Computational and Structural Biotechnology Journal 23 (2024) 1562–1571

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

A novel framework for human leukocyte antigen (HLA) genotyping using probe capture-based targeted next-generation sequencing and computational analysis

Sheng-Kai Lai [a,b], Allen Chilun Luo [b], I-Hsuan Chiu [b], Hui-Wen Chuang [c], Ting-Hsuan Chou [c], Tsung-Kai Hung [c], Jacob Shujui Hsu [c], Chien-Yu Chen [d], Wei-Shiung Yang [c,e,f], Ya-Chien Yang [g,h], Pei-Lung Chen [a,b,c,e,f,*]

[a] Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan
[b] Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan
[c] Graduate Institute of Medical Genomics and Proteomics, College of Medicine, National Taiwan University, Taipei, Taiwan
[d] Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan
[e] Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[f] Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[g] Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University, Taipei, Taiwan
[h] Department of Laboratory Medicine, National Taiwan University Hospital, Taipei, Taiwan

ARTICLE INFO

Keywords:
Human leukocyte antigen (HLA)
Next-generation sequencing (NGS)
Probe capture
Targeted sequencing
Human Pangenome Reference Consortium (HPRC)
IPD-IMGT/HLA
HISAT-genotype
HLA-VBseq
SpecHLA
T1K
QzType
DRAGEN

ABSTRACT

Human leukocyte antigen (HLA) genes play pivotal roles in numerous immunological applications. Given the immense number of polymorphisms, achieving accurate high-throughput HLA typing remains challenging. This study aimed to harness the human pan-genome reference consortium (HPRC) resources as a potential benchmark for HLA reference materials. We meticulously annotated specific four field-resolution alleles for 11 HLA genes (HLA-*A*, *-B*, *-C*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1*, *-DRB1*, *-DRB3*, *-DRB4* and *-DRB5*) from 44 high-quality HPRC personal genome assemblies. For sequencing, we crafted HLA-specific probes and conducted capture-based targeted sequencing of the genomic DNA of the HPRC cohort, ensuring focused and comprehensive coverage of the HLA region of interest. We used publicly available short-read whole-genome sequencing (WGS) data from identical samples to offer a comparative perspective. To decipher the vast amount of sequencing data, we employed seven distinct software tools: OptiType, HLA-VBseq, HISAT genotype, SpecHLA, T1K, QzType, and DRAGEN. Each tool offers unique capabilities and algorithms for HLA genotyping, allowing comprehensive analysis and validation of the results. We then compared these results with benchmarks derived from personal genome assemblies. Our findings present a comprehensive four-field-resolution HLA allele annotation for 44 HPRC samples. Significantly, our innovative targeted next-generation sequencing (NGS) approach for HLA genes showed superior accuracy compared with conventional short-read WGS. An integrated analysis involving QzType, T1K, and DRAGEN was developed, achieving 100% accuracy for all 11 HLA genes. In conclusion, our study highlighted the combination of targeted short-read sequencing and astute computational analysis as a robust approach for HLA genotyping. Furthermore, the HPRC cohort has emerged as a valuable assembly-based reference in this realm.

1. Introduction

The human leukocyte antigen (HLA) region located on chromosome 6p21 is a genomic complex that encodes HLA proteins. These proteins play pivotal roles in orchestrating both the innate and adaptive immune responses. Notably, the classical class I (*HLA-A*, *HLA-B*, and *HLA-C*) and class II (*HLA-DR*, *HLA-DQ*, and *HLA-DP*) HLA genes are the most polymorphic entities in the human genome [1]. Their significance is underscored by their association with pathogen resistance, graft rejection, and a plethora of over 100 immune-related diseases [2–4]. Many of

* Corresponding author at: Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan.
E-mail address: paylong@ntu.edu.tw (P.-L. Chen).

https://doi.org/10.1016/j.csbj.2024.03.030
Received 2 November 2023; Received in revised form 20 March 2024; Accepted 31 March 2024
Available online 2 April 2024

these conditions correlate with specific HLA alleles [5–11], emphasizing the importance of accurate HLA allele typing.

Traditional molecular HLA typing methodologies include polymerase chain reaction (PCR)-based procedures utilizing sequence-specific primers [12–15], sequence-specific oligonucleotide probes (SSO) [16–19], and sequencing-based typing (SBT) techniques [20]. Although SBT remains a trusted method because of its ability to discern HLA alleles at two-field to four-field resolutions [21,22], it is not devoid of limitations. Specifically, SBT suffers from inadequate throughput and occasionally produce ambiguous results. The advent of next-generation sequencing (NGS) offers a promising avenue to mitigate these challenges by capitalizing on the parallel sequencing capacity of numerous individual DNA molecules [23–29]. However, it's worth noting that most existing NGS-based methods are contingent upon long-range PCR of the HLA region. PCR-based methods offer the advantages of speed and low DNA input requirements. However, it is worth noting that most existing NGS-based methods are contingent on long-range PCR of the HLA region. PCR-based methods offer the advantages of speed and low DNA input requirements. However, caveats include the demand for relatively high-quality DNA, potential allele dropout in HLA class II genes, and the inability to detect novel alleles beyond the designed PCR range. In contrast, capture-based methods excel in probe coverage, extending to all regions except the HLA class II gene intron 1, significantly reducing the risk of allele dropout owing to the flexibility of probe capture. Moreover, capture-based methods have relatively low DNA quality requirements. However, their drawbacks include longer experimental processes and the need for higher quantities of DNA. Solutions in the form of probe capture-based targeted sequencing have been suggested to alleviate some PCR-associated concerns [30–33], as exemplified by the commercial product AlloSeq Tx (CareDx, Brisbane, CA, USA). Nonetheless, the precision of such capture-based NGS HLA typing methods remains largely unexplored, and the intricacies of probe designs and computational tools remain undisclosed.

In the contemporary landscape, multiple NGS data types are amenable to HLA typing analysis via a range of software tools, including HLAminer [34], HLAforest [35], ATHLATES[36], seq2HLA[37], OptiType[38], HLAssign[39], HLAreporter[40], Polysolver[41], HLA-VBseq [42], HLA*PRG[43], xHLA[44], HLA-HD[45], HLAscan[46], PHLAT [47], Kourami[48], STC-Seq[49], OncoHLA[50], HLA*LA[51], HISAT-genotype[52], arcasHLA[53], SpecHLA[54], T1K[55], DRAGEN [56], and QzType (TBG Biotechnology Corp., Taipei, Taiwan) (http://www.tbgbio.com/en/product/product_analysis_software) (supplementary-1). Several studies have addressed systematic comparisons [33,50,57–59], however a definitive assessment, particularly considering both probe capture-based and whole-genome sequencing (WGS) approaches, remain elusive. This study explored and compared multiple HLA genotyping tools using whole-genome and capture-based sequencing data, an aspect that has not been thoroughly addressed in the literature. The utilization of high-quality personal genome assembly annotations as a benchmark in our study was uncommon in prior research, further enhancing the novelty of our research. Additionally, we propose that human pan-genome reference consortium (HPRC) specimens sourced from diverse populations provide a valuable alternative to the International Histocompatibility Working Group (IHWG) reference panel and enrich the HLA reference database.

