Human Immunology 82 (2021) 801–811

Contents lists available at ScienceDirect

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

Research article

Next-generation sequencing technologies: An overview

Taishan Hu [a], Nilesh Chitnis [a,c], Dimitri Monos [a,b,*], Anh Dinh [a,b,*]

[a] Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States
[b] Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[c] Department of Surgery, Baylor College of Medicine, Houston, TX, United States

ARTICLE INFO

Article history:
Received 14 December 2020
Revised 18 February 2021
Accepted 23 February 2021
Available online 19 March 2021

Keywords:
Next-generation sequencing
Short-read sequencing
Long-read sequencing

ABSTRACT

Since the days of Sanger sequencing, next-generation sequencing technologies have significantly evolved to provide increased data output, efficiencies, and applications. These next generations of technologies can be categorized based on read length. This review provides an overview of these technologies as two paradigms: short-read, or "second-generation," technologies, and long-read, or "third-generation," technologies. Herein, short-read sequencing approaches are represented by the most prevalent technologies, Illumina and Ion Torrent, and long-read sequencing approaches are represented by Pacific Biosciences and Oxford Nanopore technologies. All technologies are reviewed along with reported advantages and disadvantages. Until recently, short-read sequencing was thought to provide high accuracy limited by read-length, while long-read technologies afforded much longer read-lengths at the expense of accuracy. Emerging developments for third-generation technologies hold promise for the next wave of sequencing evolution, with the co-existence of longer read lengths and high accuracy.

1. Introduction

Nearly 25 years after the structure of DNA was discovered, the first method for sequencing DNA was published [1,2]. This method involved the addition of chain-terminating and radioactively labeled (earlier approach) or fluorescently labeled (later approach) dideoxynucleotides to perform sequencing of a DNA strand complementary to the interrogated template strand. The fragments were then size separated and analyzed by gel electrophoresis to determine the sequence. Known as Sanger-sequencing, the method continued to improve with the introduction of capillary electrophoresis and gained wide acceptance as a "first-generation sequencing" method to sequence small and large genomes from bacteria and phages, to humans. Given that only one sequencing reaction could be analyzed at a time, the method was of limited throughput. Sequencing of diploid DNA also complicated the discernment of haploid sequences critical to many diagnostic and investigative purposes, necessitating subcloning, plating, and DNA preparation of individual subclones before sequencing. These labor-intensive processes contributed to the first human genome project, taking over a decade and costing $2.7 billion to complete [3]. Despite additional advancements permitting additional human genome sequencing for $10 million, the technology had reached its ceiling for time and cost [4]. It was evident that in order to expand our sequencing capabilities, new technologies had to be developed.

Between 2004 and 2006 "next-generation sequencing (NGS)" technologies were introduced, which transformed biomedical inquiry and resulted in a dramatic increase in sequencing data-output [5]. The significant increase in data output was due to the nanotechnology principles and innovations that allowed massively parallel sequencing of single DNA molecules. The combined features of high throughput and single-molecule DNA sequencing are hallmarks of NGS, irrespective of the sequencing platform. The technology's evolved procedures were better merged with data acquisition and analysis, freeing the community from more labor-intensive and low-efficiency historical Sanger sequencing approaches and facilitating an extraordinary increase in data output. Second-generation approaches, such as on the Illumina or Ion Torrent platforms, generally start with DNA fragmentation, DNA end-repair, adapter ligation, surface attachment, and in-situ amplification. These "short-read" sequencing technologies involve the massively parallel sequencing of short reads, whereby millions of individual sequencing reactions occur in parallel. However, by nature of being short-read technologies, sequencing data over long stretches of DNA must be reassembled, presenting challenges with structural variations or low-complexity regions.

* Corresponding authors.
E-mail addresses: monosd@chop.edu (D. Monos), dinha1@chop.edu (A. Dinh).

https://doi.org/10.1016/j.humimm.2021.02.012
0198-8859/

Third-generation sequencing, such as on the Pacific Biosciences or Oxford Nanopore technology platforms, can achieve read lengths upwards of 10 kb, well beyond Sanger or short-read sequencing technologies. These "long-read" technologies can overcome issues encountered with short-reads, such as genome-wide repeats and structural variant detection. Compared to second-generation methods, process changes include minimal library preparation steps and direct targeting of unfragmented DNA molecules in real-time, where the limiting factor is the production of high molecular weight DNA for these purposes. An early limitation of these third-generation technologies compared to second-generation methods was the accuracy of the reads, which continues to improve over time, particularly with software analysis advancements.

Considering the currently available technologies and related advantages or disadvantages, the use of short-read or long-read sequencing depends on the research or clinical application. Thus far, research and clinical needs have dictated additional technological advancements. However, the limitations of these technologies have hindered additional scientific discoveries or clinical applications. The future direction of these technologies is an evolution from limiting factors to potentiating ones, giving rise to new research and clinical applications.

This review provides an overview of available NGS technologies categorized as short versus long reads, their benefits and limitations, remaining questions, and their future. While detailed clinical applications, data analysis software, and algorithms are beyond the scope of this review and are covered elsewhere in this Special Issue, human leukocyte antigen (HLA) genotyping by NGS is referenced to illustrate each technology.

2. Short-read sequencing

Short-read NGS, or second-generation sequencing, refers to the next advancement of sequencing technologies after traditional first-generation Sanger sequencing. The common feature of short-read technologies is massive sequencing of short (250–800 bp), clonally amplified DNA molecules sequenced in parallel [6]. NGS workflow includes library preparation, sequencing, and data analysis [7].

2.1. Library preparation

A pre-requisite for NGS is a good quality library. Library preparation includes first obtaining templates corresponding to molecules of interest for sequencing and subsequently prepare the fragments to make them compatible with the sequencing platform used [8]. NGS can be broadly divided into DNA sequencing (DNA-Seq) and RNA sequencing (RNA-Seq) [9-11].

2.1.1. DNA-seq library preparation

Depending on the sequencing template, DNA-Seq can include Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Epigenome Sequencing, or Targeted Sequencing (TS) [9,12].

Two approaches for template preparation include polymerase chain reaction (PCR) and hybridization capture-based approaches. PCR-based methods are commonly used to prepare template for TS [8,13]. In its early stages, amplicon sequencing focused on a limited number of genes or exons (short-range PCR). Such was the case for early HLA genotyping, which focused on the amplification of exons encoding the antigen-recognition site (ARS), but this approach presented the same issues of typing ambiguities seen with Sanger-based typing [14]. The advent of long-range PCR (LR-PCR) lead to shotgun sequencing (sequencing of randomly broken, shorter than the amplicon, fragments) as a dominant approach, and entire genes, including intronic, untranslated, upstream, and downstream regions, could be sequenced. Thus, LR-PCR improved issues of sequence ambiguities seen with short amplicon sequencing [15,16]. It should be noted, however, that the LR-PCR-based approach, especially for HLA genotyping, is occasionally characterized by allele dropouts.

Hybridization capture-based preparation of templates is utilized for WES and TS [17,18], whereby complementary biotinylated probes are hybridized with regions of interest, which are then isolated using streptavidin-coated magnetic beads. Compared to PCR-based approaches, this target enrichment approach is a cost-effective method for TS of larger genomic regions and a greater number of genes with decreased instances of allele dropout. That said, these two methods are not mutually exclusive. A combined approach, Regional-specific extraction (RSE), has been reported to capture the large genomic region of the major histocompatibility complex (MHC) [19] and other genomic regions [20,21]. The RSE methodology involves enzymatic extension of a non-biotinylated oligonucleotide hybridized to a particular sequence and incorporating biotinylated nucleotides as it extends. The extended DNA is then captured with streptavidin-coated magnetic beads, enriched with whole-genome amplification techniques, and prepared for sequencing.

