Human Immunology 82 (2021) 812–819

Contents lists available at ScienceDirect

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

Research article

Challenges in the application of NGS in the clinical laboratory

Yuxin Yin, Carrie Butler, Qiuheng Zhang *

UCLA Immunogenetics Center, Department of Pathology & Laboratory Medicine, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 30 November 2020
Revised 25 February 2021
Accepted 29 March 2021
Available online 21 April 2021

Keywords:
Next-generation sequencing
Clinical
Immunology
Genomics
Precision medicine

ABSTRACT

Next-generation sequencing (NGS), also known as massively parallel sequencing, has revolutionized genomic research. The current advances in NGS technology make it possible to provide high resolution, high throughput HLA typing in clinical laboratories. The focus of this review is on the recent development and implementation of NGS in clinical laboratories. Here, we examine the critical role of NGS technologies in clinical immunology for HLA genotyping. Two major NGS platforms (Illumina and Ion Torrent) are characterized including NGS library preparation, data analysis, and validation. Challenges of NGS implementation in the clinical laboratory are also discussed, including sequencing error rate, bioinformatics, result interpretation, analytic sensitivity, as well as large data storage. This review aims to promote the broader applications of NGS technology in clinical laboratories and advocate for the novel applications of NGS to drive future research.

1. Introduction

Since Sanger sequencing [1] was first developed in 1977, it has been considered the "Gold Standard" sequencing method in the clinical laboratory. However, Sanger sequencing is limited by its low throughput and high cost. Next-generation sequencing (NGS) is also known as "massively parallel sequencing". It refers to the high-throughput sequencing technology which enables a large number (millions to billions) of DNA templates to be sequenced in parallel, thereby generating an unprecedented amount of genetic information in a single run [2]. The benefits of NGS include higher sequencing capacity, ability to multiplex samples, and lower sequencing cost if batched. The advances in NGS technologies have made high resolution, high throughput, and unambiguous HLA typing possible compared to traditional methods [3–14].

The human leukocyte antigen (HLA) genes are the most polymorphic genes in the human genome. HLA has many clinical implications in allogeneic transplantation, inflammation, and autoimmune diseases. The massive sequencing capacity of NGS permits its broad research and clinical applications in different areas of immunology and the field of transplantation. The degree of HLA matching between transplant recipients and donors significantly affects the outcomes of both solid organ transplant and hematopoietic stem cell transplantation (HSCT) [15–17]. The implementation of NGS technologies in routine clinical work permits in-depth characterization of the full length of HLA gene sequences; thereby provides optimal HLA matching of donor-recipient pairs for organ transplantation. In renal transplantation, a growing body of literature has highlighted the association between a greater number of HLA molecular mismatches including eplets or small groups of polymorphic amino acid mismatches, and adverse allograft outcomes which require high-resolution HLA typing of both donors and recipients [18]. Despite the matching, the expression levels of HLA genes can also have crucial roles in transplantation. Petersdorf et al. demonstrated those *HLA-DPB1* alleles with rs9277534G at 3′UTR are associated with high expression *HLA-DPB1* alleles and rs9277534A are associated with low expression allele. Recipients with rs9277534A-linked *DPB1* who received donors that carried rs9277534G-linked *DPB1* had higher risks of acute graft-versus-host disease (GVHD) [19]. In solid organ transplantation, monitoring donor-derived cell-free DNA (dd-cfDNA) after transplantation can greatly help the early diagnosis of allograft rejection. Non-invasive NGS assays have been developed to quantify dd-cfDNA without the need for prior genotyping of the donor and recipient and replace invasive biopsies [20,21]. In HSCT, NGS has also been reported to be utilized for monitoring chimerism post HSCT. The advances in NGS make it possible to sequence as many single nucleoids polymorphisms (SNP) as possible. Additionally, minimal residual disease (MRD) detection can also be incorporated into the same NGS run to monitor disease relapse [22]. Moreover, NGS permits sequencing MICA, MICB [23], and KIR [24,25] to study their roles in transplantation. Using NGS TCR repertoire sequencing, Sykes and colleagues have provided evi-

* Corresponding author.
E-mail address: jqzhang@mednet.ucla.edu (Q. Zhang).

https://doi.org/10.1016/j.humimm.2021.03.011
0198-8859/

dence for T cell clonal deletion as a mechanism of allograft tolerance in patients who received combined kidney and bone marrow transplants [26].

Traditional HLA methods, such as SBT (sequencing-based typing), SSO (sequence specific oligonucleotide probes), and SSP (sequence specific primer) usually only cover exons 2–4 for HLA class I and exons 2–3 for class II. Due to incomplete coverage of the HLA genes, phasing (cis–trans) ambiguities and ambiguities caused by polymorphic nucleotide difference(s) outside sequencing regions are common. HLA typing ambiguities were reported ranging from 24% to 53% depending on the HLA loci and the version of the HLA database used [27–29]. At UCLA Immunogenetics center, we also experienced > 50% of ambiguities using Sanger sequencing before we switched to NGS in 2014. Resolving HLA typing ambiguities in HLA laboratories is labor-intensive, costly, and significantly increases the turnaround time. The current NGS typing approaches, however, cover all exons and most introns of the HLA genes, therefore significantly reducing these ambiguities. Despite these great benefits of NGS, the implementation of the technology in clinical diagnostic laboratories has proceeded slowly. The challenges of the implementation of NGS include the high initial installation cost, challenges in training due to the complex workflow, longer turnaround time compared to existing methods, and steep learning curve. As sequencing technology continues to evolve, third-generation sequencing technology has also advanced significantly. Although third-generation sequencing provides a better option for *de novo* sequence assembly, currently the high error rate limits its application in the clinical laboratory. Therefore, this review will focus on the short-read second-generation sequencing of HLA.

2. NGS platform comparison

One of the first decisions a clinical laboratory needs to make is what type of NGS platform to purchase. There are a wide variety of instruments on the market that differ by installation cost, chemistry, read length, sequencing capacity, instrument footprint, and sequencing time. Currently, Illumina and Ion Torrent are the two major NGS platforms on the market. Both methods sequence DNA by synthesis. In the Illumina sequencing process, DNA is sequenced by incorporation of the fluorescently labeled, reversibly terminating nucleotides. Before the incorporation of the next set of nucleotides, chemical cleavage of the terminator occurs to enable the next sequencing cycle. After each incorporation, a charge-coupled device (CCD) for four-, two-channel chemistry or a complementary metal–oxidesemiconductor (CMOS) for one-channel chemistry will record the incorporated nucleotide. In the four-channel system, each of the four nucleotides is labeled with a separate dye. In each amplification cycle, each dye signal is imaged by using two lasers and four filters. The two-channel method uses two fluorescent dyes and two images to determine four nucleotides. For example, clusters seen in red or green images are interpreted as C and T bases, respectively. Clusters observed in both red and green images are flagged as A bases (appearing as yellow clusters), while unlabeled clusters are identified as G bases. In the one-channel chemistry, 4 types of nucleotides are added to the sequencing reaction with only A bases and T bases labeled with fluorescence. During the first imaging step, the light emission from DNA incorporation is recorded by the CMOS sensor. The second chemistry step removes the fluorescent label from A bases and adds a fluorescent label to C bases. The second image is also recorded by the CMOS sensor. In both chemistry steps, G bases are unlabeled. The combination of image 1 and image 2 are used to identify which bases are incorporated at each sequencing cycle (A = on/off; C = off/on; T = on/on; G = off/off). The one- and two-channel systems provide faster and more affordable sequencing methods and make NGS more applicable in clinical laboratories.

