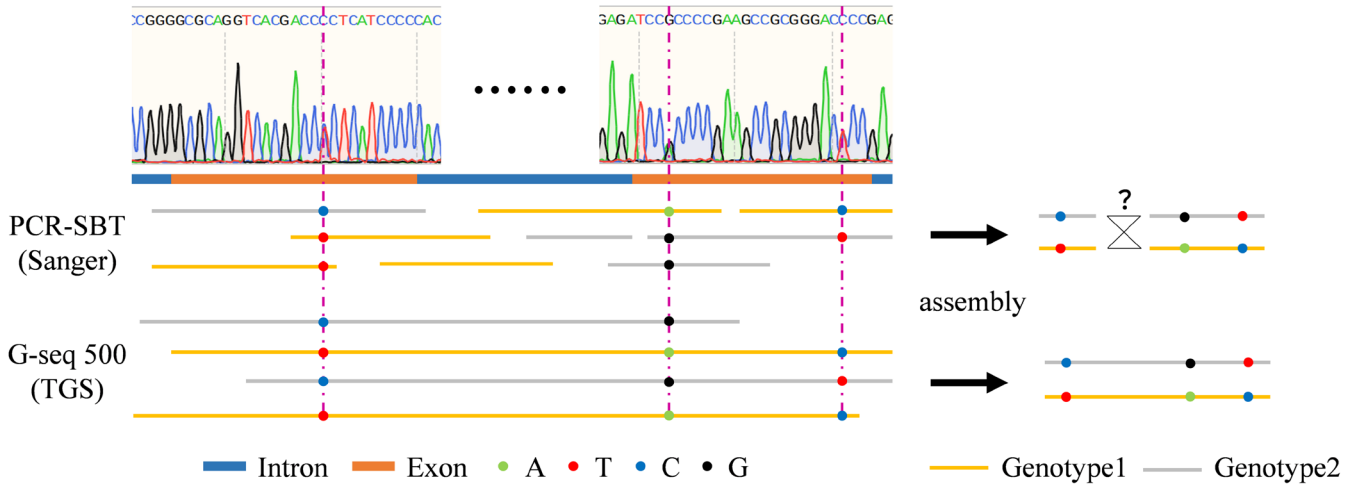


FIGURE 2 | Comparison of PCR-SBT and nanopore TGS based on the G-seq500 platform. PCR-SBT sequencing uses traditional Sanger sequencing, which is a short-read sequencing technology. Commercialised reagent kits typically amplify and sequence regions with concentrated polymorphisms. Because of its limited read length, PCR-SBT cannot span intronic regions and therefore cannot determine whether adjacent heterozygous SNPs are located on the same chromosome. In contrast, nanopore TGS based on the G-seq500 platform is a long-read sequencing technology. Long reads can span intronic regions and cover the entire gene region, thereby resolving phase ambiguity that is difficult to address with Sanger sequencing.

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Comparison of ambiguous HLA-A loci results detected by PCR–SBT and TGS.