The most common tools for studying epigenetic modifications and their impact on gene regulation are Whole-genome bisulfite sequencing (WGBS) and chromatin immunoprecipitation, followed by NGS (ChIP-Seq) [8,22,23]. WGBS enables a genome-wide analysis of DNA methylation (5mC) at base-pair resolution. Preparation of genomic samples for WGBS is commonly performed through the post-bisulfite treatment of DNA and de-tagging before index adaptor ligation for NGS sequencing [8]. ChIP-Seq allows for genome-wide mapping of DNA-binding proteins and histone modifications at base-pair resolution. To prepare samples for ChIP-Seq, formaldehyde-fixed or natural chromatin is fragmented by micrococcal nuclease (MNase) or sonication, which is further immunoprecipitated with target-specific antibody conjugated to magnetic beads. Isolated DNA from the precipitated protein-DNA complexes is used to generate libraries [8].

DNA-Seq library construction involves the core steps of fragmentation, end-repair, adaptor ligation, and size selection [24]. Fragmentation shears DNA into the optimal platform-specific size range. Three approaches for fragmentation include physical (i.e., acoustic shearing or sonication), enzymatic (i.e., fragmentase or transposase tagmentation), or chemical (heat with divalent metal cation) methods [8]. The integration of fragmentation into template preparation for WGBS and ChIP-Seq obviates additional fragmentation steps. Fragmentation, which permits representation of both small and large fragments, is a key consideration with respect to HLA genotyping on the Illumina MiSeq, a short-read sequencer. Smaller fragments contribute to high-quality sequencing data, whereas longer fragments provide distal phase information [25]. Subsequent end-repair prepares libraries for adaptor ligation by ensuring DNA fragment ends are free of overhangs that contain 5′ phosphate and 3′ hydroxyl groups, by blunt ending and dA tailing (addition of non-template deoxyadenosine 5′-monophosphate to blunted ends for ligation to adaptors with complementary dT-overhangs) for Illumina, or blunt ending only for Ion Torrent.

Adaptors include platform-specific sequences for fragment recognition by the sequencing instrument, such as DNA fragment binding to the flow cells of Illumina platforms. Adaptors also increasingly include a short, unique sequence, known as a barcode or index, to identify individual samples, allowing large numbers of samples to be pooled and sequenced simultaneously in a single run (multiplex sequencing). These barcodes can be identified and used to assign reads to individual samples during the data analysis. In the context of HLA genotyping, multiple HLA loci from a single

sample are prepared together with a single barcode and permits multiple samples to be simultaneously sequenced and differentiated during the analysis [25].

Following adaptor ligation, size selection enriches DNA fragments within a defined size range and removes contaminants to improve sequencing efficiency. Size selection can be performed using bead-based (Beckman Coulter's SPRIselect and Agencourt AMPure XP beads) or electrophoretic-based (Sage Sciences' Pippin Prep) approaches. Bead-based methods have the advantage of simultaneously concentrating pools, while electrophoretic-based methods improve precision. Proper size selection optimizes the sequencing run, increases the number of samples sequenced, provides higher-quality data, and maximizes phasing, the last of which is especially important for HLA typing [25]. The size-selected library can be sequenced directly or PCR-amplified before sequencing.

Illumina's Nextera technology offers an alternative method to prepare DNA fragments, utilizing an on-bead tagmentation library prep, which integrates the library preparation steps of DNA normalization, fragmentation, and size selection. Following tagmentation, a limited PCR is performed to integrate the adapters for sequencing and barcodes for sample indexing. This workflow is fast and simple, enabling sequencing-ready libraries to be generated in less than 90 min, with less than 15 min of hands-on time.

2.1.2. RNA-seq library preparation

RNA-Seq is useful for functional genomic studies such as differential gene expression, alternative splicing, and variant discovery [8,26-28]. RNA-Seq can be divided into Whole Transcriptome Sequencing (WTS), mRNA sequencing (mRNA-Seq), and small RNA sequencing (smRNA-Seq). Sample preparation generally includes three steps: total RNA isolation, target RNA enrichment, and reverse transcription of RNA into complementary DNA (cDNA).

Given that rRNA comprises at least 90% of total RNA extracted from mammalian cells or tissues, rRNA depletion must be included in the sample preparation for WTS [8]. Illumina Stranded Total RNA Prep with Ribo-Zero, a popular kit for preparing libraries for WTS, involves removing rRNA from total RNA, chemically fragmenting remaining RNA, and random priming for reverse transcription [29]. Subsequent end-repair, adaptor-ligation, and final PCR amplification result in a cDNA library. For mRNA-Seq sample preparation, mRNA is captured by oligo-dT magnetic beads and separated from total RNA. Library preparation for mRNA is similar to that of WTS, using the Illumina TruSeq Stranded mRNA kit [30].

Small RNAs are a class of non-coding RNAs that are shorter than 200 nucleotides. The most common and well-studied species are microRNAs (miRNA), which play critical roles in gene regulation. The library preparation of the smRNA-Seq is simple due to a 5′ terminal phosphate present in the native state of miRNA [8,31]. The library prep starts with a two-step ligation. A 3′-end blocked adenylated DNA adapter is first ligated to the RNAs using a special T4 RNA ligase 2, after which a second 5′ RNA adapter is ligated with RNA ligase 1. After ligation, reverse transcription-PCR is performed to convert ligated miRNAs into cDNAs, and the amplification product, following gel size-selection, is processed for sequencing.

2.2. Sequencing platforms

Short-read sequencing processes include two consecutive elements: clonal amplification and sequencing [5,32-34]. Clonal amplification involves solid-phase amplification of DNA fragments to produce strong, detectable signals during sequencing. The solid-phase to which single DNA fragments bind can be beads (Thermofisher's Ion Torrent) or flow cell surfaces (Illumina). Depending on the sequencing platform, emulsion PCR (Ion Torrent) or "bridging" PCR (Illumina) is used to amplify the anchored DNA fragments into millions of spatially separated template fragments.

The sequencing principle for both Illumina and Ion Torrent platforms is based on the "sequencing by synthesis (SBS)" approach, which involves DNA-polymerase-dependent nucleotide incorporation on the extended DNA chain [34]. These two platforms are reviewed below. Key characteristics of each of these two types of sequencing platforms are presented in Table 1.

2.2.1. Ion Torrent

Clonal amplification on the Ion Torrent platform is achieved by a bead-based method on Ion Sphere particles in a micro-well via emulsion PCR. In this process, adapter sequences are ligated to DNA fragments, and are subsequently captured in a water-in-oil emulsion droplet (micelle), along with a bead covered with complementary adapters, deoxynucleotides (dNTPs), primers, and DNA polymerase. Each micelle functions as a micro-PCR reactor, allowing PCR amplification independent of one another. The Ion torrent semiconductor chip consists of a flow chamber and a complementary metal-oxide semiconductors (CMOS) pH sensor. Micelles are loaded onto a semiconductor chip with microwells, and the chip is flooded with unmodified A, T, G, or C nucleotides sequentially during sequencing. Incorporating a single nucleotide results in the release of a hydrogen ion detected by the CMOS pH sensor. Ion Torrent is the first to perform semiconductor sequencing without optical sensing [33,34]. This technology delivers fast sequencing run times (between 2.5 and 4 hrs) with reads between 200 and 600 bp [35]. An intrinsic weakness of the Ion Torrent chemistry is difficulty sequencing through homopolymer regions. With the incorporation of multiple identical bases, there may be a loss of linearity of response from inaccurate measurements of voltage pulse magnitudes, which may appear as insertion/deletion errors in a single read. On the other hand, the technology has a low substitution error rate (per base rate of <0.1%) [36]. Ion Torrent platforms include the Ion PGM Dx, Ion GeneStudio S5, and Genexus instruments [37].