Unlike Illumina, multiple nucleotides can be added during a single sequencing cycle, in the Ion Torrent sequencing, a single type of nucleotides is added in each sequencing round. When a nucleotide is added to the DNA template, hydrogen ions are released. The release of hydrogen ions will change the pH of the solution and convert it into digital information (0, 1) by semiconductors underneath the well. Unlike Illumina, multiple nucleotides of the same type can be incorporated during a single sequencing cycle in Ion Torrent technology. Homopolymers are consecutive repetitions of a base in a string. For example, when the incorporation of six consecutive adenines (AAAAAA-homopolymer run) to a DNA template is recorded as six signals. Due to this sequencing feature, the Ion Torrent platform is prone to insertion and deletion mutations (indel) errors, since the six consecutive adenines can be read as seven (insertion) or five (deletion). Despite these pros and cons, both platforms provide accurate and reliable sequencing data for clinical use when there are enough coverage and adequate quality control.

Commonly used sequencers from Illumina and ThermoFisher are listed in Table 1. Laboratories may choose the right instrument based on cost (instrumental and reagents), footprint, the sample volume of the laboratory, and sequencing time to determine if turnaround time can be met. Illumina currently offers four-benchtop and one-production scale instruments. The benchtop sequencers include iSeq100, MiniSeq, MiSeq, NexSeq, and the production scale sequencer includes NovaSeq6000. The advantages of Illumina sequencers include high throughput, high accuracy sequencing data. Additionally, paired-end sequencing also provides a greater ability to identify the relative positions of polymorphism sites and characterize repetitive sequence regions by filling in gaps of consensus sequence at both ends to achieve complete overall coverage. A disadvantage of the Illumina platform is the set-up cost is higher than the Ion Torrent system. The Ion Torrent series of machines including the Ion PGM, Ion S5/S5XL, and Ion Proton (ThermoFisher) (Table 1). The advantages of the Ion Torrent technology include relatively lower sequencing cost, faster sequencing speed. The drawbacks of the Ion Torrent include lower throughput and higher sequencing error particularly in sequences with homopolymer runs. However, with enough sequencing coverages, both platforms provide accurate HLA typing results (Table 2) [4,5,7–13,29–42]. In our laboratory, we successfully detected *A*24:11 N* which has a cytosine (C) insert at exon 4 resulting in an 8C homopolymer run using the Ion Torrent platform. Accurate numbers of 16 thymine (T) in intron 2 of *DQB1*02:01:01:01* and 18 T in *DQB1*02:02:01:01* were also correctly identified with adequate coverage. However, the Ion Torrent sequencing platform had difficulty accurately elucidating 27 T in *DRB1*15:02:01:01* versus 26 T in *DRB1*15:02:01:02* in intron 5 resulting in typing ambiguities in our hands. Nonetheless, since most of the homopolymer runs located in exons are less than 10 repeats, therefore it does not affect the accuracy of the HLA typing up to the 3rd field.

Table 2 summarizes recent publications on HLA genotyping using Illumina and Ion Torrent platforms. These platforms differ by their sequencing capacity, sequencing time, chemistry, footprint, and price. The Illumina sequencing platform output ranges from less than 1 Gb (iSeq100) to over 6000 Gb (NovaSeq6000) per run, and the number of reads has increased from millions to billions. The run time of the Illumina sequencers ranges from 5 to 55 h depending on the read length and chemistry. As for Ion Torrent sequencers, the output ranges from 20 Mb (Ion PGM) to 10 Gb (Ion Proton) per run, and the reads have increased from 0.2 million to around 80 million. The run time for the Ion Torrent sequencers is much shorter (2–7 h depends on the read length and chemistry for sequencing) than the Illumina platform.

Table 1
List of main NGS instruments.

	Illumina					ThermoFisher		
								
	iSeq100	MiniSeq	MiSeq	NextSeq	NovaSeq6000	Ion PGM	Ion S5/S5XL	Ion Proton
Output Range	1.2 Gb	1.8-7.5 Gb	0.3-15 Gb	20-120 Gb	134-6000 Gb	20 Mb-2 Gb	1.2-8 Gb	10 Gb
Run Time	9-19 hr	24 hr	5-55 hr	11-29 hr	24-44 hr	2-7 hr	4-5 hr	2-4 hr
Reads per Run	4 million	8-25 million	1-25 million	130-400 million	Up to 20 billion	0.2- 6 million	3-80 million	60-80 million
Read Length	2 x 150 bp	2 x 150 bp	2 x 300 bp	2 x 150 bp	2 x 150 bp	200-400 bp	200-400 bp	200 bp
Instrument Cost	~$19.9K	~$50K	~$100K	~$250K	~$985K	~$50K	~$65K/157K	~$149K

Illumina sequencing supports the sequencing of templates from both ends (paired-end sequencing). Illumina offers 100 bp x2, 150 bp x2, 250 bp x2, and 300 bp x2 paired-end sequencing chemistry. During paired-end sequencing, two sequencing reads are generated for each library molecule from both ends (e.g., 150 bp x2). Comparing to single-end sequencing, paired-end sequencing has the advantage of detecting genomic rearrangements and repetitive sequence elements. Even though both 150 bp x2 [10,11,30,33,37,38] and 250 bp x2 [4,5,12,13,29,31,32,35–37,39] paired-end Illumina sequencing were reported successfully integrating accurate HLA typing, T. Profaizer et al., reported there was an approximately 5% decrease in unambiguous typing of samples when the PCR amplicons were fragmented into the 100–300 bp compared to 300–600 bp [43]. Illumina 250 bp x2 chemistry significantly increases sequencing time which can be challenging to reach fast turnaround time. NGS needs to be batched to be cost-effective. It can be challenging for smaller clinical laboratories to meet the turnaround time and batch the samples to reduce cost at the same time. For those laboratories, smaller instruments like iSeq100, S5 could be a good choice (Table 1).