PCR-SBT	TGS	No. of samples	Cause of ambiguity
*A*02:01 A*11:01*	*A*02:01:01*	3	[a]
*A*02:34 A*11:19*	*A*11:01:01*		
*A*24:02 A*26:01*	*A*26:01:01*	3	[a]
*A*24:06 A*26:08*	*A*24:02:01*		
*A*24:56 A*26:12*			
*A*03:01 A*26:01*	*A*26:01:01*	2	[a]
*A*03:08 A*26:13*	*A*03:01:01*		
*A*02:01 A*68:01*	*A*68:01:02*	2	[a,b]
*A*02:01 A*68:11N*	*A*02:01:01*		
*A*02:34 A*68:03*			
*A*02:35 A*68:05*			
*A*02:01 A*24:02*	*A*24:02:01*	1	[a]
*A*02:36 A*24:03*	*A*02:01:01*		
*A*02:12 A*24:13*			
*A*02:01 A*03:01*	*A*03:01:01*	1	[a]
*A*02:01 A*03:21N*	*A*02:01:01*		
*A*02:01 A*03:26*			
*A*02:24 A*03:17*			
*A*02:34 A*03:08*			
*A*02:35 A*03:108*			
*A*01:01 A*11:01*	*A*01:01:01*	1	[a]
*A*01:17 A*11:19*	*A*11:01:01*		
*A*01:01 A*30:01*	*A*01:01:01*	1	[a]
*A*01:03 A*30:01*	*A*30:01:01*		
*A*01:01 A*02:01*	*A*01:01:01*	1	[a]
*A*36:04 A*02:36*	*A*02:01:01*		
*A*02:01 A*11:01*	*A*11:01:01*	1	[a]
*A*02:34 A*11:19*	*A*02:01:01*		
*A*02:01 A*24:02*	*A*02:01:01*	1	[a]
*A*02:01 A*24:02L*	*A*24:02:01*		
*A*02:36 A*24:03*			
*A*02:12 A*24:13*			
*A*11:01 A*24:03*	*A*11:01:01*	1	[a]
*A*11:04 A*24:10*	*A*24:03:01*		
*A*24:02 A*33:03*	*A*33:03:01*	1	[a]
*A*24:02L A*33:03*	*A*24:02:01*		
*A*24:53 A*33:03*			
*A*24:57 A*33:03*			
*A*24:81 A*33:03*			
*A*24:64 A*33:03*			
*A*02:07 A*24:02*	*A*02:07:01*	1	—
*A*02:07 A*24:02L*	*A*24:02:01*		

*Note: A*68:11N*, *A*03:21N*, *A*24:02L*, *N* = null allele; *L* = low-expression allele.
[a]Phase ambiguity, requiring additional specific primers for distinction.
[b]Polymorphic sites located in other exons.

Peak resident memory usage (maximum resident set size) was 9,468,724 kB (approximately 9.0 GiB), and no swapping occurred during execution. These results indicate that the workflow is primarily CPU-intensive, with moderate memory requirements and some disk I/O, and does not require GPU acceleration. Although benchmarking was performed on a high-performance server, the observed resource usage suggests that, for a workload of this scale, the pipeline can be executed on a standard multi-core CPU workstation.

4 | Discussion

Given that PCR-SBT remains widely employed in HLA genotyping laboratories, this study systematically compared conventional PCR-SBT with nanopore-based TGS using the G-seq500 platform for high-resolution HLA class I genotyping. While both methods showed complete concordance at the two-field typing level, the TGS-based approach demonstrated advantages in typing resolution, ambiguity reduction and overall interpretability of results.

PCR-SBT is an effective method for HLA genotyping and, when combined with common and well-documented (CWD) allele reference tables [28], can resolve a large proportion of HLA class I genotyping cases. However, due to the concentration of HLA polymorphisms primarily in Exons 2–4, commercial PCR-SBT kits usually amplify and sequence only these regions. When complex polymorphisms occur in other regions of the HLA class I gene, incomplete coverage may result in ambiguous allele combinations. In this study, 63.27% of samples exhibited multiple allele combinations, with the maximum number of combinations being 8 (at the HLA-B locus) and the minimum being 2, illustrating the limitations of PCR-SBT in fully capturing HLA gene complexity.

As of July 2026, the number of recognised HLA alleles in the IPD-IMGT/HLA database has reached 46,652. As more polymorphic sites are discovered, nucleotide differences between alleles are likely to decrease, making highly homologous alleles more difficult to distinguish [14]. This trend further contributes to the increasing ambiguity in short-read–based genotyping methods. Phase uncertainty remains a major

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Comparison of ambiguous HLA-B loci results detected by PCR–SBT and TGS.