2. Materials and methods

2.1. HLA nomenclature

HLA allele names consist of four fields of digits separated by colons, each serving a distinct purpose. The first field signifies serologic or sequence similarity, the second field represents a unique designation based on the order of discovery, the third field indicates synonymous variation, and the fourth field denotes variation in the non-coding region. Notably, discrepancies in the digits of the first two fields of an allele imply one or more nucleotide substitutions resulting in differences in the encoded protein [60].

2.2. Utilization of international reference materials with high populational diversity

We acquired 44 specimens from the Coriell Institute for Medical Research (https://www.coriell.org/), which were sourced directly from the HPRC. These specimens were assembled using PacBio HiFi long-read technology, producing superior quality triophased personal assembly files that are accessible for reference [61].(https://github.com/human-pangenomics/HPP_Year1_Assemblies).

After comprehensive annotation, as detailed in the following section, these specimens offered benchmark HLA allele types for evaluating the precision and reliability of diverse HLA-focused wet- and dry-lab methodologies.

2.3. HLA allele annotation for the HPRC samples

The HPRC-derived personal genome assembly FASTQ files were subjected to HLA annotation using two distinct tools: gAIRR-annotate [62] and minimap2 [63]. The gAIRR-annotate procedure, which was originally designed to annotate adaptive immune receptor repertoire (AIRR) allele types, designates the PacBio HiFi personal assemblies as references and employs BWA-MEM (v0.7.17) for the alignment of known AIRR allele sequences to the references. The resulting BAM files were then subjected to a selection process to identify the V(D)J genes within the assemblies. Therefore, by substituting the AIRR allele sequences with all full-length HLA allele sequences provided by the IMGT/HLA database, we successfully determined the HLA genotypes in the assemblies.

Conversely, minimap2 (v2.24) functions by aligning the HLA coding sequences (CDSs) to the PacBio HiFi Trio Phased Assemblies. It then prioritizes the HLA genotypes possessing sequences that exhibit a perfect match or have the highest alignment score as HLA genotypes for each assembly.

When disparate allele annotations were offered by these two software tools, a subsequent step was initiated to extract relevant contigs from the assembly. The extracted contig(s) were further analyzed using the commercial software Geneious (Biomatters, Auckland, New Zealand), involving alignment with HLA allele sequences possessing the highest alignment scores to identify the correct result or potential novel loci. A thorough manual review ensured that all inconsistencies or ambiguities stemming from the software outputs were judiciously addressed and reconciled. This stringent approach vouches for the veracity and dependability of benchmark findings, amplifying the integrity of HLA genotyping evaluations, and augmenting our understanding of genetic diversity within this pivotal region.

2.4. Probe design and synthesis

Our probe design utilized an HLA-related sequence from the GRCh37 reference genome. These sequences were identical between GRCh37 and GRCh38, with differences only in the chromosomal position. In addition, we complemented the design with a complete set of known HLA sequences from the IPD-IMGT/HLA Database. (https://www.ebi.ac.uk/ipd/imgt/hla/download/). As of October 2023, the IMGT/HLA database has reported 26,341 HLA class I and 11,175 HLA class II alleles, totaling 37,516 alleles. We strategically positioned the probe-binding regions within stable intronic areas or at the exon-intron junctions of the HLA genes, as illustrated in Fig. 1. Additional probes were designed to accommodate alleles with significant sequence differences from the GRCh37 reference genome within the same genomic region. Caution was exercised to circumvent the incorporation of probes into repetitive elements, notably long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs) situated in introns, with special

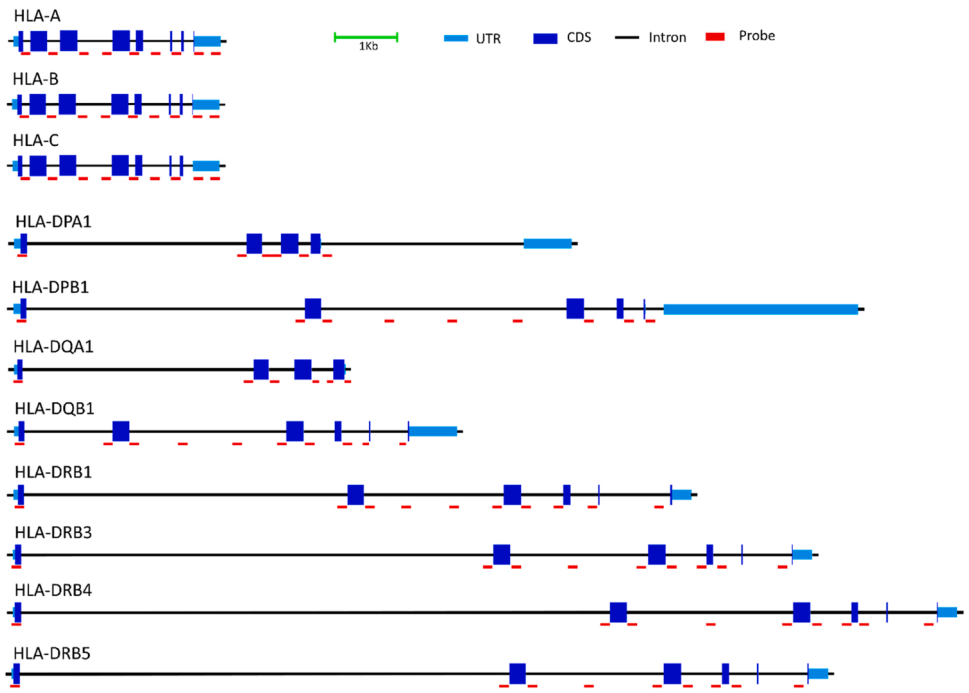

Fig. 1. A schematic representation highlighting the relative positions of HLA CDS and probes. This schematic is drawn to scale based on the lengths of the components depicted. Notably, HLA class II intron1 possess an abundance of repetitive elements (omitted for clarity), complicating probe design within these regions.

attention paid to HLA class II intron 1.