2.2.2. Illumina

llumina's NGS technology is based on SBS with a fluorescent-labeled reversible terminator technology [5,34]. Prior to sequencing, clonal amplification (cluster generation) of DNA libraries occurs through "bridge amplification" PCR, which occurs on the sequencing flow cell and is controlled by the sequencing instrument. Sequencing is based on the optical readout of incorporating fluorescent nucleotides coupled to a reversible terminator by a DNA polymerase. A single fluorescently labeled reversible terminator-bound dNTP is incorporated into the nucleic acid chain during each sequencing cycle, and the resulting fluorescent signal is imaged. The terminator and fluorescent dye are cleaved from the incorporated dNTP to allow the next labeled dNTP to be added. An animation of this technology can be found on the Illumina website [38]. In addition, Illumina NGS platforms are capable of paired-end sequencing, sequencing that occurs from both ends of a DNA fragment, which generates high-quality sequence data with in-depth coverage and high numbers of reads. Paired-end sequencing is useful for HLA typing because the sequencing of both ends of a long fragment permits phasing across distal polymorphic positions. The Illumina MiSeq instrument, commonly used for clinical HLA typing, favors fragment sizes of 350–500 bases long, but fragments 600–900 bases or longer are optimal for phasing distal polymorphisms, which is permitted by paired-end sequencing [25,39].

Illumina's reversible terminator technology, along with paired-end sequencing, makes it the most accurate base-by-base sequencing technology on the market, with an error rate of 0.1% (primarily substitution errors, very rarely insertions/deletions). Illumina has five benchtop sequencers with differing sequencing outputs and

Table 1
Short-read Sequencing Platforms and various characteristics.

Company	Illumina								ThermoFisher		
System Platform	iSeq	Miniseq	MiSeq	NextSeq550	NextSeq 1000&2000	NovaSeq 6000	MiSeqDx	NextSeq550 Dx	GeneStudio S5	Genexus	Ion PGM-Dx
Sequencing Principle	Sequence by Synthesis										
Detection	Fluorescent								Ion		
Applications	Small WGS, TS, Small RNA sequencing		Small WGS, TS, ChIP-Seq, Small RNA sequencing	TS, small WGS, exome and transcriptome sequencing		TS, WGS, WES, transcriptome and epigenome sequencing	TS, Small WGS	TS, exome and transcriptome sequencing	TS, epigenetic, exome, and transciptome sequencing	TS	TS
Maximum Read length (bases)	2 × 150 bp		2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2× 300 bp	2× 150 bp	600 bp	400 bp	200 bp
Flow cells/device	1					2	1				
Output (per flow cell)	1.2 Gb	7.5 Gb	15 Gb	120 Gb	330 Gb	3000 Gb	≥5 Gb	≥90 Gb	15 Gb	24 Gb	1 Gb
Sequencing Run time	9.5-19 hr	5-24 hr	4-56 hr	11-29 hr	11-48 hr	13-44 hr	24 hr	≤35 hr	4.5-21.5 hr	14-31 hr	4.4 hr
Accuracy/Quality Score	Q30≥ 80% (2 × 150 bp)		Q30≥ 70% (2 × 300 bp)	Q30≥ 75% (2 × 150 bp)		Q30≥ 75% (2 × 250 bp)	Accuracy >99.66%, Q30> 80%	Accuracy ≥99.98%, Q30≥75%	Accuracy ≥99%	Accuracy ≥99%	Accuracy ≥99%
Equipment Cost (USD)	$19,900	$49,500	$99,000	$275,000	$335,000	on request					
Dimensions	42.5 × 30.5 × 33 cm	45.6 × 48 × 51.8 cm	68.6 × 56.5 × 52.3 cm	53.3 × 63.5 × 58.4 cm	92 × 120 × 118 cm	80 × 94.5 × 165.6 cm	68.6 × 56.5 × 52.3 cm	54 × 69 × 58 cm	54.2 x 80.6 x 50.9 cm	81.5 × 106.5 × 167.8 cm	61 × 53 × 51 cm
Weight	15.9 kg	45 kg	57.2 kg	83 kg	141 kg	481 kg	54.4 kg	84.4 kg	63.5 kg	204.1 kg	30 kg
Advantages	High accuracy with good depth of coverage										
Disadvantages	Long running time with phasing difficulties										

WGS = whole genome sequencing, WES = whole exome sequencing. TS = targeted sequencing

total reads per run (iSeq, MiniSeq, MiSeq, NextSeq 550, NextSeq 1000, and NextSeq 2000) [40] and two *in-vitro* diagnostic instruments (MiSeqDx and NextSeq 550Dx) [41]. A weakness of Illumina platforms is the relatively long run time, with a total sequencing time of up to 56 h for MiSeq using a reagent kit version 3 with a read length of 2 × 300 bp. More recent models, such as MiniSeq, NextSeq 550, and NovaSeq 6000, use two-channel SBS technology with two images per cycle to make all four base calls, which reduces sequencing time and data processing while maintaining high quality and accuracy [42].

2.3. Data analysis

Given that massively parallel sequencing produces large volumes of data, streamlined bioinformatics data analysis, and data management are essential for implementing these technologies. The data analysis workflow is typically subdivided into primary, secondary, and tertiary analyses [43-45].

Primary analysis is usually performed by the instrument software following sequencing and involves base-calling for each clonally amplified DNA fragment. Quality control procedures, such as

read filtering and trimming, also happen in this phase. Sequence information is recorded along with the quality scores (Phred values) and stored in a FASTQ format. Specifically for the Illumina platform, paired-end sequencing generates two FASTQ files linked through sequence identifiers. Sequencing run quality on the Illumina instruments is assessed using three indicators: cluster density, percentage of clusters passing filters, and percentage of base calls with a quality score of at least Q30 (1 in 1000 probability of an incorrect base call).

Secondary analysis includes read alignment and variant calling. Short reads, either single-end or paired-end, are stored in FASTQ files and are first aligned against the human reference genome (currently Genome Reference Consortium Human Build 38). The Burrows-Wheeler Aligners (BWA) tool, commonly used for fast and accurate alignment, is based on a hash-table algorithm allowing for gapped alignment [46]. The alignment results are stored in a binary alignment/map (BAM) format. Alignments can be viewed using user-friendly and freely available software, such as the Interactive Genome Viewer (IGV) [47]. Another quality indicator for evaluation at this stage is coverage, which includes depth of coverage (number of times a base is sequenced) and breadth of coverage (percentage of a reference genome covered). Coverage ensures that sensitivity and specificity are sufficient for supporting variant detection [48].

After read alignment, the next step is variant calling. Since reads are aligned, the variants, SNPs, indels, or larger structural variants, can be identified by comparing the sample to a reference genome. Open source tools, such as GATK and Freebayes [49,50], are available for this analysis. The sequence variation data are stored in the Variant Call Format (VCF).