PCR-SBT	TGS	No. of samples	Cause of ambiguity
*B*15:01 B*51:01*	*B*15:01:01*	2	[a]
*B*15:01N B*51:01*	*B*51:01:01*		
*B*15:04 B*51:06*			
*B*15:08 B*52:01*			
*B*15:24 B*78:02*			
*B*15:38 B*51:02*			
*B*35:43 B*52:02*			
*B*15:56 B*51:07*			
*B*35:01 B*51:01*	*B*35:01:01*	2	[a]
*B*35:11 B*51:09*	*B*51:01:01*		
*B*53:01 B*78:02*			
*B*35:24 B*51:02*			
*B*53:06 B*78:04*			
*B*08:01 B*35:01*	*B*08:01:01*	1	[a]
*B*08:35 B*35:34*	*B*35:01:01*		
*B*08:23 B*35:29*			
*B*13:02 B*35:01*	*B*35:01:01*	1	[a]
*B*13:01 B*35:37*	*B*13:02:01*		
*B*46:01 B*81:01*	*B*81:02:01*	1	[b]
*B*46:01 B*81:02*	*B*46:01:01*		
*B*35:03 B*51:01*	*B*51:01:01*	1	[a]
*B*35:34 B*51:13*	*B*35:03:01*		
*B*53:04 B*78:02*			
*B*39:01 B*40:06*	*B*39:01:01*	1	[a]
*B*39:06 B*40:11*	*B*40:06:01*		
*B*15:01 B*48:01*	*B*48:01:01*	1	[a]
*B*15:01N B*48:01*	*B*15:01:01*		
*B*15:07 B*48:03*			
*B*35:01 B*39:05*	*B*39:05:01*	1	[a]
*B*35:06 B*39:07*	*B*35:01:01*		
*B*15:02 B*51:01*	*B*15:02:01*	1	[a]
*B*15:13 B*78:02*	*B*51:01:01*		

*Note: B*15:01N*, *N* = null allele.
[a]Phase ambiguity, requiring additional specific primers.
[b]Polymorphic sites located in other exons.

challenge, as heterozygous alleles with adjacent SNPs cannot be reliably assigned to the same chromosome without additional amplification or specialised primers. In our dataset,

TABLE 4 | Comparison of ambiguous HLA-C loci results detected by PCR–SBT and TGS.

PCR-SBT	TGS	No. of samples	Cause of ambiguity
*C*03:03 C*07:02*	*C*03:03:01*	4	[a]
*C*03:04 C*07:56*	*C*07:02:01*		
*C*01:02 C*08:01*	*C*08:01:01*	2	[b]
*C*01:02 C*08:22*	*C*01:02:01*		
*C*04:01 C*15:02*	*C*04:01:01*	3	[a,b]
*C*04:09N C*15:02*	*C*15:02:01*		
*C*04:10 C*15:03*			
*C*04:82 C*15:02*			
*C*04:04 C*15:08*			
*C*04:01 C*07:02*	*C*07:02:01*	2	[b]
*C*04:09N C*07:02*	*C*04:01:01*		
*C*04:01 C*06:02*	*C*04:01:01*	1	[b]
*C*04:09N C*06:02*	*C*06:02:01*		
*C*04:82 C*06:02*			
*C*05:01 C*08:01*	*C*05:01:01*	1	[a]
*C*08:02 C*08:10*	*C*08:01:01*		
*C*03:04 C*07:02*	*C*03:04:01*	1	[a]
*C*03:40 C*07:29*	*C*07:02:01*		
*C*03:35 C*07:10*			
*C*01:02 C*04:01*	*C*04:01:01*	1	[a]
*C*01:17 C*04:10*	*C*01:02:01*		

(Continues)

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | (Continued)

PCR-SBT	TGS	No. of samples	Cause of ambiguity
*C*04:01* *C*04:01*	*C*04:01:01*	1	[b]
*C*04:01* *C*04:09N*	*C*04:01:01*		
*C*04:09N* *C*04:09N*			
*C*03:04* *C*16:02*	*C*03:04:01*	1	[a]
*C*03:07* *C*16:01*	*C*16:02:01*		
*C*04:01* *C*14:02*	*C*04:01:01*	1	[b]
*C*04:09N* *C*14:02*	*C*14:02:01*		

*Note: C*04:09N*, *N* = null allele.
[a]Phase ambiguity, requiring additional specific primers for distinction.
[b]Polymorphic sites located in other exons.

26 ambiguous allele combinations were attributable solely to unresolved phasing, underscoring the necessity of long-read sequencing for resolving phasing issues.

TGS technologies—exemplified by G-seq long-read sequencing—have shown considerable potential in clinical applications [29–31], particularly in the development of platelet donor databases to reduce the incidence of immune-mediated platelet transfusion refractoriness, where high-resolution genotyping is critical for accurate donor–recipient matching [32]. Nanopore sequencing platforms, known for their long-read lengths, flexibility and real-time sequencing capabilities, are increasingly gaining attention for HLA genotyping [21, 28, 33].