The constructed panel comprised 297 probes, providing coverage across the entirety of the class I HLA genes (*HLA-A*, *-B*, *-C*) and the majority (except for certain intron 1 regions) of the class II HLA genes (*HLA-DRB1*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1*, *-DRB3*, *-DRB4*, *-DRB5*). We selected these 11 HLA genes because of their clinical relevance in organ transplantation, which necessitates the matching of *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPB1*, *HLA-DQB1*, and *HLA-DRB1*. The remaining five genes have been associated with various immune-related diseases in numerous studies, rendering them of significant research value. The emphasis placed on these genes by most commercial kits and analytical software underscores their importance. It is important to recognize that certain parts of the HLA class II intron 1 were deliberately omitted from probe coverage because of the dense presence of repetitive elements. The length of intron 1 varies among HLA class II genes, from 1.1 kbp (*HLA-DQB1*) to 9.5 kbp (*HLA-DRB4*). Therefore, among the HLA class II genes, only *HLA-DQB1* has been fully sequenced. The total length of these 11 HLA genes was approximately 100 kbp; after subtracting the lengths of some intron 1 regions without the designed probes, the target region covered by the probes was approximately 57 kbp. Roche (Basel, Switzerland) supplied bespoke 120-bp probes. The bed file related to the probes is provided in Supplementary-2 file.

2.5. Library preparation, probe capture-based enrichment, and sequencing

To initiate the procedure, we took one microgram of genomic DNA and subjected it to fragmentation via the Covaris instrument (Covaris, Inc., Woburn, MA, USA), yielding fragments with an average length of 800 bps. This was followed by a series of preparatory steps, including end-repair, size selection, dA-tailing, and index ligation, using the TruSeq DNA PCR-free Low-Throughput Library Prep Kit (Illumina, Inc., San Diego, CA, USA).

Our probe-based capture methodology adhered to the KAPA HyperCap workflow v3.0 (Roche, Basel, Switzerland), utilizing the KAPA HyperCapture Reagent kit and KAPA HyperCapture Bead kit. Briefly, the libraries of fragmented genomic DNA, each uniquely indexed for every sample, were pooled before hybridizing them with Roche-customized capture probes.

This mixture was incubated at 55 ℃ for a duration of 16 h. The ensuing day entailed a purification phase, which was followed by amplification of the captured samples via ligation-mediated polymerase chain reaction (LM-PCR) over 14 cycles. After purification, the enriched libraries were quantified using the KAPA Library Quantification Kit (Roche, Basel, Switzerland) and sequenced. Sequencing was performed using the NovaSeq 6000 S4 Reagent Kit v1.5 (Illumina, Inc., San Diego, CA, USA), producing 150 bp paired-end reads.

This hybridization mix was incubated at 55 ℃ for a duration of 16 h. The next day entailed a clean-up phase followed by amplification of the captured sample via LM-PCR over 14 cycles. After clean-up, the target-enriched libraries were quantified using the KAPA Library Quantification Kit (Roche) and used for sequencing. Sequencing was performed using the NovaSeq 6000 S4 Reagent Kit v1.5 (Illumina, Inc.), yielding two–150 bp paired-end reads.

2.6. Computational analysis

Given that the primary objective of this study was not to undertake a comprehensive comparison of all available HLA analysis tools, we chose a subset of tools. The selection criteria were based on recent advancements, capacity to interpret a wide range of genes (both class I and class II), and typing resolution (three or four fields). Additionally, although OptiType is only capable of typing HLA class I, it is commonly used in comparative studies of HLA typing tools. Therefore, we included this in our analysis. In the end, We conducted a comparative evaluation of five open-source software tools: OptiType (v1.1.3), HLA-VBseq (v1), HISAT-genotype (v1.3.2), SpecHLA (v1.0.3), and T1K (v1.0.2), alongside two proprietary software platforms, DRAGEN (v4.2.4)[56] and QzType (v1.0.8665.38443) (TBG Biotechnology Corp., Taiwan) (http://www.tbgbio.com/en/product/product_analysis_software). Each software package was used in adherence to the official user guidelines. Raw FASTQ files were used as the inputs. These original FASTQ files were obtained from two distinct sources: one set originated from our capture-based sequencing experiments, whereas the other was derived

from publicly available WGS data (https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/). The results were classified into two- and three-field resolutions and then juxtaposed with the benchmark results to gauge their consistency (Fig. 2). The percentage of consistency for each tool was calculated as the number of correctly predicted alleles divided by the total number of alleles. The raw FASTQ files generated by capture-based sequencing can be downloaded from the following link (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA1078907&o=acc_s%3Aa); the relevant bash commands and typing results are provided in Supplementary-3 and − 4 files. Fig. 3 and Fig. 4.

Whenever there were inconsistencies in the reported HLA genotypes, we conducted a joint analysis using two (T1K and DRAGEN) or three (T1K, DRAGEN, and QzType) tools together with manual elucidation. When the joint analysis identified inconsistent HLA genotypes, we performed an alignment using BWA-MEM (v0.7.17) on the filtered FASTQ files from the HLA-relevant regions and corresponding full-length HLA allele sequences. Subsequently, we utilized Integrative Genomics Viewer (IGV) to meticulously inspect the SNP positions carried by inconsistent HLA, ensuring that the reads in that region support the results provided by the tool. An example explanation is provided in the Supplementary-5 file.

3. Results

3.1. HLA allele annotation of 44 HPRC personal assemblies

The benchmark results for the HPRC personal assemblies were generated through a consensus approach using two annotation methods. The gAIRR-annotate method leverages the full-length HLA gene sequences, thereby providing comprehensive four-field results. In contrast, minimap2 relies solely on the HLA CDS, limiting its resolution to three fields. When there was a discrepancy between the results of the two annotation techniques, long contiguous contigs containing the relevant HLA genes were extracted from the PacBio HiFi assembly. Subsequently, the commercial software Geneious (Biomatters, Auckland, New Zealand) was used to superimpose both sets of results onto the isolated contigs to verify accurate annotation.

During our benchmarking activities, we identified a significant limitation pertaining to the use of minimap2 for the alignment. It is imperative to highlight that minimap2's alignment approach occasionally overlooks small exons. Consequently, this limitation affected alignment scores and lengths, especially when contrasting *HLA-A** 02:01:01:01 and *HLA-A** 02:844. The omission of the last 5 bp small CDS of class I exon8 in the alignment indicated that the alignment parameters for *HLA-A** 02:01:01:01 were frequently less optimal than those for *HLA-A** 02:844. Such disparities in the alignment results occasionally prompted our designed filtering system to produce erroneous outputs for *HLA-A** 02:844.

The comprehensive four-field-resolution HLA genotypes of the 44 HPRC samples are provided in the Supplementary-6 file.