3. Library preparation

The current NGS wet bench workflow generally includes the following steps: DNA extraction, library preparation, library enrichment, and sequencing. Samples from multiple patients can be pooled into one run by labeling the individual sample with a unique barcode. Several commercial HLA typing reagents are available on the market, all of which have achieved comparable accuracy and relatively easy multiplexing workflow covering classical HLA class I (*HLA-A, -B, -C*) and class II (*HLA-DRB1, -DRB345, -DQA1, -DQB1, -DPA1, -DPB1*) genes [8–13,44,45]. These reagents can be divided into two NGS typing approaches. The long-range PCR-based library construction protocols will have some variations due to different NGS platforms. However, the major library construction involves two main steps: 1) DNA fragmentation; 2) attachment of adaptor and barcoding. The long-range PCR approach requires high quality of DNA compared to probe-based assays. Poor quality DNA can cause PCR bias in heterozygous samples, which results in allele imbalance and allele drop out under extreme scenarios [4,46]. The second approach of NGS HLA typing uses HLA-specific probes to capture HLA genes through hybridization. The probe-based HLA typing was first reported by Wittig and colleagues [34]. They demonstrated no HLA allele dropout of 357 samples in eight HLA loci tested (*HLA-A*, *HLA-C, HLA-B, HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1,* and *HLA-DPB1*). The benefit of capture-based typing is there is no need for long-range PCR, therefore saving PCR time and avoiding PCR amplification bias. However, it has an intrinsic bias in capturing AT/GC-rich areas, and highly repetitive sequences may be underrepresented. Since HLA class II has large introns that tend to be full of repeat elements, it is challenging to design probes with enough specificity to capture sequences from these areas. Therefore, the probe-based assay only covers the full exons of HLA class II genes. The lack of coverage in introns can result in typing ambiguities in HLA class II genotyping, but mostly in the 4th field of typing [47].

Due to the short sequencing length of NGS (400–600 bp), genomic DNA needs to be fragmented into smaller fragments before sequencing and adaptors will be added to each sequencing fragment. Library construction includes DNA fragmentation, adaptor ligation, and barcoding. Nearly all NGS library preparation methods use some type of DNA size-selection and purification before sequencing. The size of the target DNA fragments in the final library is an important parameter for NGS library construction. This process involves removing unwanted fragment sizes that will interfere with downstream library preparation steps, sequencing, or analysis. There are several approaches to DNA size selection. Two commonly used methods are bead-based selection [48,49] and gel electrophoresis-based selection [50]. Ampure is a commonly used bead-based size selection method. Ampure beads bind reversibly to DNA in the presence of polyethylene glycol (PEG) and salt. Both salt and PEG are important to form DNA precipitate on Ampure beads dependent on their concentration and the beads to DNA ratio. Pippin Prep is a gel electrophoresis system that can separate DNA fragments ranging from 100 bp to 1.5 kb depending on agarose concentration, the target sizes can be entered in software, and DNA fractions are collected automatically. PippinPrep has 5 lanes and PippinHT has 24 lanes for higher throughput. Since the DNA library is already barcoded, multiple samples can be pooled into one lane as long as it does not exceed the capacity of the DNA load. Bead-based selection is easier to be incorporated into an automated library preparation system. If done manually,

Table 2
Publications on HLA typing using Illumina and Ion Torrent platforms.

Year	Author	Genes/HLA Locus	PCR/ Capture	Sequencer	Sequencing length	Coverage range	Sequencing time (estimated)	# of samples/run
2012	Wang C et al. [30]	*HLA-A, -B, -C, -DRB1*	Long PCR	GAIIx, HiSeq, MiSeq (Illumina)	100x2, 150x2	99.3% of alleles meet the minimum coverage of 100x, and the majority of them are beyond 900x	~ 24 h for MiSeq	180/lane 2700/ HiSeq run
2013	Hosomichi K et al. [5]	*HLA-A, -B, -C, -DRB1, -DQB, -DPB1*	Long PCR	MiSeq (Illumina)	250X2	HLA class I genes have higher average coverage 3,405 × compared to HLA class II genes 1,157×	~39 h	/
2014	Lange V et al. [31]	*HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1*	Exon based PCR	MiSeq (Illumina)	250X2	1000 − 3000x coverage (median) with only 9 sample exons (out of 20,388) below 250x	~39 h	384
2014	Ehrenberg PK et al. [32]	*HLA-A, -B, -C, -DRB1*	Long PCR	MiSeq (Illumina)	250x2	Exon average coverage: 26.8–3570	~39 h	96
2015	Zhou M et al. [33]	*HLA-A, -B, -C, -DRB1, -DQB1*	PCR	HiSeq (Illumina)	150x2	Average exonic regions is 500×, and 96% of the exonic regions depth > 100×	~27 h	2068/lane 16544/ run
2015	Wittig M et al. [34]	*HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, -DPA1, DPB1*	Capture	HiSeq (Illumina)	100x2	Unique start point coverage is > 50x	~10 days	48 samples/lane
2015	Ozaki Y et al. [7]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQB1, -DPB1*	Long PCR	Ion Torrent PGM (ThermoFisher)	Average read length is 273.3 ± 9.9 bp	Average depth per locus: 156.9 ± 48.5 to 418.3 ± 143.0	~4–7 h	46
2015	Barone JC et al. [8]	*HLA-A, -B, -C, -DRB1, -DQB1, -DPB1*	Long PCR	Ion Torrent PGM (ThermoFisher)	The mean read length is 274.5 ± 18 bp	Mean coverage for exons within each locus ranged from 89 to 745	~4–7 h	48
2016	Weimer ET et al. [29]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	MiSeq (Illumina)	250x2	The average is 280 for all HLA loci	39 ~ hours	24
2016	Yin Y et al. [4]	*HLA-A, -B, -C, -DRB1, -DQB1*	Long PCR	MiSeq (Illumina)	250x2	The average coverage depth is > 500 × across all loci	~39 h	96
2016	Duke JL et al. [12]	*HLA-A, -B, -C, -DRB1, -DQB1*	Long PCR	MiSeq (Illumina)	250x2	The average depth of coverage is 1875-3212x, while the minimum depth is 50x	39 ~ hours	up to 48
2016	Norman PJ et al. [35]	*HLA-A, -B, -C, KIR*	Capture	HiSeq, MiSeq (Illumina)	101x2, 250X2, 300x2	/	~56 h for MiSeq	/
2017	Ehrenberg PK et al. [36]	*HLA-A, -B, -C, -DRB1, -DQB1, -DPB1*	Long PCR	MiSeq (Illumina)	250X2	All exons: 12-2070x	~39	96
2017	Profaizer T et al. [37]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	MiSeq (Illumina)	150x2, 250X2	The average is >200x	~19 vs ~39 h	24
2017	Montgomery MC et al. [38]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	MiSeq (Illumina)	150x2	/	~19 h	17
2017	Schöfl G et al. [39]	*HLA-A, -B, -C, -DRB1, -DQB1, -DPB1, CCR5, RBO + RHD, KIR*	Exon based PCR	MiSeq, HiSeq (Illumina)	250x2	98.6% of all HLA amplicons achieved > 100x	~3 days for HiSeq (Rapid mode)	up to 10,000/ HiSeq run
2017	Norman PJ et al. [40]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1, GPX5, MUC22, C4A/ C4B, BTB9*	Capture	HiSeq (Illumina)	/	Mean read depth of 59.4 (SD 18.4) and with 99.6% of the coverage > 10×	/	up to 96
2018	Thorstenson YR et al. [10]	*HLA-A, -B, -C, - DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	MiSeq, NextSeq (Illumina)	150x2	/	~24 h for MiSeq, ~29 h for NextSeq	up to 384 samples on NextSeq; up to 24 samples on MiSeq
2018	Closa L et al. [13]	*HLA-A, -B, -C, KIR*	Long PCR	MiSeq (Illumina)	150x2, 250X2	Mean coverage for HLA class I is 345.08–378.01x	24/40 h	up to 96
2019	Creary LE et al. [11]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	MiSeq, NextSeq (Illumina)	150x2	/	~24 h for MiSeq, ~29 h for NextSeq	up to 384 samples on NextSeq; up to 24 samples on MiSeq
2020	Cargo M et al. [9]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	Ion Torrent S5 XL (ThermoFisher)	/	/	~4 h	24
2020	Liu C et al. [41]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	Ion Torrent S5 (ThermoFisher)	Median read length is 261 bp	The minimum key-exon coverage is > 100x for most samples across all loci	~4 h	16–48
2020	Truong L et al. [42]	*HLA-A, -B, -C, -DRB1, -DRB3/ 4/5, -DQA1, -DQB1, -DPA1, -DPB1*	Long PCR	Ion Torrent S5 XL (ThermoFisher)	Mean read-length is 260 bp ± 10%	0.24% alleles with less than 20x	~4 h	up to 48