Tertiary NGS analysis involves variant annotation and interpretation. This analysis usually involves functional annotation of discovered variants (e.g., SNP, INDEL, and CNV interpretation) to determine their biological and pathological functions. ANNOVAR and VAT are common tools for this analysis [51-53].

Regarding HLA typing, specialized software programs have been developed for secondary and tertiary analyses. Due to the high density of polymorphisms throughout the length of the HLA genes, alignment to the human reference genome is inefficient for accurately determining the HLA alleles present within a sample, and instead rely on alignment to the IMGT/HLA database, a dictionary of all known HLA alleles [54]. Depending on the protocol, coverage profiles may not be uniform throughout the amplicon, and lack of coverage in key regions such as exons may affect the accuracy of the typing. Software analysis programs typically have built-in filters to define the minimum coverage required for accurate typing, although certain circumstances may permit going below this threshold, such as when the polymorphisms of two alleles of a locus are phased, or when the region with low coverage does not affect the genotyping (i.e., introns, untranslated regions) [25]. An additional consideration for HLA typing data analysis is the assessment for adequate allele balance to detect issues due to allele dropout, caused either by preferential amplification due to technical issues or a disease state whereby one of the two alleles has been eliminated.

Similar to DNA-Seq analysis, RNA-Seq data analysis involves primary (base calling), secondary (reads mapping and transcriptome reconstruction), and tertiary (expression quantification and differential expression analysis) analyses [27,55,56]. However, RNA-Seq analysis can be more challenging given the complexities of alternative splicing and the dynamic range of gene expression. Read mapping is a crucial first step in transcriptome profiling since over 90% of human transcripts cross exon–intron junctions. Instead of the conventional mapping tool BWA, the more commonly used tools are TopHat and STAR [57-59]. After alignment, the mapped reads can be assembled into transcripts. The reconstruction of transcripts from short-read data can be challenging, for which two strategies have been employed: reference-guided and reference-independent approaches, the former assembling overlapping reads on the reference into transcripts using tools like Cufflinks [60], and the latter making *de novo* reconstruction of transcripts using tools such as Trinity [61]. Many application tools are available for transcript quantification and differential expression analysis, such as RSEM, Cufflinks, edgeR, and DESeq2 [60,62,63]. Differentially expressed genes can be visualized using heatmaps and clustering. More advanced analyses are usually required for functional interpretation of differential expressed genes, such as gene ontology (GO), pathway, and network enrichment analysis.

2.4. Benefits and limitations

The hallmark of NGS is high-throughput, multiplexed, and clonal sequencing [6]. NGS has several advantages over traditional Sanger sequencing [64]. Gene and sample multiplexing dramatically reduce the cost per sample. Additionally, NGS can resolve most phasing problems encountered in Sanger sequencing due to clonal sequencing of haploid fragments. Short-read sequencing has dominated the current sequencing market. With the advancement of NGS technologies, the third-generation sequencing platforms have gained more attention given longer read lengths and real-time single-molecule sequencing [65]. Compared to long-read sequencing, short-read sequencing limitations include longer running times and difficulties with *de novo* assembly, haplotype phasing, and identification of transcript isoforms and structural variants. An advantage of short-read sequencing is that many computational tools are designed and dedicated to short-read data mining. Until recently, short-read sequencing also had the advantage of being more accurate than long-read sequencing.

3. Long-read sequencing

Compared to short-read sequencing technologies, long-read technologies, also called third-generation sequencing technologies, can generate sequences > 10 kb directly from native DNA. While early iterations of these technologies were plagued with inaccuracies, more recent modifications and improvements have enabled much higher accuracy and offer exciting possibilities to sequence large DNA molecules, such as for the diagnosis of genetic diseases [66]. Long-read sequencing is particularly desirable for HLA genotyping because it would permit complete phasing of alleles and address the issue of typing ambiguities. Two primary long-read technologies, Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT), which operate on different principles, are reviewed. Key characteristics of each of these two types of sequencing platforms are presented in Table 2.

3.1. Pacific biosciences (PacBio)

3.1.1. Technology

Pacific Biosciences (PacBio) or SMRT (Single Molecule, Real-Time) sequencing is a third-generation sequencing method. In this sequencing process, the DNA to be sequenced exists as a single-stranded circular DNA, termed SMRTbell template. The SMRTbell template is generated by ligation of hairpin adaptors (SMRTbell adapters) to both ends of the double-stranded DNA (dsDNA) template molecule. The sequencing reaction occurs in an “SMRT Cell” chip with many small pores called zero-mode waveguides (ZMW), each approximately 70 nm in diameter and 100 nm in depth. PacBio sequencing platforms include the older PacBio RS II system, which has 150,000 ZMWs per SMRT cell, and the newer

Table 2
Long-read Sequencing Platforms and various characteristics.

Company	Pacific Biosciences			Oxford Nanopore			
System Platform	Sequel	Sequel II	Sequel IIe	Flongle	MinION	GridION	PromethION
Sequencing Principle	PacBio Single Molecule Sequencing			Nanopore Single molecule Sequencing			
Detection	Fluorescent			Electrical Conductivity			
Applications	Whole genome *de novo* assembly, variant detection, structural variation detection, full length transcript sequencing, targeted/amplicon sequencing, metagenomics sequencing			DNA, amplicons, cDNA, Direct RNA sequencing			
Maximum Read length (bases)	300 kb			Longest read so far: > 4 Mb			
Flow cells/device	12 SMRT Cells 1M can be used at a time, and 8 SMRT Cell 8M can be used serially			1 (126 channels per flow cell)	1 (512 channels per flow cell)	5 (512 channels per flow cell)	24 or 48 (3000 channels per flow cell)
Output (per flow cell)	75 Gb	600 Gb		1 - 2 Gb[a]	10 - 30 - 50 Gb[a]		100 - 200 - 300 Gb[a]
Sequencing Run time	Up to 20 hr	Up to 30 hr		1 min - 16 hr	1 min - 72 hr		
Accuracy/Quality Score	Number of HiFi Reads >99% Accuracy: Up to 5,000,000 reads	Number of HiFi Reads >99% Accuracy: Up to 4,000,000 reads		**Single Molecule:** R9 modal Accuracy >98.3%, R10 modal Accuracy >97.5%. New chemistry Accuracy >99% (coming soon) **Consensus:** R9.4.1: Current best Q45 (>99.99%) R10: Current Best Q50 (99.999%)			
Equipment Cost (USD)	approximately $525,000			$1,460 (12 flow cells included)	$9,300	$69,955	24 flow cells: $335,455 48 flow cells: $530,000
Dimensions	92.7 x 91.4 x 167.6 cm			105 x 23 x 8 mm	Mk1b: 105 x 23 x 33 mm; Mk1c: 140 x 30 x 116 mm	365 x 220 x 360 mm	Sequencer: 590 x 190 x 430 mm; Data Acquisition unit: 178 x 440 x 470 mm
Weight	362 kg			20 g	Mk1b: 87 g Mk1c: 450 g	11 kg	Sequencer: 28 kg; Data Acquisition unit: 25 kg
Advantages	Very long reads can help resolve ambiguities; no DNA amplification required, comparatively faster turnaround time			Fast-sequencing; Small instrument footprint; Portability; Real-time data analysis			
Disadvantages	Sequencing equipment is expensive, which could be cost-prohibitive for smaller clinical laboratories, large footprint of the equipment, historically higher error rate (continues to improve)			Historically higher error rate (continues to improve)			

[a]Yields depend on sample and preparation methods. Outputs in the following ranges per flow cell using latest chemistries and protocols

sequel system, which has a million ZMWs per SMRT cell [67-70]. An individual DNA polymerase molecule immobilized in each ZMW enables the sequencing of a single SMRTbell template.