In this study, we employed G-seq500, a novel nanopore sequencing platform. By leveraging a sequencing-by-synthesis strategy with real-time signal acquisition and incorporating a novel nanopore, an optimised polymerase, and specially modified nucleotide strands, the system enables accurate base identification during DNA synthesis [21]. The G-seq500 platform used in this study integrates advanced nanopore design, an optimised polymerase, and base-modification chemistry to improve raw-read accuracy while retaining the advantages of long-read sequencing, thereby providing a solid foundation for high-quality HLA genotyping. To date, the G-seq500 sequencing platform has been successfully applied to the genotyping of *Vibrio parahaemolyticus*, demonstrating clear advantages in cost-effectiveness, turnaround time and high-resolution discrimination [21]. The platform's ability to generate reads up to 5 kb—a length sufficient to fully span each individual HLA class I gene, including HLA-A, HLA-B and HLA-C at ~3.5 kb each—facilitates the direct phasing of SNPs and enables detection of polymorphisms beyond canonical exons. This study demonstrated that all samples analysed by TGS produced unique genotyping results, with typing resolution improving from two-field to four-field resolution, significantly enhancing the reliability of results. This capability is critical for achieving precise matching between donors and recipients and reducing immune rejection.

We developed the HLATyper bioinformatics pipeline to complement the experimental platform. HLATyper provides an integrated and automated workflow for high-resolution HLA genotyping using TGS data. The tool suite incorporates multiple steps, including dynamic reference sequence selection, sequence cleaning, exact-sequence abundance filtering for candidate allele selection, and fine-grained exon-level alignment to reduce error accumulation and phase ambiguity inherent to long-read sequencing data.

In conclusion, the integrated approach combining the G-seq500 nanopore sequencing platform with the HLATyper analytical pipeline establishes a robust framework for high-resolution HLA class I genotyping. By generating full-length reads that span the entire HLA-A, -B and -C genes, this strategy effectively eliminates the phase ambiguity inherent to traditional PCR-SBT and enables assignment of four-field typing through comprehensive sequence alignment. This integrated strategy represents a valuable and reliable supplementary method applicable to clinical scenarios requiring precise compatibility matching. The HLATyper pipeline is fully open-source, with the source code available at https://gitee.com/sculab/HLATyper.

However, this study has several limitations, including a relatively small sample size and the lack of standardised reference materials (e.g., IHW cell lines with confirmed four-field typing for HLA-A, -B and -C) for quality control. Although the G-seq500 platform generated full-length reads for each of the HLA-A, -B and -C genes, the four-field typing results reported in this study should be interpreted as the closest database match based on full-length sequence alignment, rather than experimentally confirmed true alleles, due to the lack of certified reference cell lines with validated four-field typing. We emphasise that the reported four-field result represents the closest matching sequence rather than a definitive assignment and should be interpreted together with the accompanying variant-site table. Therefore, future research will focus on expanding the sample cohort, obtaining IHW reference cell lines for rigorous external validation, and extending HLATyper to HLA class II genotyping. Additionally, we plan to validate the efficacy and stability of HLATyper in multicentre clinical settings to further support its clinical implementation and standardisation.

Author Contributions

Investigation: Hong-Hong He and Run-Yu Xu. Methodology: Run-Yu Xu, Qiu-Lei Zhong and Yan Yu. Project administration: Ming-Yuan Wang, Miao He and Long-Hai Tang. Writing – original draft: Hong-Hong He and Run-Yu Xu. Writing – review and editing: Yan Yu, Miao He and Ming-Yuan Wang. Funding acquisition: Hong-Hong He, Long-Hai Tang and Miao He.

Funding

This work was supported by a grant from the Gusu Talent Program (grant number: (2022) 160); the Suzhou Science and Technology Plan Project (grant numbers: SKY2023077; SKYD2023164; SYW2024145); the Jiangsu Provincial Health Commission Project (grant number: MQ2024021); the Suzhou Municipal Clinical Medical Expert

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Team Introduction Program 2026; and the Key R&D Support Plan of the Science and Technology Bureau of Chengdu (grant number: 2026-YF05-01325-SN).