Note: "/" indicates data not applicable.

bead-based selection takes a little longer in terms of hands-on time compared to gel electrophoresis-based selection methods, while gel electrophoresis-based selection takes a little longer to run.

Because NGS includes parallel sequencing of many patient samples with multiple loci in a single run, it is crucial to accurately determine the amount of DNA input of each patient and loci equally, therefore nucleic acid quantification is very important to ensure each allele tested has a balanced representation of data. Excess DNA in the libraries may cause high cluster density with signals from one cluster bleeding into adjacent clusters; or more than one library template attached to a bead or flow cell, leading to mixed sequences. The PhiX control which consists of balanced base composition at ~ 45% GC and ~ 55% AT, can serve as an ideal Illumina sequencing control for quality monitoring including cluster generation and sequencing. Commonly used DNA qualification and quantification equipment are listed in Table 3. The NanoDrop method is a UV absorbance-based method, which may overestimate the concentration. Qubit is a benchtop fluorometer that uses the binding of a fluorescent dye to DNA to detect DNA concentration. It allows for highly sensitive quantitation of DNA detecting as little as 10 pg/μl (Table 3). Agilent Bioanalyzer or TapeStation not only can be used for DNA concentration measurement but also can be used for detecting the size distribution of DNA fragments. The Bioanalyzer is less expensive but requires using a chip with a capacity of up to 12 samples. The TapeStation uses strip tubes, which allows for greater flexibility in the number of samples analyzed at a time. If only a few samples will be run at a time, the TapeStation would be more practical and cost-efficient especially when the cost of consumables is considered. However, the NanoDrop, Qubit, Agilent Bioanalyzer, and TapeStation are considered low throughput methods. qPCR quantification method [51] is highly sensitive (1 pg/ul), which can run up to 96 samples at a time. There are many additional options, such as a plate reader, which can accommodate 96 well plates. It has both fluorescence and UV absorbance detection methods. The clinical laboratory can choose appropriate instruments based on the lab space and budget.

This complex library construction is a major limitation of NGS. It is labor-intensive and has a longer turnaround time in comparison to SSP and SSO [52], which can result in a few hours. Implementation of an automated library preparation system in the clinical laboratory significantly reduces hands-on time and human errors due to the complexity of the library preparation. Automated library preparation systems including small benchtop instruments such as KingFisher Duo Prime, Flex, and Presto (ThermoFisher) and high throughput systems including Biomek FX/iSeries (Beckman Coulter), NGS STAR (Hamilton), Scicione NGSx (PerkinElmer), etc., to meet a range of throughput needs. Significant cost reduction on NGS library construction is expected with the introduction of automated platforms [4]. Laboratories can reduce cost by streamlining workflow, choosing the most cost-effective library preparation methods, and increasing the number of samples per run. For example, the automated library construction can be set up during the night, thus it not only saves hands-on time of the technologists but also saves the turnaround time of the test. An automated workflow will dramatically reduce hands-on time, eg. to construct 24 samples for NGS sequencing, it takes 1–2 h of hands-on time using the manual method in our laboratory while the automated method only takes 10–20 min in our hands. Currently, with automation and multiple shifts, NGS HLA typing can be reported within 2–3 days. This turnaround time will have to be assessed to determine if it meets the needs of each lab's clinical workflow.

After library preparation, both Illumina and Ion Torrent require clonal amplification to generate enough signal for detection. Clonal amplification and sequence both happen on the flow cell of the Illumina platform, therefore the machine running time includes both procedures. While Ion Torrent clonal amplification is not done on the sequencer, the machine running time only includes actual sequencing time. Therefore, the running time of iSeq100 and Ion Torrent S5/S5XL is similar.

4. Understanding NGS sequencing data

High quality and unbiased sequencing data are the keys to accurate HLA typing results. After sequencing, FASTQ files that contain raw NGS sequencing reads, and sequence quality scores will be generated. In FASTQ files, each nucleotide base is associated with a quality score (Q score). Phred quality score (Q) indicates the probability that a given base is called incorrectly by the sequencer [53,54]. Q30 is indicating that the probability of an incorrect base call was less than 1 in 1000. This nominally corresponds to a 99.9% correct rate of the base calling. In most commercial software, the FASTQ files undergo a series of bioinformatics processes to assess read length distribution, quality scores, GC contents, and demultiplexing. The raw reads of the FASTQ file are then reassembled into a full sequence by aligning to a known reference sequence or by *de novo* assembly to generate BAM files, which is sequencing alignment data stored in a binary format. Most of the commercial software of the HLA typing kits can provide relatively accurate HLA genotype results [45,55].

One of the greatest advantages of NGS HLA typing is it significantly reduced cis/trans ambiguities inherent in Sanger sequencing, resulting in one pass testing. However, since IPD-IMGT/HLA database [56] has incomplete sequence coverage beyond exons 2 and 3 of many of the HLA alleles, and with the ever-growing number of HLA alleles, HLA typing ambiguity will likely exist, but to a much smaller extent. NGS library has an average insert size of less than 1000 bp [4,57], meaning that polymorphisms>1000 bp apart cannot be phased, which may still result in phasing ambiguities. For example, HLA *DRB1*04:01:01:01* and *DRB1*04:01:01:02* only differ in intron 1 (8.2 kb), and the length of GT repeats in intron 2. There is a ~ 6.5 kb homolog sequence between *DRB1*04:01:01:01* and *DRB1*04:01:01:02* without polymorphic sites, therefore, when intron 1 is not included in the sequencing

Table 3
List of DNA Quantification machines.