The SMRTbell library is loaded in the SMRT Cell, the polymerase binds to the adapter of the SMRTbell, and replication begins. Four fluorescently labeled A, T, G, and C nucleotides with unique emission spectra are used during replication. The binding of a nucleotide to the polymerase produces a signature light pulse, which occurs during replications across all ZMWs in the SMRT cell and is captured in a "movie." Light pulses are interpreted as nucleotide sequences, and the sequence obtained from each ZMW is called a "Continuous Long Read" (CLR). Since the hairpin adaptors make the DNA template circular, the polymerase can continue sequencing through the adapter to replicate the second DNA strand.

Sequencing of a DNA strand once is referred to as a "pass," and the DNA may be sequenced multiple times as "passes." The CLR sequence can be broken down during analysis, whereby the adapter sequence is discarded, and the DNA template existing between the adapters is retained, which is referred to as a subread. Thus, each pass produces a subread, and multiple subreads are produced from multiple passes. Each polymerase molecule has a limited lifetime for which it can effectively sequence the SMRTbell template; therefore, more passes are possible over shorter DNA templates compared to longer DNA templates [67-70]. An animation of this technology can be found on the PacBio website [71].

If multiple subreads exist from a CLR, these can be collapsed down to create a single-molecule circular consensus sequence (CCS, also known as HiFi), whereby the accuracy of the CCS will be improved over the individual accuracy of each subread because the random errors in each subread will be corrected by the other subreads. Given the improvements to the technology that have been accomplished with the Sequel II system, highly accurate, long reads are possible, with reports of up to 99.8% accuracy. Additionally, this method has a precision rate of 99.91% for single-nucleotide variants (SNVs), insertions and deletions (95.98%), and structural variants (95.99%) when the human HG002/NA24385 genome was sequenced [72]. Although HiFi sequencing provides shorter reads than traditional PacBio sequencing, which can provide up to 60 kb reads [67], HiFi reads are more accurate [72] and still provide many folds longer reads (13.5 ± 1.2 kb) compared to short-read methods. HiFi sequencing is available on the Sequel IIe system [73].

PacBio technology can also be used for RNA sequencing by a technique termed Iso-Seq. Using the Iso-Seq method, entire transcripts, including any isoforms, can be sequenced. In this method, RNA is converted to cDNA, and HiFi sequencing is used to generate sequencing data. Reference sequences are then used to identify transcript isoforms.

3.1.2. Data analysis

Analysis of PacBio data typically requires software that can assemble large-sized reads accurately. For *de novo* assembly, PacBio sequence data can be analyzed with software such as HGAP (used for assembling the reads and polishing), Falcon (diploid assembly program), Canu (useful for single-molecule sequencing data with high noise), Sprai (used to assemble larger contigs), Celera® Assembler (allows assembly of subreads), MHAP (detection of overlaps) in addition to SMRT® Analysis suite offered by Pacific Biosciences [74].

RNA sequencing analysis can be done by the Iso-seq analysis tool offered by PacBio. Additionally, tools such as SQANTI, TAMA, DNAnexus, Comptutomics, and LoReAn are useful for this analysis [75].

3.1.3. Advantages and disadvantages

As a long-read technology, the PacBio technology has several advantages. Generated reads can be very long. It has been reported that the top 5% of reads can be greater than 135 kb in length [76]. Given this feature, phasing polymorphic genes such as the HLA genes is possible. With this method, data is collected in real-time, therefore offering a faster turnaround time than second-generation methods. Additionally, while targeted short-read methods frequently require PCR amplification, PacBio does not require DNA amplification and thus avoids pitfalls (AT and GC rich regions amplification difficulty) associated with PCR. This method can be used for DNA and RNA sequencing to identify DNA modifications N^6-methyladenine (m^6A) and N^4-methylcytosine (m^4C) and novel isoforms.

Until more recently, a significant disadvantage of the PacBio method was the high level of error (~14%). To overcome high error-rate, hybrid sequencing approaches that combine short-read and PacBio methods have been used [77]. As reviewed above, the high accuracy of HiFi sequencing has led to decreased error rates seen with this technology. Additional disadvantages, particularly compared to Oxford Nanopore Technology sequencing, as reviewed below, are the more considerable financial investment required for start-up and the comparatively large size of these instruments.

More recent reports describe the use of unique molecular identifiers (UMI) for PacBio CCS and Oxford Nanopore Technologies (ONT) to provide high-accuracy single-molecule consensus sequences of large genomic regions. UMIs are oligonucleotides with sequences of random bases that tag each template molecule in the sample for subsequent sorting and analysis of reads to filter out chimeras. This approach has been reported to decrease the consensus error rate of PacBio CCS and ONT to 0.0041% and 0.0007%, respectively [78].

3.2. Oxford nanopore technology (ONT)

3.2.1. Technology

Oxford Nanopore Technology (ONT) sequencing can generate reads greater than 1 Mb [79] and computationally stitched together, greater than 2 Mb [80]. ONT sequencing is based on the passage of single-stranded nucleic acid (DNA or RNA) through a staphylococcal α-hemolysin (αHL) protein pore [81]. Adaptor ligation to double-stranded DNA facilitates its capture by the protein pore. The libraries are loaded onto a flow cell containing a membrane embedded with hundreds to thousands of nanopores. A preloaded motor enzyme on the adapter at the 5′ end, along with an applied ion current, moves the single strand through the pore. The passage of each nucleotide through the pore results in a characteristic disruption in ion current detected by sensors and is recorded. Animations of this technology can be viewed on the ONT media resources webpage [82].

This technology's pore chemistry allows the uninterrupted traversing of long sequences, with the limiting factor being the preparation of high molecular weight DNA [79,83,84], which determines standard long reads (10–100 kb) vs. ultra-long reads (>100 kb). Both standard long and ultra-long reads are reported to have an accuracy of 87–98%, and about 91% and 93% of homopolymers at least five bases long are called accurately in raw reads, respectively [66]. The 92–93% accuracy for ONT reads limits this technology for single nucleotide variant calling [83]. Sequencing errors are likely related to the motor enzyme and fluctuation of the nucleotide procession rate, whereby faster rates may result in missed bases, and shorter rates may result in a single base being detected as a repetitive sequence. Additionally, the pore's physical properties are such that up to five neighboring bases on the DNA strand may affect the instruments' ion current levels. The accuracy of ONT raw reads is dependent on the base-calling (translation of the electrical signal to DNA sequence) algorithm that is used, which continues to improve over time [85].

Standard long reads can be generated on three ONT platforms, which differ in their flow cell capacity. The MinION, a portable, pocket-sized device, holds one flow cell, and the GridION holds up to five flow cells of the same type with 512 channels, each channel containing four nanopores, totaling 2048 nanopores per flow cell. Given that there is only one active pore per channel that generates sequences at a given time, during which the other three are inactive, the number of channels corresponds to the number of DNA molecules that can be simultaneously sequenced [86]. Please see Figure 1D of the reference by Ip et al. 2015 for an illustration of this concept. The PromethION can hold up to 48 flow cells of a different kind, with 12,000 nanopores divided over 3000 channels. Consequently, the PromethION, which has approximately six times

as many channels, can deliver six-fold more throughput per flow cell and generate 50–100 Gb long-read data per flow cell vs. 2–20 Gb of long-read data compared to the two other platforms [66]. For applications where low-throughput is adequate, a flow cell dongle, or "Flongle," is compatible with the MinION or GridION and uses a flow cell with 126 nanopores, each with its own channel, which permits sequencing with all nanopores. The Flongle allows for low-cost and rapid sequencing. When coupled with the MinION, the Flongle allows for a portable approach to sequencing in various fields. In contrast to technologies such as the MiSeq, sequencing analysis can occur in real-time, shortening turnaround times and permitting clinically impactful testing not previously possible. Recent reports describe turnaround times of less than six hours for high-resolution HLA typing that would permit the HLA typing of deceased donors [87,88].