Ethics Statement

This study was reviewed and approved by the Medical Ethics Committee of Suzhou Blood Center (approval No. SZBC202309).

Consent

Written informed consent was obtained from all participants in the study.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data generated or analysed during this study are included in this published article.

References

1. S. Buhler, M. Nørgaard, R. Steffensen, et al., "High Resolution HLA Genotyping With Third Generation Sequencing Technology-A Multicentre Study," *HLA* 104, no. 2 (2024): e15632.

2. L. E. Hudson and R. L. Allen, "Leukocyte Ig-Like Receptors - A Model for MHC Class I Disease Associations," *Frontiers in Immunology* 7 (2016): 281.

3. S. Medhasi and N. Chantratita, "Human Leukocyte Antigen (HLA) System: Genetics and Association With Bacterial and Viral Infections," *Journal of Immunology Research* 2022 (2022): 9710376.

4. T. Profaizer, A. Pole, C. Monds, J. C. Delgado, and E. Lázár-Molnár, "Clinical Utility of Next Generation Sequencing Based HLA Typing for Disease Association and Pharmacogenetic Testing," *Human Immunology* 81, no. 7 (2020): 354–360.

5. M. Meneghini, A. Perona, E. Crespo, et al., "On the Clinical Relevance of Using Complete High-Resolution HLA Typing for an Accurate Interpretation of Posttransplant Immune-Mediated Graft Outcomes," *Frontiers in Immunology* 13 (2022): 924825.

6. C. Crane, M. Niemann, B. Dale, et al., "High-Resolution HLA Genotyping Improves PIRCHE-II Assessment of Molecular Mismatching in Kidney Transplantation," *Human Immunology* 85, no. 3 (2024): 110813.

7. H. Xin, J. Li, H. Sun, et al., "Benchmarking of 5 Algorithms for High-Resolution Genotyping of Human Leukocyte Antigen Class I Genes From Blood and Tissue Samples," *Annals of Translational Medicine* 10, no. 11 (2022): 633.

8. H. Feng, Y. Jin, and B. Wu, "Strategies for Neoantigen Screening and Immunogenicity Validation in Cancer Immunotherapy (Review)," *International Journal of Oncology* 66, no. 6 (2025): 43.

9. J. Cha, M. Hur, H. Kim, et al., "Comparative Assessment of Risk and Turn-Around Time Between Sequence-Based Typing and Next-Generation Sequencing for HLA Typing," *Diagnostics (Basel)* 14, no. 16 (2024): 1793.

10. I. C. Baek, E. J. Choi, D. H. Shin, H. J. Kim, H. Choi, and T. G. Kim, "Allele and Haplotype Frequencies of Human Leukocyte Antigen-A, -B, -C, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPA1, and -DPB1 by Next Generation Sequencing-Based Typing in Koreans in South Korea," *PLoS One* 16, no. 6 (2021): e0253619.

11. C. G. Haider, B. Sun, Y. Wang, et al., "Optimized Multiplex PCR-NGS for Comprehensive HLA Genotyping in Chinese Populations: Resolving Ambiguities at High Resolution," *Frontiers in Immunology* 16 (2025): 1551173.

12. J. L. Duke, C. Lind, K. Mackiewicz, et al., "Determining Performance Characteristics of an NGS-Based HLA Typing Method for Clinical Applications," *HLA* 87, no. 3 (2016): 141–152.

13. E. Arrieta-Bolanos, A. C. M. Lima, M. Andreani, et al., "HLA Matching in Contemporary Haematopoietic Cell Transplantation: Recommendations From the EBMT Practice Harmonisation and Guidelines Committee," *Bone Marrow Transplantation* 61, no. 8 (2026): 1015–1023.

14. X. Lian, D. Zhang, Y. Wen, et al., "Application of ONT in Ambiguous Results of the HLA-DPB1 Genotyping," *Journal of Clinical Transfusion and Laboratory Medicine* 26, no. 3 (2024): 365–370.

15. N. El-Lagta, L. Truong, F. Ayora, et al., "Revolutionising High Resolution HLA Genotyping for Transplant Assessment: Validation, Implementation and Challenges of Oxford Nanopore Technologies' Q20(+) Sequencing," *HLA* 104, no. 4 (2024): e15725.