	NanoDrop	Qubit	Bioanalyzer	TapeStation	qPCR
Quantification method	UV absorbance	Fluorescence	Fluorescence	Fluorescence	Fluorescence
Specific for DNA or RNA	Non-Discriminatory	Quantifies DNA and RNA independently	Quantifies DNA and RNA independently	Quantifies DNA and RNA independently	Quantifies DNA and RNA independently
Accuracy at low concentrations	Significantly overestimates concentration	Accurately quantify 10 pg/μl	Accurately quantify 5 pg/μl	Accurately quantify 10 pg/μl	Accurately quantify 1 pg/μl
Throughput	Low	Low	Medium 12 samples/run	Medium 16 samples/run	High 96 samples/run
Cost	Low	Low	High	High	High

strategy, or when the DNA fragmentation is too short, there will be typing ambiguities between the two alleles. However, like this example, most of these types of ambiguities can be reported as G group clinically, or if it is a rare allele, it can be ruled out by allele frequency.

Coverage is an important concept in NSG. Coverage depth refers to the number of times that a base has been sequenced, while coverage breadth refers to the percentage of a target gene sequenced [58]. For accurate HLA typing, a balanced depth of coverage and read breadth is equally important. Theoretically, each heterozygous position should be present in 50% of reads. Uneven depth of coverage across HLA genes can be caused by poor DNA quality, PCR dropout or allelic imbalance, libraries with too small of fragments to achieve phasing, and inability to sequence high GC regions, regions with short tandem repeats, such as exon 1 of the *DPB1* with $(AAGG)_{(4-17)}$ [59] and *DRB1* genes with $(GT)_{(7-27)}(GA)_{(5-30)}$ [57] or homopolymer runs. The analysis software may inappropriately perceive the low-coverage heterozygous allele as background "noise", leading to the assignment of homozygosity with the preferentially amplified allele [60,61]. For a laboratory to have the sensitivity of detecting 5% of the imbalanced allele, a minimum depth of 400 coverages will result in 20x coverages at the imbalanced position. Albrecht et al. recommend at least 10 coverages are required to counterbalance random noise [62]. Song et al. demonstrated accurate SNP detection could reach 97% when read depth was greater than 10 and recommended setting read depth at 20 to ensure accuracy [63]. Currently, literature reported NGS coverages ranging from 10 to a couple of thousands with an average coverage in hundreds (Table 2). Allele imbalance usually is not a significant problem for heterogynous alleles that differ with multiple polymorphic sites. However, it can cause allele dropout for alleles with similar SNP for example, *DRB1*15:01:01* and *DRB1*15:03:01* only differ at exon 2 where *DRB1*15:01:01* carries a T while *DRB1*15:03:01* carries a C at amino acid (AA) Codon position 30 (TAC vs. CAC), particularly both alleles are commonly associated with *DQB1*06:02*. Currently, there is no consensus on the minimum coverage requirement in a clinical setting, therefore each laboratory has to set its parameters to achieve accurate results. However, increasing the average coverage will reduce the number of samples that can be performed on each sequencing run, therefore increasing the cost. Generally, the number of samples that can be run depends on expected coverages, eg. the number of samples per run = mapped sequences / [size of all 11 HLA loci $\times$ 2 (allele) $\times$ average coverage expected for each allele].

Allele drop out is a major concern for long-range PCR NGS protocol, particularly when the DNA quality and quantity are not sufficient. Allele drop out can be caused by: 1) PCR primer design problems; 2) a novel allele containing a polymorphic site at the primer binding site; and 3) extreme allele amplification bias due to different GC content between two heterozygous alleles. However, allele imbalance is not the only criteria for correct allele assignment. For example, *DQB1*02:01* and *DQB1*02:02* have 300 bp insertion with high GC content which is absent in other DQB1 alleles. This may result in inefficient PCR amplification of *DQB1*02* when they co-amplified with other DQB1 alleles causing allele imbalance or drop out. In this case, using the knowledge of *DR17-DQ2* and *DR7-DQ2* association will help to identify potential allele drop out. Even the presence of a minor allele only represents 1–2% of a heterozygous sample and has low coverage of around 10, additional investigation using SSP, SSO, or SBT should be performed to confirm the homozygosity as the software may consider it as a background "noise". However, when a minor allele only represents 1–2% of the heterozygous sample but has high coverages > 100x, it should be considered a true signal, not background. The knowledge of HLA allele frequency, *HLA B-C, DRB1-DQA1-DQB1* linkage are also very important to determine if an allele drop out is present and if additional testing is needed. Normally, it is a good practice to set the maximum allele imbalance between 20% and 25% [64].

With the broad use of long-range PCR and hybrid-capture-based HLA typing methods in the clinical laboratory, a large number of novel intron polymorphisms or polymorphisms outside antigen recognition sites were detected. However, it probably does not affect routine clinical practice and it is up to the individual laboratory to decide if they would like to further investigate these novel alleles and submit them to the IMGT/HLA Database. The submission of novel alleles to the IMGT/HLA Database requires accurate consensus sequences reported by two independent PCR reactions to prevent unspecific PCR artifacts [62].

5. NGS data storage

The volume of data generated by an NGS analysis pipeline is truly massive in comparison with SSP, SSO, and SBT. Sophisticated bioinformatics, fast data processing capabilities, and data streaming can be costly. Laboratories need to prepare for a marked increase in data analysis and storage requirements and be prepared for financial investment in computers and local or cloud storage. NGS data such as FASTQ and BAM files, need to be retained and stored by the laboratory to provide complete traceability of patient test results. Generally, storage of these files for a minimum of 3 years is recommended to comply with Clinical Laboratory Improvement Amendment policies and should be HIPPA compliant. Data storage and processing may be performed locally on 1) external hard drives or a local secure server; 2) cloud services such as the BaseSpace Sequence Hub of Illumina. BaseSpace Sequence Hub is a security-first platform that has been certified for HIPAA compliance. It includes end-to-end encryption, auditing, and fine-grained access control; and 3) through a third party, such as Amazon Web Services (AWS). AWS is a secure cloud-based service and storage that is in line with HIPAA compliance requirements.