Also generated through these platforms are ultra-long reads, which are reads over 100 kb and reportedly up to several megabases long [80]. Ultra-long reads permitted the complete phasing of the 4 Mb major histocompatibility complex region, which would not otherwise be accomplished with short-read sequencing [83]. In contrast to standard long reads, however, the throughput for ultra-long reads is lower, with generally 500 Mb to 2 Gb of ultra-long read data per flow cell on the MinION and GridION. To date, the generation of ultra-long reads has not been successful with the PromethION [84], likely due to the lack of compatible sequencing kits for such long DNA fragments [66].

Beyond DNA sequencing, ONT may be used to sequence RNA and detect DNA and RNA modifications. Similar to PacBio, ONT can sequence full-length RNA as cDNA [89,90]. However, ONT also has the ability to use native RNA, which has advantages over cDNA synthesis by including long transcripts that may otherwise be missed and avoiding PCR amplification biases [90]. Modifications to native DNA or RNA are detected from the characteristic disruptions in ion current that result from these modifications.

As described earlier, the error rates of ONT and PacBio sequencing are significantly higher than current second-generation methods. An acceptable error rate is the one that our software systems can potentially neutralize computationally. The current status of ONT in the context of Immunogenetics is that our software systems can handle the error rate of these platforms [88]. However, there is uncertainty associated with this approach. The error rates of PacBio HiFi sequencing are approaching appropriateness for clinical utility [91]. Any technological advancements that will reduce the error rates are welcome.

3.2.2. Data analysis

After introducing this technology, ONT launched the MinION Access Programme (MAP), a beta-testing program for a developer community of more than 1000 laboratories that allowed research groups to evaluate the base throughput, read quality, and performance of the platform [92]. This community has developed several computational approaches and algorithms to manage and analyze ONT data, including base-calling, read mapping, *de novo* assembly, and variant detection and discovery.

The operating software of ONT devices is MinKNOW, which carries out data acquisition, real-time analysis, and local base-calling to produce raw signal data (FAST5 file) that can be used for base-calling. Examples of base-calling software include Albacore, Guppy, Scrappie, and Flappy [93]. Improved base-calling algorithms, which are applied to raw data, allow sequencing accuracy to increase [85].

A developing unique feature of ONT sequencing is adaptive sampling of reads during the sequencing experiment [94,95], which allows the operator to specify regions of the genome to target based on software configuration. If the DNA strand entering the pore is not of interest, it is flagged by the real-time data-streaming and ejected from the nanopore. As the DNA strand of interest is processed, the system identifies it as the target strand, and sequencing continues. Adaptive sampling, therefore, enables the system only to select the strands of interest and reject those that are irrelevant, resulting in high coverage of sequencing data for the region of interest without the need for extensive sample preparation.

3.2.3. Advantages and disadvantages

The general advantages of ONT as a long-read technology compared to short-read technologies are similar to those of PacBio. These include better phasing of polymorphic genes, detection and proper characterization of structural rearrangements, data collection in real-time, and faster turnaround times. Additionally, since native DNA is used, any errors introduced during the DNA amplification process of short-reads are eliminated. When compared to PacBio or second-generation sequencing platforms, ONT instruments have the advantages of being lower cost, portable, and of significantly smaller size, characteristics that can be very useful for low-income settings or field applications [96].

General disadvantages of ONT as a long-read technology include the fact that the data is subject to signal-to-noise constraints that result in a higher level of error (2–15%) relative to that seen with short-read technologies. Modifications to the nanopore protein to allow for slower nucleotide procession rates have permitted better data acquisition and improved error rates. Furthermore, these high-error rates decrease over time with improved base calling models [97]. Additionally, ONT errors are primarily systematic, which cannot be resolved as easily as random errors from increased coverage [97].

4. Conclusion and future directions

Since the first fully sequenced human genome, sequencing underwent a paradigm shift with technological changes that enabled massive data output and improved efficiencies. Ultimately, the goal of ongoing technological advancement is to better serve research and clinical applications. For a sequencing technology to be used in a clinical setting, the technology must provide sensitive, specific, and reproducible results within a certain timeframe and at a reasonable cost. Massive amounts of generated data must be reliably distilled and appropriately formatted to provide clinically meaningful and actionable results. The ability to satisfy these factors has enabled short-read sequencing to traverse from the research to the clinical setting readily, including detecting rare variants in hereditary disorders [98], diagnosis and management of patients with solid tumors or hematologic malignancies [99,100], and HLA typing for transplant, pharmacogenomics, and disease associations [15,39,101-103].

Despite highly accurate sequencing data that permits the described applications, short-read sequencing technologies remain limited due to PCR amplification and short read-lengths. This technological limitation halts ongoing clinical progress. Third-generation technologies afford much longer read-lengths from native DNA, permitting a more uniform, direct assessment of the sequence and circumventing the limitations of short-read technologies. The main limiting factor of these third-generation technologies was a high-error rate that made them clinically unsuitable. Continual refinement of these third-generation technologies and bioinformatics tools to increase accuracy holds promise for the next wave of advancements with significant impact. Not only will there be cost-effective, highly accurate long reads with a faster turnaround in the clinical setting, but more importantly, critical genetic information will be available as complete haplotypes will be generated with all detected variants resolved.

Long-read sequencing technologies can reliably detect structural variations frequently associated with pathological phenotypes, a task that is practically impossible with short-read technologies.

With the ability of these technologies to provide both accurate sequencing of long DNA fragments and information on methylation patterns and chromosomal inactivation, these technologies will add a dimension to interpretation beyond the base-pair characterization of the DNA sequence, bringing us closer to the physiological state of the cell. Meanwhile, advances characterizing the different RNA species of the cell using RNA-seq technologies will further promote our understanding of gene expression and its role in the physiology of different cell types and their interactions. Although DNA sequencing technologies have been used extensively for clinical immunogenetics to advance the field, RNA characterization technologies that have been used for research Immunogenetics [104,105] have yet to be used for immunogenetics diagnostic applications. It is anticipated that advancements in RNA profiling and the study of interactive relationships between different regions within the MHC, and therefore of gene expression, will promote our understanding of the role of other genes and regions besides HLAs, and provide a more comprehensive picture of MHC-controlled immune responses. The gained knowledge would naturally lead to new clinical applications.

In conclusion, new developments in sequencing technologies that hold the promise of robust, reproducible, accurate sequencing of long fragments at even higher throughput and lower cost will likely be the next wave of sequencing platforms [106,107]. This third generation of sequencers may form the basis for yet another wave of unprecedented discoveries, as the characterization of the genomes and transcriptomes of many organisms and large populations will become feasible, promoting not only scientific discoveries but also other applications with significant economic and health-related impact.

Declaration of Competing Interest

Dimitri Monos receives royalties from, is a consultant to, and owns options in Omixon Inc. The remaining authors have no conflicts of interest to disclose.