3.2. Evaluation of software tools for analyzing targeted/panel short-read NGS data

A total of 44 samples were compared in our analysis. Therefore, the percentage of consistency for each tool was calculated as the number of correctly predicted alleles divided by 88. HLA-VBseq, HISAT, and SpecHLA consistently underperformed compared to other software tools in the two- and three-field resolution categories (Table 1). Specifically, for HLA-VBseq at the three-field resolution, its accuracy rates were moderate, with *HLA-DQA1* registering the highest accuracy at 96.59% and *HLA-DRB1* the lowest at 73.86%. The two-field resolution showed marginal improvement, with *HLA-C* achieving a pinnacle accuracy of 97.73%. HISAT outperformed HLA-VBseq in both resolution categories, with high concordance rates for the majority of HLA genes. The standout genes were *HLA-A*, *HLA-C*, and *HLA-DQA1*, each reaching a concordance rate of 98.86% at two-field resolution. SpecHLA demonstrated the highest accuracy for *HLA-C* at 98.86%, whereas *HLA-DQA1* exhibited the lowest accuracy at 79.55% for the three-field resolutions. Transitioning to a two-field resolution showed a slight improvement, with *HLA-C* achieving a perfect accuracy of 100%. OptiType performed comparably to HISAT and SpecHLA for HLA class I typing and slightly outperformed HLA-VBseq. However, none of these genes achieved a 100% accuracy. Compared with T1K, QZtype, and DRAGEN, the performance of OptiType was slightly inferior.

T1K outperformed most other tools, showing remarkable concordance rates, especially at two-field resolution. *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, and *HLA-DRB5* showed accuracy of 100%. Even in the context of *HLA-DRB1* typing, T1K achieved an accuracy of 97.73%.

QzType, at a two-field resolution, displayed solid performance across the board, with most genes achieving 100% concordance, except for *HLA-DQB1* at 95.35%. At three-field resolution, while maintaining high accuracy, QzType displayed a slight decrease in *HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1* typing.

DRAGEN showed reasonably good accuracy at a two-field resolution, with *HLA-B*, *HLA-C*, *HLA-DQA1*, and *HLA-DQB1* achieving 100% accuracy. However, it is important to note that the highest resolution was

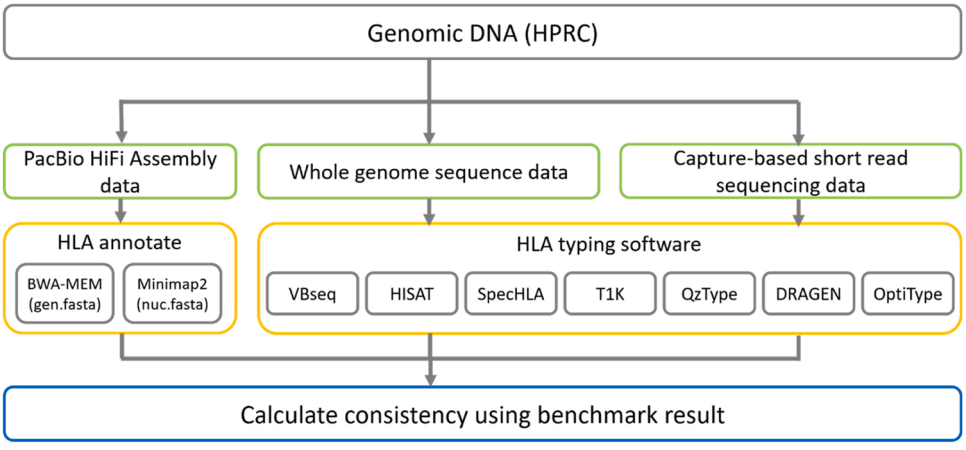

Fig. 2. Schematic representation of the experiment workflow illustrates the multi-faceted approach employed. Each sample from the HPRC dataset yields three distinct sets of data: firstly, benchmark results derived from the annotation of PacBio HiFi assembly; secondly, HLA typing results from the public available WGS data; thirdly, HLA typing results from capture-based sequencing data. These results, acquired using seven distinct computation tools, are then juxtaposed with the corresponding benchmarks. The percentage of consistency for each tool is the number of correctly predicted alleles divided by the total number of alleles.

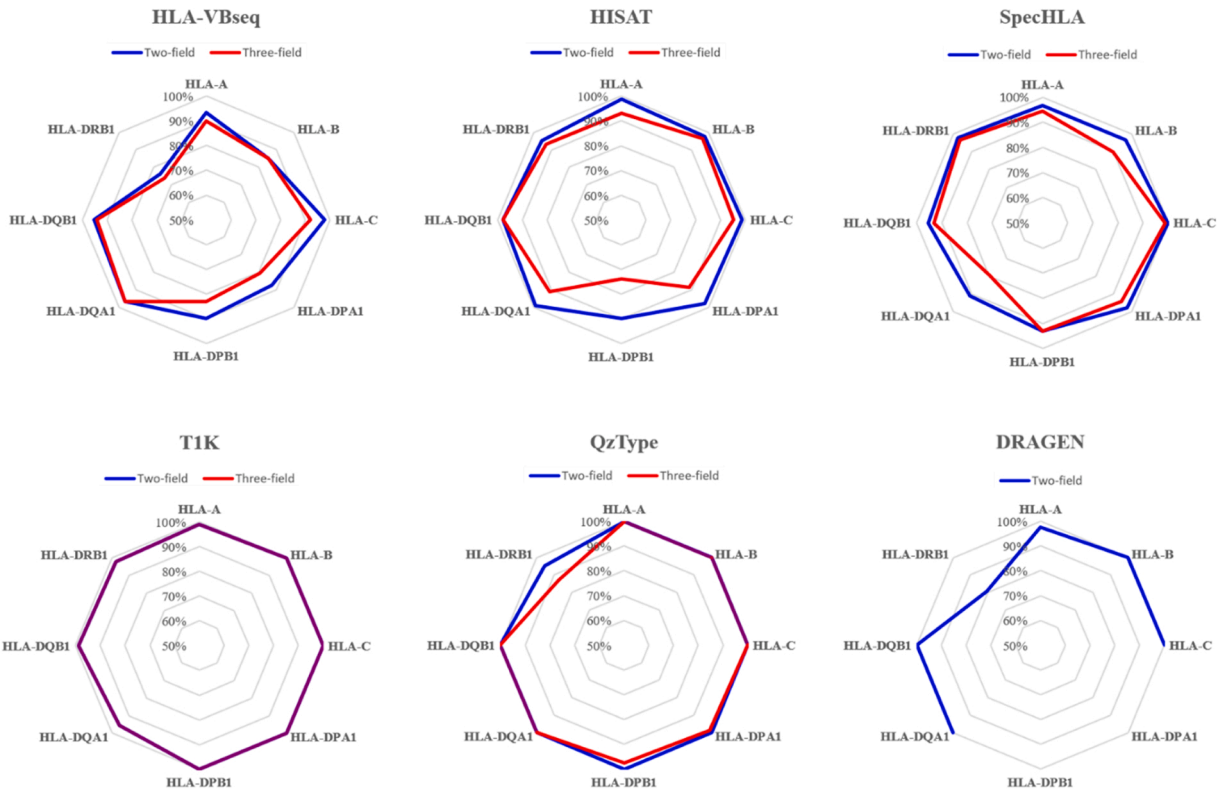

Fig. 3. Depiction of typing accuracy across eight HLA genes using five computational tools. This blue trajectory signifies the accuracy achieved at two-field resolution, whereas the red trajectory delineates the accuracy observed at three-field resolution. In the radar plot, each vertex represents one of the HLA genes, and the proximity of a trajectory to a regular octagon shape indicates higher accuracy. Notably, *HLA-DRB3/4/5* are not included in the plot because not all software tools can interpret these genes. As OptiType can only produce HLA class I typing at two-field resolution, it was not included in the depiction.

only two-field, and some HLA genes (*HLA-DPA1*, *-DPB1*, *-DRB3*, *-DRB4*, and *-DRB5*) could not be analyzed using DRAGEN.