6. Validation and quality control

Validation contributes to the understanding of the strengths, weaknesses, and limitations of the testing performed in the laboratory. It is critical to validate the entire NGS process from end to end (DNA extraction through the analysis) [31]. Similar to all clinical laboratory tests, sensitivity, specificity, and reproducibility should be established for NGS HLA typing. Ensure the validation includes a sufficient number of samples and no less than 50 samples for all loci, and a minimum of 80% of the blind samples concordant at all tested loci at least in the 1st and 2nd fields [65]. Validation should include as many HLA alleles as possible and ideally include samples with homoplymer runs if using the Ion Torrent sequencing platform. Laboratories have to ensure the validation includes all sample types (e.g., blood, buccal swab) the lab routinely uses for clinical testing. Notably, long-range PCR based NGS HLA sequencing has an issue of allele dropout especially for buccal swab samples due to low quality and/or quantity of buccal swab DNA. PCR primer-based bias is subject to variations in DNA quantity or quality or caused by sequence-specific variations [4]. Laboratories also have to establish appropriate quality control, e.g., fragment size, average read length, sequencing quality, coverage death, coverage breadth, and allele imbalance threshold. When the QC metric does not meet the criteria, an additional investigation needs to be performed particularly when there are exonic mismatches, not enough coverages, uncommon *HLA B-C, DR-DQ,* and *DRB1-DRB345* linkages, rare alleles, novel alleles, or extreme allelic imbalance.

The cost of validation also needs to be considered in developing an NGS assay which includes reagents and labor. In addition to the

cost of validation, the personnel cost of implementation and training should also be considered. Personnel requirements to run NGS may vary across different states. Depending on the staff's molecular background, the training process may take weeks to months to reach proficiency in both the benchwork and analysis of results.

7. Conclusion and future development

The choice of sequencing platforms and reagents largely depends on sample volume, turnaround time, cost, and workflow in a clinical laboratory. Currently, commercial reagents fulfill accurate HLA typing despite some phasing ambiguities of some HLA class II genes [4,57]. For short-read NGS, sequences from homologous regions will have poor mapping quality, and reads belonging to the homolog regions can be mapped to wrong places causing inadequate coverages. For repetitive areas, a unique sequence flanking the repetitive region is required to reliably map a sequencing read. This problem can be resolved by a longer sequence read [66,67]. Two new sequencing platforms referred to as third-generation sequencers, can provide longer sequencing reads of a single molecule: Pacific Biosciences (PacBio) SMRT (single molecule real-time) (Menlo Park, California), and Oxford Nanopore (Oxford, United Kingdom). The PacBio sequencer uses multiple tiny wells each of which has a DNA polymerase attached to the bottom with one long DNA fragment [68]. A fluorescence signal from a tagged nucleotide corresponding to the addition of one specific nucleotide is detected when a base is added to the DNA strand [68]. The Oxford Nanopore uses a protein pore inserted into a membrane to measure the ionic current fluctuations when single-stranded nucleic acids pass through biological nanopores [69]. The individual nucleotides (A, C, G, or T) confer different resistances to the stretch of nucleic acid within the pore [69,70]. Importantly, neither of the two platforms requires an amplification step thus reducing the background noise. Both platforms can perform long reads (averaging 14–40 Kbp for the PacBio and 8–100 Kbp for the Nanopore, with the current record of Nanopore sequencing at ~2.3 Mb [71]), which can overcome issues of repetitive regions. However, both platforms have higher error rates compared to second-generation sequencing methods [68,72]. For PacBio, replicated sequencing of the same molecule can overcome the high error rate as the errors are random [73,74]. However, replicated sequencing cannot overcome the errors on Nanopore as its errors are biased [75]. The accuracy of reads produced by both platforms has dramatically increased recently. It is claimed the error rate has been reduced to <1% for PacBio [76] and <5% for Nanopore [77]. Yet these sequencers have limited adoption in the clinical laboratory, due to their higher price, lower throughput, and possibly owing to the challenges of clinically validating instruments with high intrinsic error rates.