Acknowledgment

We thank Jamie L. Duke, Deborah Ferriola, and Timothy L. Mosbruger for their review of this manuscript and helpful suggestions.

References

[1] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid, Nature 171 (1953) 737.
[2] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, PNAS 74 (1977) 5463.
[3] NHGRI. The Cost of Sequencing a Human Genome. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
[4] S. Levy, G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, et al., The Diploid Genome Sequence of an Individual Human, PLoS Biol. 5 (2007) e254.
[5] E.R. Mardis, Next-generation sequencing platforms, Annu Rev Anal Chem (Palo Alto Calif) 6 (2013) 287.
[6] T. Tucker, M. Marra, J.M. Friedman, Massively parallel sequencing: the next big thing in genetic medicine, Am. J. Hum. Genet. 85 (2009) 142.
[7] Illumina. Next-Generation Sequencing for Beginners. https://www.illumina.com/science/technology/next-generation-sequencing/beginners.html
[8] S.R. Head, H.K. Komori, S.A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D.R. Salomon, et al., Library construction for next-generation sequencing: overviews and challenges, Biotechniques 56 (2014) 61.
[9] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57.
[10] G. Lightbody, V. Haberland, F. Browne, L. Taggart, H. Zheng, E. Parkes, et al., Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application, Briefings Bioinf. 20 (2019) 1795.
[11] J. Shendure, S. Balasubramanian, G.M. Church, W. Gilbert, J. Rogers, J.A. Schloss, et al., DNA sequencing at 40: past, present and future, Nature 550 (2017) 345.
[12] J.M. Rizzo, M.J. Buck, Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing, Cancer Prevention Research 5 (2012) 887.
[13] E.L. van Dijk, Y. Jaszczyszyn, C. Thermes, Library preparation methods for next-generation sequencing: tone down the bias, Exp Cell Res 322 (2014) 12.
[14] G. Bentley, R. Higuchi, B. Hoglund, D. Goodridge, D. Sayer, E.A. Trachtenberg, et al., High-resolution, high-throughput HLA genotyping by next-generation sequencing, Tissue Antigens 74 (2009) 393.
[15] 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.
[16] C.D.M. Meldrum, R.W. Tothill, Next-generation sequencing for cancer diagnostics: a practical perspective, Clin Biochem Rev 32 (2011) 177.
[17] W.S. Liang, K. Stephenson, J. Adkins, A. Christofferson, A. Helland, L. Cuyugan, et al., Whole Exome Library Construction for Next Generation Sequencing, Methods Mol Biol 1706 (2018) 163.
[18] L. Mamanova, A.J. Coffey, C.E. Scott, I. Kozarewa, E.H. Turner, A. Kumar, et al., Target-enrichment strategies for next-generation sequencing, Nat Methods 7 (2010) 111.
[19] J. Dapprich, D. Ferriola, K. Mackiewicz, P.M. Clark, E. Rappaport, M. D'Arcy, et al., The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity, BMC Genomics 17 (2016) 486.
[20] T. Gupta, F.L. Marlow, D. Ferriola, K. Mackiewicz, J. Dapprich, D. Monos, et al., Microtubule Actin Crosslinking Factor 1 Regulates the Balbiani Body and Animal-Vegetal Polarity of the Zebrafish Oocyte, PLoS Genet. 6 (2010) e1001073.
[21] S.E. Pinney, K. Ganapathy, J. Bradfield, D. Stokes, A. Sasson, K. Mackiewicz, et al., Dominant form of congenital hyperinsulinism maps to HK1 region on 10q, Hormone research in paediatrics 80 (2013) 18.
[22] S. Sarda, S. Hannenhalli, Next-generation sequencing and epigenomics research: a hammer in search of nails, Genomics & informatics 12 (2014) 2.
[23] E. Meaburn, R. Schulz, Next generation sequencing in epigenetics: insights and challenges, Semin Cell Dev Biol 23 (2012) 192.
[24] J. Podnar, H. Deiderick, S. Hunicke-Smith, Next-generation sequencing fragment library construction, Curr Protoc Mol Biol (2014;107:7 17 1.).
[25] 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.
[26] J. Podnar, H. Deiderick, G. Huerta, S. Hunicke-Smith, Next-Generation Sequencing RNA-Seq Library Construction, Curr Protoc Mol Biol (2014;106:4 21 1.).
[27] M.I. Love, S. Anders, V. Kim, W. Huber, RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000Research 2015;4:1070.
[28] K.-O. Mutz, A. Heilkenbrinker, M. Lönne, J.-G. Walter, F. Stahl, Transcriptome analysis using next-generation sequencing, Curr. Opin. Biotechnol. 24 (2013) 22.
[29] Illumina. Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/RNA/illumina-stranded-total-rna-reference-1000000124514-01.pdf
[30] Illumina. TruSeq Stranded mRNA Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-mrna-workflow/truseq-stranded-mrna-workflow-reference-1000000040498-00.pdf
[31] Illumina. TruSeq® Small RNA Library Prep Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqsmallrna/truseq-small-rna-library-prep-kit-reference-guide-15004197-02.pdf
[32] E.L. van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, Ten years of next-generation sequencing technology, Trends Genet 30 (2014) 418.
[33] S.E. Levy, R.M. Myers, Advancements in Next-Generation Sequencing, Annu Rev Genomics Hum Genet 17 (2016) 95.
[34] S. Goodwin, J.D. McPherson, W.R. McCombie, Coming of age: ten years of next-generation sequencing technologies, Nat. Rev. Genet. 17 (2016) 333.
[35] ThermoFisher. Ion GeneStudio S5 Next-Generation Sequencing Series Specifications. https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-s5-ngs-targeted-sequencing/ion-s5-specifications.html
[36] B. Merriman, D Team IT, Rothberg JM: Progress in Ion Torrent semiconductor chip based sequencing, Electrophoresis 33 (2012) 3397.
[37] ThermoFisher. Ion Torrent Next-Generation Sequencing Instruments. https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence.html
[38] Illumina. Explore Illumina sequencing technology https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html
[39] 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.