3.3. Evaluation of software tools for analyzing WGS NGS data

QzType is unable to process WGS data; therefore, it was not included in the comparison of WGS methods.

It can be observed that the overall accuracy of WGS analysis is generally lower compared to the capture-based panel. Notably, in SpecHLA, two specimens were excluded from WGS analysis, thereby reducing the sample size to 42. This exclusion was due to insufficient reads in the HLA region, leading to an inability to determine the HLA genotype.

T1K demonstrated a higher accuracy at a three-field resolution, outperforming the other three software tools. At two-field resolution, T1K fell slightly behind DRAGEN, except for *HLA-DRB1*.

3.4. Joint analysis efficacy of software tools for WGS or targeted sequencing data

Among the available software tools, QzType, T1K, and DRAGEN exhibited commendable performance in the analysis of short-read NGS data, albeit with occasional discrepancies in HLA genotype determination. Considering these inconsistencies, we advocate a joint analytical framework. This approach involves meticulous curation of these inconsistent HLA genotypes, accomplished by realigning NGS reads to the purported alleles. Depending on the nature of the NGS procedures, such as WGS or panel-based targeted sequencing, the analysis can incorporate either a combination of tools (T1K and DRAGEN) or a trio (T1K, DRAGEN, and QzType).

Our empirical findings underscore the fact that a joint analysis employing these two tools considerably enhances accuracy. Remarkably, utilizing the three tools for joint analysis culminated in an impeccable 100% accuracy rate in both two-field and three-field typing across all 11 HLA genes, as detailed in Table 1, and illustrated in Fig. 5.

This table provides a systematic comparison of seven distinct HLA typing software tools, all of which utilize HPRC samples and differ in their field resolution for typing (either two- or three-field). This evaluation encompasses both the resolution and accuracy rates of specific HLA genes. DRAGEN produces results solely for two fields and is incapable of typing the *HLA-DPA1*, *-DPB1, -DRB3, -DRB4 and -DRB5* alleles. OptiType produces only two-field HLA class I results. _ joint2: joint analysis of T1K and DRAGEN. _joint3: joint analysis of T1K, DRAGEN, and QzType.

4. Discussion

The HLA region, comprising a vast array of genes pivotal for the immune response, is the subject of intensive research because of its direct implications in transplant immunology, pharmacogenomics, and disease predisposition. The accuracy with which HLA alleles are determined can critically influence clinical decisions hypersensitivity reactions. However, given the extensive polymorphism of the HLA region, achieving impeccable accuracy remains challenging. Recent advances in NGS technologies have ushered in a new era of HLA genotyping, promising higher resolution and accuracy than traditional methods. However, as our study highlights, even state-of-the-art NGS methodologies, when used in isolation, have limitations, particularly when applied to short-read data. Variances between software tools, each built on its own alignment and imputation strategies, can lead to sporadic discrepancies in HLA allele calls. This is particularly concerning in clinical settings where even minor inaccuracies can culminate in catastrophic consequences.

Probe-capture-based targeted sequencing has emerged as a

Fig. 4. Comparative accuracy analysis of four typing tools employing WGS and panel-based methods across eight HLA genes.The blue trajectory denotes the accuracy of two-field typing utilizing panel data, the light blue trajectory indicates the accuracy of two-field typing with WGS data, the red trajectory captures the accuracy of three-field typing using panel data, and the light red trajectory signifies the accuracy of three-field typing derived from WGS data. Notably, QzType is absent from the representation due to its incapability of analyzing WGS data.

promising alternative to WGS and long-range PCR-based NGS for HLA genotyping. The advantages of this approach are manifold. Compared with WGS, targeted sequencing offers an unparalleled depth of coverage for regions of interest, which is essential given the intricate polymorphic nature of the HLA system. The QzType analysis results confirmed 100% coverage in the HLA CDS region, with over 90% of the CDS region reaching a depth of 300 (supplementary-7). This ensures that even minute variants are not missed, thus bolstering confidence in the allele calls. Additionally, targeted sequencing circumvents the prodigious costs associated with sequencing an entire genome, making it a more cost-effective choice for laboratories with a singular focus on HLA or related regions. The reduction in sequenced regions also expedites both sequencing and data analysis times, allowing for a quicker turnaround, which is a paramount feature in time-sensitive clinical scenarios. In juxtaposition with long-range PCR-based NGS, probe capture-based methods offer a distinct advantage in terms of flexibility, allowing for the addition of probes as needed or the integration of other panels of interest, as well as providing broader coverage of sequencing regions. As a testament to this adaptability, our method permitted the integration of probes specific to *KIR* and *AIRR* alongside the primary HLA panel. Comprehensive sequencing methods and results for *AIRR* genes have been published in another paper by our team [62], while sequencing of *KIR* genes is currently underway. Both results exhibited robust coverage and sufficient depth. This amalgamated approach provides a comprehensive view of the immunogenomic landscape in a single sequencing run, opening avenues for intricate studies linking the *HLA*, *KIR*, and *AIRR* genes to immunological responses and disorders. Furthermore, a recurrent issue with long-range PCR is allelic dropout, a phenomenon in which one allele in a heterozygous sample is preferentially amplified over the other. Probe capture-based methods mitigate the issue of allele dropout as well as the necessity of high-quality and ample genomic DNA.

However, capture-based sequencing using short reads has several limitations. In the case of Class II Intron 1 s, which contain numerous repetitive elements such as SINEs or LINEs that are widely spread throughout the genome, designing probes to capture these regions may result in the capture of an excessive number of fragments unrelated to the HLA region, thereby increasing the complexity and time required for subsequent analyses. Moreover, the sequencing system used in this experiment employs 150 bp paired-end short read sequencing, and the obtained short reads may not fully span all repetitive elements because of the read length limit. Consequently, it is challenging to distinguish

Table 1
Evaluation of short-read NGS analysis accuracy for 44 HPRC samples.