Currently, the implementation of NGS technology in routine diagnostics laboratories is expected to continuously grow. The benefits of sequencing introns and untranslated regions of HLA genes will greatly increase our knowledge in transplant immunology and may lead to improved outcomes. With the development of long-read third-generation sequencing, better HLA haplotype phasing and understanding the roles of other immune regulatory genes in the HLA region will be possible and are essential for precision patient care.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463.
[2] S. Yohe, B. Thyagarajan, Review of Clinical Next-Generation Sequencing, Arch. Pathol. Lab. Med. 141 (2017) 1544.
[3] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, et al., Comparison of next-generation sequencing systems, J. Biomed. Biotechnol. 2012 (2012) 251364.
[4] Y. Yin, J.H. Lan, D. Nguyen, N. Valenzuela, P. Takemura, Y.T. Bolon, et al., Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: results from over 10,000 donor recruitment samples, PLoS ONE 11 (2016) e0165810.
[5] K. Hosomichi, T.A. Jinam, S. Mitsunaga, H. Nakaoka, I. Inoue, Phase-defined complete sequencing of the HLA genes by next-generation sequencing, BMC Genomics 14 (2013) 355.
[6] E.A. Trachtenberg, C.L. Holcomb, Next-generation HLA sequencing using the 454 GS FLX system, Methods Mol. Biol. 1034 (2013) 197.
[7] Y. Ozaki, S. Suzuki, K. Kashiwase, A. Shigenari, Y. Okudaira, S. Ito, et al., Cost-efficient multiplex PCR for routine genotyping of up to nine classical HLA loci in a single analytical run of multiple samples by next generation sequencing, BMC Genomics 16 (2015) 318.
[8] J.C. Barone, K. Saito, K. Beutner, M. Campo, W. Dong, C.P. Goswami, et al., HLA-genotyping of clinical specimens using Ion Torrent-based NGS, Hum. Immunol. 76 (2015) 903.
[9] M. Cargou, M. Ralazamahaleo, L. Blouin, I. Top, V. Elsermans, M. Andreani, et al., Evaluation of the AllType kit for HLA typing using the Ion Torrent S5 XL platform, HLA 95 (2020) 30.
[10] Y.R. Thorstenson, L.E. Creary, H. Huang, V. Rozot, T.T. Nguyen, F. Babrzadeh, et al., Allelic resolution NGS HLA typing of Class I and Class II loci and haplotypes in Cape Town, South Africa. Hum Immunol 79 (2018) 839.
[11] L.E. Creary, S. Gangavarapu, K.C. Mallempati, G. Montero-Martin, S.J. Caillier, A. Santaniello, et al., Next-generation sequencing reveals new information about HLA allele and haplotype diversity in a large European American population, Hum. Immunol. 80 (2019) 807.
[12] J.L. Duke, C. Lind, K. Mackiewicz, D. Ferriola, A. Papazoglou, A. Gasiewski, et al., Determining performance characteristics of an NGS-based HLA typing method for clinical applications, HLA 87 (2016) 141.
[13] L. Closa, F. Vidal, M.J. Herrero, J.L. Caro, Design and Validation of a Multiplex KIR and HLA Class I Genotyping Method Using Next Generation Sequencing, Front. Immunol. 9 (2018) 2991.
[14] C. Lind, D. Ferriola, K. Mackiewicz, S. Heron, M. Rogers, L. Slavich, et al., Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing, Hum. Immunol. 71 (2010) 1033.
[15] J.J. Kim, S.V. Fuggle, S.D. Marks, Does HLA matching matter in the modern era of renal transplantation?, Pediatr. Nephrol. (2019).
[16] S.J. Lee, J. Klein, M. Haagenson, L.A. Baxter-Lowe, D.L. Confer, M. Eapen, et al., High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation, Blood 110 (2007) 4576.
[17] Y. Morishima, K. Kashiwase, K. Matsuo, F. Azuma, S. Morishima, M. Onizuka, et al., Biological significance of HLA locus matching in unrelated donor bone marrow transplantation, Blood 125 (2015) 1189.
[18] C. Wiebe, V. Kosmoliaptsis, D. Pochinco, I.W. Gibson, J. Ho, P.E. Birk, et al., HLA-DR/DQ molecular mismatch: A prognostic biomarker for primary alloimmunity, Am. J. Transplant. 19 (2019) 1708.
[19] E.W. Petersdorf, M. Malkki, C. O'HUigin, M. Carrington, T. Gooley, M.D. Haagenson, et al., High HLA-DP Expression and Graft-versus-Host Disease, N. Engl. J. Med. 373 (2015) 599.
[20] V. Garcia Moreira, B. Prieto Garcia, J.M. Baltar Martin, F. Ortega Suarez, F.V. Alvarez, Cell-free DNA as a noninvasive acute rejection marker in renal transplantation, Clin. Chem. 55 (2009) 1958.
[21] I. De Vlaminck, H.A. Valantine, T.M. Snyder, C. Strehl, G. Cohen, H. Luikart, et al., Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection, Sci. Transl. Med. 6 (2014) 241ra77.
[22] J.M. Lee, Y.J. Kim, S.S. Park, E. Han, M. Kim, Y. Kim, Simultaneous monitoring of mutation and chimerism using next-generation sequencing in myelodysplastic syndrome, J. Clin. Med. (2019) 8.
[23] A. Klussmeier, C. Massalski, K. Putke, G. Schafer, J. Sauter, D. Schefzyk, et al., High-throughput MICA/B genotyping of over two million samples: workflow and allele frequencies, Front. Immunol. 11 (2020) 314.
[24] B. Maniangou, N. Legrand, M. Alizadeh, U. Guyet, C. Willem, G. David, et al., Killer immunoglobulin-like receptor allele determination using next-generation sequencing technology, Front. Immunol. 8 (2017) 547.
[25] B. Maniangou, C. Retiere, K. Gagne, Next-generation sequencing technology a new tool for killer cell immunoglobulin-like receptor allele typing in hematopoietic stem cell transplantation, Transfus. Clin. Biol. 25 (2018) 87.
[26] H. Morris, S. DeWolf, H. Robins, B. Sprangers, S.A. LoCascio, B.A. Shonts, et al., Tracking donor-reactive T cells: Evidence for clonal deletion in tolerant kidney transplant patients, Sci. Transl. Med. 7 (2015) 272ra10.
[27] S.D. Adams, K.C. Barracchini, D. Chen, F. Robbins, L. Wang, P. Larsen, et al., Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification, J. Transl. Med. 2 (2004) 30.
[28] C.E. Voorter, E. Mulkers, P. Liebelt, E. Sleyster, E.M. van den Berg-Loonen, Reanalysis of sequence-based HLA-A, -B and -Cw typings: how ambiguous is today's SBT typing tomorrow, Tissue Antigens 70 (2007) 383.