16. H. Choi, E. J. Choi, H. J. Kim, et al., "A Walk Through the Development of Human Leukocyte Antigen Typing: From Serologic Techniques to Next-Generation Sequencing," *Clinical Transplantation and Research* 38, no. 4 (2024): 294–308.

17. Y. Wang, Y. Zhao, A. Bollas, Y. Wang, and K. F. Au, "Nanopore Sequencing Technology, Bioinformatics and Applications," *Nature Biotechnology* 39, no. 11 (2021): 1348–1365.

18. C. Liu, "A Long Road/Read to Rapid High-Resolution HLA Typing: The Nanopore Perspective," *Human Immunology* 82, no. 7 (2021): 488–495.

19. J. A. Geo, R. Ameen, S. Al Shemmari, and J. Thomas, "Advancements in HLA Typing Techniques and Their Impact on Transplantation Medicine," *Medical Principles and Practice* 33, no. 3 (2024): 215–231.

20. F. B. Pedersen, A. W. Hauge, J. F. Hansen, L. S. Andersen, S. Blomsterberg, and H. Bruunsgaard, "Cost-Effective and Highly Scalable Typing of HLA Classes I and II Genes of up to 96 Individuals Using Nanopore Sequencing," *HLA* 105, no. 4 (2025): e70164.

21. R. Zhou, Y. Qi, L. Zhou, et al., "MtlA Gene Sequencing Based on 3rd -Generation Sequencing for Vibrio Typing," *BMC Microbiology* 25, no. 1 (2025): 347.

22. T. Shiina, S. Suzuki, Y. Ozaki, et al., "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, no. 4 (2012): 305–316.

23. L. Wang, Q. Wang, Z. Tang, et al., "Concurrent Classical and Hypervirulent *Klebsiella pneumoniae* Infection in Distinct Host Niches Revealed by a Rapid Nanopore Whole-Genome and Plasmid Sequencing Method," *Frontiers in Cellular and Infection Microbiology* 15 (2025): 1633833.

24. D. J. Barker, G. Maccari, X. Georgiou, et al., "The IPD-IMGT/HLA Database," *Nucleic Acids Research* 51, no. D1 (2023): D1053–D1060.

25. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic Local Alignment Search Tool," *Journal of Molecular Biology* 215, no. 3 (1990): 403–410.

26. P. J. A. Cock, T. Antao, J. T. Chang, et al., "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics," *Bioinformatics* 25, no. 11 (2009): 1422–1423.

27. N. P. Brown, C. Leroy, and C. Sander, "MView: A Web-Compatible Database Search or Multiple Alignment Viewer," *Bioinformatics* 14, no. 4 (1998): 380–381.

28. N. Chen, W. Zhang, L. Dong, et al., "Explore of Nanopore Sequencing Technology in Ambiguities of HLA Genotyping," *Chinese Journal of Blood Transfusion* 38, no. 3 (2025): 309–315.

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29. M. Devriese, J. Rouquie, S. Da Silva, et al., "Single Locus HLA Sequencing With the Nanopore Technology for HLA Disease Association Diagnosis," *HLA* 103, no. 3 (2024): e15424.

30. M. Devriese, S. Da Silva, M. Le Mene, et al., "Two-Field Resolution On-Call HLA Typing for Deceased Donors Using Nanopore Sequencing," *HLA* 103, no. 3 (2024): e15441.

31. J. Pugh, "The Current State of Nanopore Sequencing," *Methods in Molecular Biology* 2632 (2023): 3–14.

32. N. Anukul, P. Jenjaroenpun, C. Sirikul, et al., "Ultrarapid and High-Resolution HLA Class I Typing Using Transposase-Based Nanopore Sequencing Applied in Pharmacogenetic Testing," *Frontiers in Genetics* 14 (2023): 1213457.

33. C. Scarano, I. Veneruso, R. R. De Simone, G. Di Bonito, A. Secondino, and V. D'Argenio, "The Third-Generation Sequencing Challenge: Novel Insights for the Omic Sciences," *Biomolecules* 14, no. 5 (2024): 568.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** TGS-based high-resolution (four-field) genotyping of HLA class I loci.

20592310, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tan.71051 by University Of British Columbia, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License