HPRC_panel	N	Resolution	*HLA-A*	*HLA-B*	*HLA-C*	*HLA-DPA1*	*HLA-DPB1*	*HLA-DQA1*	*HLA-DQB1*	*HLA-DRB1*	*HLA-DRB3*	*HLA-DRB4*	*HLA-DRB5*
HLA-VBseq	44	Two-field	93.18%	85.23%	97.73%	87.50%	89.77%	96.59%	95.45%	76.14%	N/A	N/A	N/A
		Three-field	89.77%	85.23%	92.05%	80.68%	82.95%	96.59%	94.32%	73.86%	N/A	N/A	N/A
HISAT	44	Two-field	98.86%	97.73%	98.86%	97.73%	89.77%	98.86%	97.73%	95.45%	N/A	N/A	N/A
		Three-field	93.18%	96.59%	95.45%	88.64%	73.86%	90.91%	97.73%	93.18%	N/A	N/A	N/A
SpecHLA	44	Two-field	96.59%	96.59%	100.00%	97.73%	93.18%	90.91%	95.45%	97.73%	N/A	N/A	N/A
		Three-field	94.32%	89.77%	98.86%	94.32%	93.18%	79.55%	93.18%	96.59%	N/A	N/A	N/A
T1K	44	Two-field	98.86%	100.00%	100.00%	100.00%	100.00%	95.45%	98.86%	97.73%	97.73%	86.36%	100.00%
		Three-field	98.86%	100.00%	100.00%	100.00%	100.00%	95.45%	98.86%	97.73%	97.73%	87.50%	100.00%
QzType	44	Two-field	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	95.45%	100.00%	100.00%	98.84%
		Three-field	100.00%	100.00%	100.00%	98.86%	97.73%	100.00%	100.00%	87.50%	100.00%	100.00%	98.84%
DRAGEN	44	Two-field	97.73%	100.00%	100.00%	N/A	N/A	100.00%	100.00%	80.68%	N/A	N/A	N/A
OptiType	44	Two-field	98.86%	96.59%	98.86%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HPRC_WGS	N	Resolution	*HLA-A*	*HLA-B*	*HLA-C*	*HLA-DPA1*	*HLA-DPB1*	*HLA-DQA1*	*HLA-DQB1*	*HLA-DRB1*	*HLA-DRB3*	*HLA-DRB4*	*HLA-DRB5*
HLA-VBseq	44	Two-field	76.14%	84.09%	94.32%	95.45%	94.32%	92.05%	93.18%	77.27%	N/A	N/A	N/A
		Three-field	73.86%	84.09%	89.77%	80.68%	87.50%	90.91%	92.05%	75.00%	N/A	N/A	N/A
HISAT	44	Two-field	93.18%	96.59%	98.86%	97.73%	84.09%	97.73%	95.45%	78.41%	N/A	N/A	N/A
		Three-field	86.36%	95.45%	95.45%	87.50%	67.05%	90.91%	94.32%	76.14%	N/A	N/A	N/A
SpecHLA	42	Two-field	95.24%	92.86%	100.00%	97.62%	88.10%	97.62%	97.62%	97.62%	N/A	N/A	N/A
		Three-field	91.67%	85.71%	98.81%	91.67%	88.10%	96.43%	96.43%	97.62%	N/A	N/A	N/A
T1K	44	Two-field	98.86%	100.00%	98.86%	97.73%	98.86%	100.00%	95.45%	93.18%	97.73%	72.73%	75.00%
		Three-field	98.86%	100.00%	98.86%	96.59%	98.86%	100.00%	95.45%	92.05%	97.73%	72.73%	75.00%
DRAGEN	44	Two-field	98.86%	100.00%	100.00%	N/A	N/A	100.00%	100.00%	71.59%	N/A	N/A	N/A
OptiType	44	Two-field	95.45%	95.45%	98.86%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Joint analysis	N	Resolution	*HLA-A*	*HLA-B*	*HLA-C*	*HLA-DPA1*	*HLA-DPB1*	*HLA-DQA1*	*HLA-DQB1*	*HLA-DRB1*	*HLA-DRB3*	*HLA-DRB4*	*HLA-DRB5*
WGS_joint2	44	Two-field	98.86%	100.00%	100.00%	N/A	N/A	100.00%	100.00%	96.59%	N/A	N/A	N/A
Panel_joint2	44	Two-field	98.86%	100.00%	100.00%	N/A	N/A	100.00%	100.00%	97.73%	N/A	N/A	N/A
Panel_joint3	44	Two-field	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
		Three-field	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

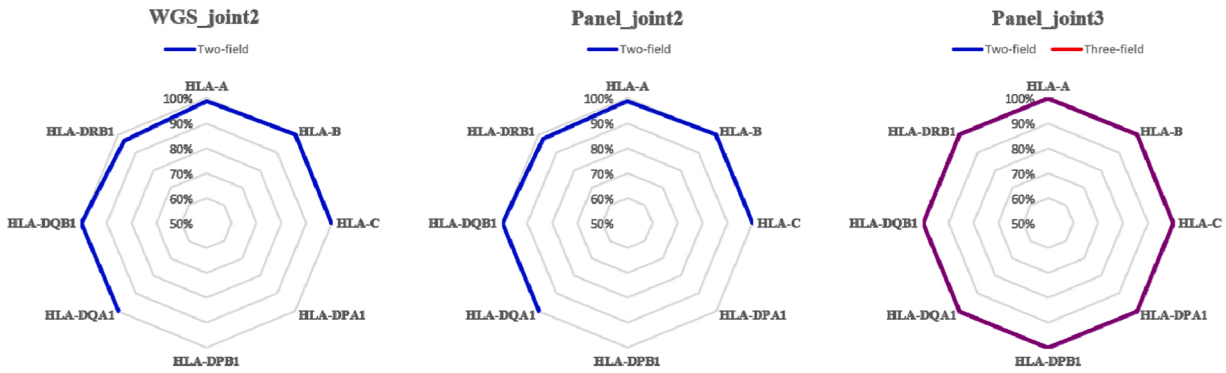

Fig. 5. Performance of joint analysis of software tools for WGS or targeted sequencing data "_joint2" refers to the joint analysis using T1K and DRAGEN. However, for *HLA-DPA1* and *HLA-DPB1*, only the results from T1K are available, so the lines representing these genes are not displayed in the plot. Conversely, "_joint3" pertains to the joint analysis of T1K, DRAGEN, and QzType. For *HLA-DPA1* and *HLA-DPB1* in this analysis, the comparison is solely between T1K and QzType.

which reads originate from Class II Intron 1 repeat regions from other genomic regions. This limitation restricts the accuracy of the capture method in reflecting the diversity of Class II Intron 1 s. However, combining the capture method with long-read NGS technologies may help circumvent this issue. Regarding computational analysis, the current limitation is the absence of a single software capable of achieving 100% accuracy for all 11 HLA genes, particularly for the relatively complex *HLA-DRB1*. The current approach for addressing this challenge involves combining multiple well-performing software tools with manual curation.