[29] E.T. Weimer, M. Montgomery, R. Petraroia, J. Crawford, J.L. Schmitz, Performance characteristics and validation of next-generation sequencing for human leucocyte antigen typing, J. Mol. Diagn. 18 (2016) 668.
[30] C. Wang, S. Krishnakumar, J. Wilhelmy, F. Babrzadeh, L. Stepanyan, L.F. Su, et al., High-throughput, high-fidelity HLA genotyping with deep sequencing, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 8676.
[31] V. Lange, I. Bohme, J. Hofmann, K. Lang, J. Sauter, B. Schone, et al., Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing, BMC Genomics 15 (2014) 63.
[32] P.K. Ehrenberg, A. Geretz, K.M. Baldwin, R. Apps, V.R. Polonis, M.L. Robb, et al., High-throughput multiplex HLA genotyping by next-generation sequencing using multi-locus individual tagging, BMC Genomics 15 (2014) 864.
[33] M. Zhou, D. Gao, X. Chai, J. Liu, Z. Lan, Q. Liu, et al., Application of high-throughput, high-resolution and cost-effective next generation sequencing-based large-scale HLA typing in donor registry, Tissue Antigens 85 (2015) 20.
[34] M. Wittig, J.A. Anmarkrud, J.C. Kassens, S. Koch, M. Forster, E. Ellinghaus, et al., Development of a high-resolution NGS-based HLA-typing and analysis pipeline, Nucleic Acids Res. 43 (2015) e70.
[35] P.J. Norman, J.A. Hollenbach, N. Nemat-Gorgani, W.M. Marin, S.J. Norberg, E. Ashouri, et al., Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing, Am. J. Hum. Genet. 99 (2016) 375.
[36] P.K. Ehrenberg, A. Geretz, R.K. Sindhu, T. Vayntrub, M.A. Fernandez Vina, R. Apps, et al., High-throughput next-generation sequencing to genotype six classical HLA loci from 96 donors in a single MiSeq run, HLA 90 (2017) 284.
[37] T. Profaizer, E. Lazar-Molnar, A. Pole, J.C. Delgado, A. Kumanovics, HLA genotyping using the Illumina HLA TruSight next-generation sequencing kits: A comparison, Int. J. Immunogenet. 44 (2017) 164.
[38] M.C. Montgomery, R. Petraroia, E.T. Weimer, Buccal swab genomic DNA fragmentation predicts likelihood of successful HLA genotyping by next-generation sequencing, Hum. Immunol. 78 (2017) 634.
[39] G. Schofl, K. Lang, P. Quenzel, I. Bohme, J. Sauter, J.A. Hofmann, et al., 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned, BMC Genomics 18 (2017) 161.
[40] P.J. Norman, S.J. Norberg, L.A. Guethlein, N. Nemat-Gorgani, T. Royce, E.E. Wroblewski, et al., Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II, Genome Res. 27 (2017) 813.
[41] C. Liu, B.F. Duffy, E.T. Weimer, M.C. Montgomery, J.E. Jennemann, R. Hill, et al., Performance of a multiplexed amplicon-based next-generation sequencing assay for HLA typing, PLoS ONE 15 (2020) e0232050.
[42] L. Truong, B. Matern, L. D'Orsogna, P. Martinez, M.G.J. Tilanus, D. De Santis, A novel multiplexed 11 locus HLA full gene amplification assay using next generation sequencing, HLA 95 (2020) 104.
[43] T. Profaizer, E.M. Coonrod, J.C. Delgado, A. Kumanovics, Report on the effects of fragment size, indexing, and read length on HLA sequencing on the Illumina MiSeq, Hum. Immunol. 76 (2015) 897.
[44] M. Kotowski, A. Bogacz, J. Bartkowiak-Wieczorek, A. Bukowska, N. Surowiec, K. Dziewanowski, et al., The importance of new generation sequencing (NGS) HLA typing in renal transplantation-preliminary report, Transplant Proc 50 (2018) 1605.
[45] K. Hosomichi, T. Shiina, A. Tajima, I. Inoue, The impact of next-generation sequencing technologies on HLA research, J. Hum. Genet. 60 (2015) 665.
[46] M.J. Gandhi, D. Ferriola, Y. Huang, J.L. Duke, D. Monos, Targeted next-generation sequencing for human leukocyte antigen typing in a clinical laboratory: metrics of relevance and considerations for its successful implementation, Arch. Pathol. Lab. Med. 141 (2017) 806.
[47] M.A. Jones, S. Bhide, E. Chin, B.G. Ng, D. Rhodenizer, V.W. Zhang, et al., Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation, Genet. Med. 13 (2011) 921.
[48] S. Bruinsma, J. Burgess, D. Schlingman, A. Czyz, N. Morrell, C. Ballenger, et al., Bead-linked transposomes enable a normalization-free workflow for NGS library preparation, BMC Genomics 19 (2018) 722.
[49] M.M. DeAngelis, D.G. Wang, T.L. Hawkins, Solid-phase reversible immobilization for the isolation of PCR products, Nucl. Acids Res. 23 (1995) 4742.
[50] M.A. Quail, Y. Gu, H. Swerdlow, M. Mayho, Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation, Electrophoresis 33 (2012) 3521.
[51] H. Andreasson, U. Gyllensten, M. Allen, Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis, Biotechniques 33 (2002) 402.
[52] H. Dunckley, HLA typing by SSO and SSP methods, Methods Mol. Biol. 882 (2012) 9.
[53] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res 8 (1998) 175.
[54] B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res 8 (1998) 186.
[55] K. Osoegawa, T.A. Vayntrub, S. Wenda, D. De Santis, K. Barsakis, M. Ivanova, et al., Quality control project of NGS HLA genotyping for the 17th International HLA and Immunogenetics Workshop, Hum. Immunol. 80 (2019) 228.
[56] J. Robinson, D.J. Barker, X. Georgiou, M.A. Cooper, P. Flicek, S.G.E. Marsh, IPD-IMGT/HLA database, Nucl. Acids Res. 48 (2020) D948.
[57] J.H. Lan, Y. Yin, E.F. Reed, K. Moua, K. Thomas, Q. Zhang, Impact of three Illumina library construction methods on GC bias and HLA genotype calling, Hum. Immunol. 76 (2015) 166.
[58] D. Sims, I. Sudbery, N.E. Ilott, A. Heger, C.P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses, Nat. Rev. Genet. 15 (2014) 121.
[59] H.W.M. van Deutekom, W. Mulder, E.H. Rozemuller, Accuracy of NGS HLA typing data influenced by STR, Hum. Immunol. 80 (2019) 461.
[60] S. Gao, D. Wang, Y. Xu, Z. Deng, Analysis for 2 samples with HLA-DQB1 allele dropout at exon 2 in sequence-based typing, Zhonghua Yi Xue Yi Chuan Xue Za Zhi 31 (2014) 496.
[61] A. Balas, F. Garcia-Sanchez, J.L. Vicario, HLA-A allele dropout in sequence-specific oligonucleotide probe typing due to intronic polymorphism in the novel A*31:01:02:02 allele, Tissue Antigens 86 (2015) 56.
[62] V. Albrecht, C. Zweiniger, V. Surendranath, K. Lang, G. Schofl, A. Dahl, et al., Dual redundant sequencing strategy: Full-length gene characterisation of 1056 novel and confirmatory HLA alleles, HLA 90 (2017) 79.
[63] K. Song, L. Li, G. Zhang, Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology, Sci. Rep. 6 (2016) 35736.
[64] W.C. Nelson, C.W. Pyo, D. Vogan, R. Wang, Y.S. Pyon, C. Hennessey, et al., An integrated genotyping approach for HLA and other complex genetic systems, Hum. Immunol. 76 (2015) 928.
[65] D. Monos, M.J. Maiers, Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): Consequences, opportunities and challenges, Hum. Immunol. 76 (2015) 883.
[66] S. Suzuki, S. Ranade, K. Osaki, S. Ito, A. Shigenari, Y. Ohnuki, et al., Reference grade characterization of polymorphisms in full-length HLA Class I and II Genes with short-read sequencing on the ION PGM system and long-reads generated by single molecule, real-time sequencing on the PacBio platform, Front. Immunol. 9 (2018) 2294.
[67] D. De Santis, L. Truong, P. Martinez, L. D'Orsogna, Rapid high-resolution HLA genotyping by MinION Oxford nanopore sequencing for deceased donor organ allocation, HLA 96 (2020) 141.
[68] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Genomics Proteomics Bioinformatics 13 (2015) 278.
[69] M. Jain, H.E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community, Genome Biol. 17 (2016) 239.
[70] J. Schreiber, Z.L. Wescoe, R. Abu-Shumays, J.T. Vivian, B. Baatar, K. Karplus, et al., Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 18910.
[71] A. Payne, N. Holmes, V. Rakyan, M. Loose, BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files, Bioinformatics 35 (2019) 2193.
[72] M. Miyamoto, D. Motooka, K. Gotoh, T. Imai, K. Yoshitake, N. Goto, et al., Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes, BMC Genomics 15 (2014) 699.
[73] S. Koren, M.C. Schatz, B.P. Walenz, J. Martin, J.T. Howard, G. Ganapathy, et al., Hybrid error correction and de novo assembly of single-molecule sequencing reads, Nat. Biotechnol. 30 (2012) 693.
[74] X. Jiao, X. Zheng, L. Ma, G. Kutty, E. Gogineni, Q. Sun, et al., A benchmark study on error assessment and quality control of CCS reads derived from the PacBio RS, J Data Mining Genomics Proteomics (2013;4.).
[75] C. Li, K.R. Chng, E.J. Boey, A.H. Ng, A. Wilm, N. Nagarajan, INC-Seq: accurate single molecule reads using nanopore sequencing, GigaScience 5 (2016) 34.
[76] A.M. Wenger, P. Peluso, W.J. Rowell, P.C. Chang, R.J. Hall, G.T. Concepcion, et al., Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome, Nat. Biotechnol. 37 (2019) 1155.
[77] M. Jain, S. Koren, K.H. Miga, J. Quick, A.C. Rand, T.A. Sasani, et al., Nanopore sequencing and assembly of a human genome with ultra-long reads, Nat. Biotechnol. 36 (2018) 338.