Although the benefits of probe capture-based targeted sequencing are pronounced, it is paramount to emphasize that achieving 100% HLA genotyping accuracy is *sine qua non*, especially in clinical practice. The repercussions of suboptimal accuracy are daunting and range from graft rejection in transplant scenarios to adverse drug reactions in pharmacogenomic applications. The repercussions of imperfect HLA typing accuracy may significantly increase the risk of organ transplant rejection, adversely affect the recipient's recovery process, and potentially lead to various post-transplant complications, jeopardizing overall patient outcomes. To address these concerns, our study pivoted on an integrative software analysis approach. By harnessing the combined strengths of multiple bioinformatics tools, we demonstrate the superiority of a multi-tool framework for mitigating errors. This was particularly evident when the individual tools yielded discordant results. The integration of the results from QzType, T1K, and DRAGEN led to an unparalleled 100% accuracy rate across all evaluated HLA genes. This layered redundant approach amplifies the confidence in allele calls and sets a benchmark for future NGS-based HLA studies.

The crux of HLA genotyping does not just rest on sequencing or analysis methodologies, but extends to the reference materials employed. Traditional references, such as those provided by the IHWG, are esteemed for their comprehensive sample repositories and diverse population representations, thus serving as longstanding benchmarks within the HLA community. Nonetheless, these studies are not devoid of caveats. They encompass inconsistencies in resolution within the provided standard genotypes, incomplete coverage of HLA genes, potential inaccuracies, and a propensity for the selection of class II-focused cell lines with homozygous alleles. In this context, we championed the use of HPRC resources. The superior-quality tri-phased personal assembly files from the HPRC, obtained using PacBio HiFi long-read technology, resolve standard genotypes with four-field resolution, establish haplotype information, and offer an alternative to the IHWG standard, contributing substantially to the diversity of the HLA community. Although these assemblies benefit from diverse technologies and high coverage depths, the possibility of assembly errors remains low. However, rare assembly errors cannot be overlooked, because a single-base difference can distinguish one HLA allele from another. Consequently, assembly errors may compromise the benchmark results. Nevertheless, our targeted sequencing and analysis operations, along with the annotation of HPRC personal assemblies, were conducted independently, thereby minimizing the chance of shared errors. In our analysis, the results from both the probe capture-based and WGS approaches were compared against the annotated benchmark. For each of the 44 samples, at least two – three tools consistently yielded results that were compatible with the benchmark, affirming the absence of assembly errors in the HLA region. Given the consistent release of high-quality data by the HPRC, the transition to using them as benchmark materials is both logical and inevitable.

While utilizing HLA resources from the 1000 Genomes Project, similar to the T1K's study [55], could indeed expand the test cohort and improve result representativeness, we opted not to pursue this avenue. This decision was based on concerns regarding the lower reliability and resolution of the ground truth mentioned in the T1K's paper, as opposed to our utilization of personal genome assembly annotations. Moreover, the data available from the 1000 Genomes Project did not align with our focus on WGS or probe capture-based NGS, which further influenced our choice. However, with the prospect of receiving further high-quality data from the HPRC, we are eager to evaluate the efficacy of our method using samples sourced from the 1000 Genomes Project. This will enable us to comprehensively evaluate its performance. Assessing our method across various sample cohorts will enrich our understanding of its adaptability and efficacy in diverse populations.

5. Conclusion

HLA genotyping is complex, both technically and biologically. Our study serves as a testament to the efficacy of probe capture-based targeted sequencing, especially when complemented by an integrative software analysis approach. The flexibility of incorporating probes for diverse regions, from *HLA* to *KIR* and *AIRR* genes, further accentuates the versatility of this method. With the continuous evolution of sequencing technologies and bioinformatics tools, the pursuit of accurate HLA genotyping remains an ongoing journey. Our findings pave the way for future research in this ever-evolving field. In addition, the implementation of our tailored annotation workflow further reinforces the accuracy of HLA typing when utilizing long-read sequencing data from internationally recognized reference samples. The steady release of high-quality trio phased assemblies by HPRC signifies a potential shift in the landscape of standard reference materials, with HPRC samples poised to assume a prominent role. These findings not only affirm the reliability of our methodology but also contribute crucial resources to propel future research endeavors in the field of HLA-related studies.

CRediT authorship contribution statement

Chien-Yu Chen: Resources, Writing – review & editing. **Wei-Shiung**

Yang: Resources, Writing – review & editing. **YA-CHIEN YANG:** Resources, Writing – review & editing. **Pei-Lung Chen:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **Sheng-Kai Lai:** Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Allen Chilun Luo:** Data curation, Methodology. **I-Hsuan Chiu:** Data curation, Methodology. **Hui-Wen Chuang:** Formal analysis, Software. **Ting-Hsuan Chou:** Formal analysis, Software. **Tsung-Kai Hung:** Software, Validation. **Jacob shujui Hsu:** Resources, Software.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We thank to National Core Facility for Biopharmaceuticals (NCFB, MOST 108–2319-B-492−001) for support and National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources. Also, we thank to Taiwan AI Labs for providing online website computational resources, TBG Diagnostics Limited for providing license of QzType. This study was supported by Taiwan Ministry of Science and Technology grants (MOST108–2622-B-002–004-CC2, MOST 108–2314-B-002–069-MY3).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.03.030.

References

[1] Vandiedonck C, Knight JC. The human Major Histocompatibility Complex as a paradigm in genomics research. Brief Funct Genom 2009;8(5):379–94.
[2] Trowsdale J. The MHC, disease and selection. Immunol Lett 2011;137(1):1–8.
[3] Trowsdale J, Knight JC. Major Histocompatibility Complex Genomics and Human Disease. Annu Rev Genom Hum Genet 2013;14(1):301–23.
[4] Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 2004;64(6):631–49.
[5] Chen B, et al. Role of HLA-B27 in the pathogenesis of ankylosing spondylitis (Review). Mol Med Rep 2017;15(4):1943–51.
[6] Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun 2015;64:13–25.
[7] Chen P-L, et al. Genetic determinants of antithyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. Nat Commun 2015;6(1):7633.
[8] Chen P-L, et al. Comprehensive Genotyping in Two Homogeneous Graves' Disease Samples Reveals Major and Novel HLA Association Alleles. PLOS ONE 2011;6(1): e16635.
[9] Furukawa H, et al. Human leukocyte antigens and systemic lupus erythematosus: a protective role for the HLA-DR6 alleles DRB1*13:02 and *14:03. PloS One 2014;9 (2). e87792-e87792.
[10] Niu, Z., Y. Zhang P Fau - Tong, and Y. Tong, Value of HLA-DR genotype in systemic lupus erythematosus and lupus nephritis: a meta-analysis. (1756–185X (Electronic)).
[11] Furukawa H, et al. Human Leukocyte Antigens and Systemic Lupus Erythematosus: A Protective Role for the HLA-DR6 Alleles DRB1*13:02 and *14:03. PLOS ONE 2014;9(2):e87792.
[12] Olerup O, Aldener A, Fogdell A. HLA-DQB1 and -DQA1 typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 h. Tissue Antigens 1993;41(3):119–34.
[13] Bunce M, Taylor CJ, Welsh KI. Rapid HLA-DQB typing by eight polymerase chain reaction amplifications with sequence-specific primers (PCR-SSP). Hum Immunol 1993;37(4):201–6.
[14] Sadler AM, et al. Low-resolution DNA typing for HLA-B using sequence-specific primers in allele- or group-specific ARMS/PCR. Tissue Antigens 1994;44(3): 148–54.
[15] Bunce M, et al. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens 1995;46(5):355–67.
[16] Wordsworth BP, et al. HLA-DR typing using DNA amplification by the polymerase chain reaction and sequential hybridization to sequence-specific oligonucleotide probes. Immunogenetics 1990;32(6):413–8.
[17] Scharf SJ, Griffith RL, Erlich HA. Rapid typing of DNA sequence polymorphism at the HLA-DRB1 locus using the polymerase chain reaction and nonradioactive oligonucleotide probes. Hum Immunol 1991;30(3):190–201.
[18] Kimura A. Eleventh International Histocompatibility Workshop Reference Protocol for the HLA DNA-Typing Technique. HLA 1991:397–419.
[19] Gao X, Jakobsen IB, Serjeantson SW. Characterization of the HLA-A polymorphism by locus-specific polymerase chain reaction amplification and oligonucleotide hybridization. Hum Immunol 1994;41(4):267–79.
[20] Nowak J, Mika-Witkowska R, Graczyk-Pol E. Genetic Methods of HLA Typing. In: Witt M, Dawidowska M, Szczepanski T, editors. in *Molecular Aspects of Hematologic Malignancies: Diagnostic Tools and Clinical Applications*. Springer Berlin Heidelberg: Berlin, Heidelberg; 2012. p. 325–39.
[21] Smith L, Fidler S. *HLS Typing Technologies*, in *The HLA Complex in Biology and Medicine: A Resource Book*. In: Mehra NK, editor. Jaypee Brothers Medical Publishers (P) Ltd: New Delhi, India; 2010. p. 175–87.
[22] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74(12):5463–7.
[23] De Santis D, et al. 16(th) IHIW: review of HLA typing by NGS. Int J Immunogenet 2013;40(1):72–6.
[24] Bentley G, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens 2009;74(5):393–403.
[25] Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet 2010; 11(1):31–46.
[26] Gabriel C, et al. HLA typing by next-generation sequencing - getting closer to reality. Tissue Antigens 2014;83(2):65–75.
[27] Jiao Y, et al. High-sensitivity HLA typing by Saturated Tiling Capture Sequencing (STC-Seq). BMC Genom 2018;19(1). 50-50.
[28] Carapito R, Radosavljevic M, Bahram S. Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. Hum Immunol 2016;77(11):1016–23.
[29] Liu L, et al. Comparison of next-generation sequencing systems. BioMed Res Int 2012;2012.
[30] Brown NK, et al. Reduced PCR-generated errors from a hybrid capture-based NGS assay for HLA typing. Hum Immunol 2021;82(4):296–301.
[31] Wittig M, et al. High-Resolution HLA-Typing by Next-Generation Sequencing of Randomly Fragmented Target DNA. Methods Mol Biol 2018;1802:63–88.
[32] Inoko H, et al. OR15 New NGS HLA typing by targeted enrichment procedure, capture method. Hum Immunol 2016;77:13.
[33] Liu P, et al. Benchmarking the Human Leukocyte Antigen Typing Performance of Three Assays and Seven Next-Generation Sequencing-Based Algorithms. Front Immunol 2021;12:652258.
[34] Warren RL, et al. Derivation of HLA types from shotgun sequence datasets. Genome Med 2012;4(12):95.
[35] Kim HJ, Pourmand N. HLA typing from RNA-seq data using hierarchical read weighting [corrected]. PLoS One 2013;8(6):e67885.
[36] Liu C, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. Nucleic Acids Res 2013;41(14):e142.
[37] Boegel S, et al. HLA typing from RNA-Seq sequence reads. Genome Med 2012;4 (12):102.
[38] Szolek A, et al. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics 2014;30(23):3310–6.
[39] Wittig M, et al. Development of a high-resolution NGS-based HLA-typing and analysis pipeline. Nucleic Acids Res 2015;43(11):e70.
[40] Huang Y, et al. HLAreporter: a tool for HLA typing from next generation sequencing data. Genome Med 2015;7(1):25.
[41] Shukla SA, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol 2015;33(11):1152–8.
[42] Nariai N, et al. HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data. BMC Genom 2015;16(Suppl 2). S7-S7.
[43] Dilthey AT, et al. High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. PLoS Comput Biol 2016;12 (10):e1005151.
[44] Xie C, et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proc Natl Acad Sci USA 2017;114(30):8059–64.
[45] Kawaguchi S, et al. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat 2017;38(7):788–97.
[46] Ka S, et al. HLAscan: genotyping of the HLA region using next-generation sequencing data. BMC Bioinforma 2017;18(1):258.
[47] Bai Y, Wang D, Fury W. PHLAT: inference of High-Resolution HLA Types from RNA and Whole Exome Sequencing. Methods Mol Biol 2018;1802:193–201.
[48] Lee H, Kingsford C. Kourami: graph-guided assembly for novel human leukocyte antigen allele discovery. Genome Biol 2018;19(1):16.
[49] Jiao Y, et al. High-sensitivity HLA typing by Saturated Tiling Capture Sequencing (STC-Seq). BMC Genom 2018;19(1):50.
[50] Sverchkova A, et al. Improved HLA typing of Class I and Class II alleles from next-generation sequencing data. HLA 2019;94(6):504–13.
[51] Dilthey AT, et al. HLA*LA-HLA typing from linearly projected graph alignments. Bioinformatics 2019;35(21):4394–6.
[52] Kim D, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37(8):907–15.
[53] Orenbuch R, et al. arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics 2020;36(1):33–40.
[54] Wang S, et al. SpecHLA enables full-resolution HLA typing from sequencing data. Cell Rep Methods 2023;3(9):100589.
[55] Song L, et al. Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. Genome Res 2023;33(6):923–31.

[56] Miller NA, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med 2015;7:100.
[57] Larjo A, et al. Accuracy of programs for the determination of human leukocyte antigen alleles from next-generation sequencing data. Front Immunol 2017;8:1815.
[58] Bauer DC, et al. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. Brief Bioinf 2018;19(2):179–87.
[59] Thuesen NH, et al. Benchmarking freely available HLA typing algorithms across varying genes, coverages and typing resolutions. Front Immunol 2022;13:987655.
[60] Hurley CK. Naming HLA diversity: A review of HLA nomenclature. Hum Immunol 2021;82(7):457–65.
[61] Liao WW, et al. A draft human pangenome reference. Nature 2023;617(7960):312–24.
[62] Lin MJ, et al. Profiling genes encoding the adaptive immune receptor repertoire with gAIRR Suite. Front Immunol 2022;13:922513.
[63] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34(18):3094–100.