[40] Illumina. Benchtop Sequencers. https://www.illumina.com/systems/sequencing-platforms.html
[41] Illumina. Bringing next-generation sequencing to clinical labs. https://www.illumina.com/systems/ivd-instruments.html
[42] Illumina. Faster sequencing and data processing. https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html
[43] R. Pereira, J. Oliveira, M. Sousa, Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics, J Clin Med 9 (2020).
[44] S. Klasberg, V. Surendranath, V. Lange, G. Schöfl, Bioinformatics Strategies, Challenges, and Opportunities for Next Generation Sequencing-Based HLA Genotyping, Transfusion Medicine and Hemotherapy 46 (2019) 312.
[45] M.P. Dolled-Filhart, M. Lee, Ou-yang C-w, Haraksingh RR, Lin JC-H: Computational and Bioinformatics Frameworks for Next-Generation Whole Exome and Genome Sequencing, The Scientific World Journal 2013 (2013) 730210.
[46] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics (Oxford, England) 25 (2009) 1754.
[47] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Briefings Bioinf. 14 (2013) 178.
[48] 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.
[49] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res 20 (2010) 1297.
[50] S. Sandmann, A.O. de Graaf, M. Karimi, B.A. van der Reijden, E. Hellström-Lindberg, J.H. Jansen, et al., Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data, Sci. Rep. 7 (2017) 43169.
[51] H. Yang, K. Wang, Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR, Nat. Protoc. 10 (2015) 1556.
[52] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res 38 (2010) e164.
[53] L. Habegger, S. Balasubramanian, D.Z. Chen, E. Khurana, A. Sboner, A. Harmanci, et al., VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment, Bioinformatics (Oxford, England) 28 (2012) 2267.
[54] J. Robinson, D.J. Barker, X. Georgiou, M.A. Cooper, P. Flicek, S.G.E. Marsh, IPD-IMGT/HLA Database, Nucleic Acids Res. 48 (2019) D948.
[55] F. Ji, R.I. Sadreyev, RNA-seq: Basic Bioinformatics Analysis, Curr Protoc Mol Biol 124 (2018) e68.
[56] M. Griffith, J.R. Walker, N.C. Spies, B.J. Ainscough, O.L. Griffith, Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud, PLoS Comput. Biol. 11 (2015) e1004393.
[57] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (2013) R36.
[58] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics (Oxford, England) 29 (2013) 15.
[59] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics (Oxford, England) 25 (2009) 1105.
[60] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, et al., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nat. Protoc. 7 (2012) 562.
[61] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644.
[62] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.
[63] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[64] M. Bahassi el, P.J. Stambrook, Next-generation sequencing technologies: breaking the sound barrier of human genetics, Mutagenesis 29 (2014) 303.
[65] E.L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, The Third Revolution in Sequencing Technology, Trends Genet 34 (2018) 666.
[66] G.A. Logsdon, M.R. Vollger, E.E. Eichler, Long-read human genome sequencing and its applications, Nat Rev Genet 21 (2020) 597.
[67] A. Rhoads, K.F. Au, PacBio Sequencing and Its Applications, Genomics Proteomics Bioinformatics 13 (2015) 278.
[68] PACBIO. Introducing the Sequel System: The Scalable Platform for SMRT Sequencing. https://www.pacb.com/blog/introducing-the-sequel-system-the-scalable-platform-for-smrt-sequencing/
[69] PACBIO. SEQUENCE WITH CONFIDENCE. https://www.pacb.com/wp-content/uploads/SMRT-Sequencing-Brochure-Delivering-highly-accurate-long-reads-to-drive-discovery-in-life-science.pdf
[70] PACBIO. PacBio RS II Sequencing System. https://www.mscience.com.au/upload/pages/pacbio/pacbio_rs_ii_brochure.pdf
[71] PacBio. SMRT SEQUENCING. https://www.pacb.com/smrt-science/smrt-sequencing/
[72] 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.
[73] PACBIO. Sequel IIe System Sequencing evolved. https://www.pacb.com/wp-content/uploads/Product-Brochure-Sequel-IIe-System-Sequencing-evolved.pdf
[74] Github. Software packages compatible with PacBio® data. https://github.com/PacificBiosciences/DevNet/wiki/Compatible-Software#denovo%20accessed%2012-4-20
[75] PacBio. HUMAN RNA SEQUENCING. https://www.pacb.com/applications/rna-sequencing/human/
[76] PacBio. Sequel II System v8.0 & SMRT Link v8.0 Technical Overview. https://www.pacb.com/wp-content/uploads/Sequel-II-System-v8.0-and-SMRT-Link-v8.0-Technical-Overview-Customer-Training.pdf
[77] B. Berbers, A. Saltykova, C. Garcia-Graells, P. Philipp, F. Arella, K. Marchal, et al., Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus, Sci. Rep. 10 (2020) 4310.
[78] S.M. Karst, R.M. Ziels, R.H. Kirkegaard, E.A. Sørensen, D. McDonald, Q. Zhu, et al., High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing, Nat. Methods (2021).
[79] K.H. Miga, S. Koren, A. Rhie, M.R. Vollger, A. Gershman, A. Bzikadze, et al., Telomere-to-telomere assembly of a complete human X chromosome. bioRxiv 2019:735928.
[80] A. Payne, N. Holmes, V. Rakyan, M. Loose, BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files, Bioinformatics 35 (2019) 2193.
[81] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, H. Bayley, Continuous base identification for single-molecule nanopore DNA sequencing, Nat. Nanotechnol. 4 (2009) 265.
[82] Nanopore. Nanopore Media Resources. https://nanoporetech.com/about-us/for-the-media
[83] 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.
[84] K. Shafin, T. Pesout, R. Lorig-Roach, M. Haukness, H.E. Olsen, C. Bosworth, et al., Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes, Nat. Biotechnol. 38 (2020) 1044.
[85] F.J. Rang, W.P. Kloosterman, J. de Ridder, From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy, Genome Biol. 19 (2018) 90.
[86] Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain Met al. : MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Res 2015;4:1075.
[87] 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.
[88] T.L. Mosbruger, A. Dinou, J.L. Duke, D. Ferriola, H. Mehler, I. Pagkrati, et al., Utilizing nanopore sequencing technology for the rapid and comprehensive characterization of eleven HLA loci; addressing the need for deceased donor expedited HLA typing, Hum. Immunol. 81 (2020) 413.
[89] S. Oikonomopoulos, Y.C. Wang, H. Djambazian, D. Badescu, J. Ragoussis, Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations, Sci Rep 6 (2016) 31602.
[90] A. Byrne, A.E. Beaudin, H.E. Olsen, M. Jain, C. Cole, T. Palmer, et al., Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells, Nat. Commun. 8 (2017) 16027.
[91] N.P. Mayor, J.D. Hayhurst, T.R. Turner, R.M. Szydlo, B.E. Shaw, W.P. Bultitude, et al., Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study, Biology of Blood and Marrow Transplantation 25 (2019) 443.
[92] A. Magi, R. Semeraro, A. Mingrino, B. Giusti, R. D'Aurizio, Nanopore sequencing data analysis: state of the art, applications and challenges, Briefings Bioinf. 19 (2018) 1256.
[93] R.R. Wick, L.M. Judd, K.E. Holt, Performance of neural network basecalling tools for Oxford Nanopore sequencing, Genome Biol. 20 (2019) 129.
[94] A. Payne, N. Holmes, T. Clarke, R. Munro, B. Debebe, M. Loose, Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. bioRxiv 2020:2020.02.03.926956.
[95] S. Kovaka, Y. Fan, B. Ni, W. Timp, M.C. Schatz, Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED, Nat. Biotechnol. (2020).
[96] J. Quick, N.J. Loman, S. Duraffour, J.T. Simpson, E. Severi, L. Cowley, et al., Real-time, portable genome sequencing for Ebola surveillance, Nature 530 (2016) 228.
[97] W.R. McCombie, J.D. McPherson, E.R. Mardis, Next-Generation Sequencing Technologies. Cold Spring Harb Perspect Med 9 (2019).
[98] S.S. Jamuar, E.C. Tan, Clinical application of next-generation sequencing for Mendelian diseases, Hum Genomics 9 (2015) 10.
[99] K.E. Fisher, L. Zhang, J. Wang, G.H. Smith, S. Newman, T.M. Schneider, et al., Clinical Validation and Implementation of a Targeted Next-Generation Sequencing Assay to Detect Somatic Variants in Non-Small Cell Lung, Melanoma, and Gastrointestinal Malignancies, J Mol Diagn 18 (2016) 299.
[100] N. Wagle, M.F. Berger, M.J. Davis, B. Blumenstiel, M. Defelice, P. Pochanard, et al., High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing, Cancer Discov 2 (2012) 82.
[101] P.T. Illing, A.W. Purcell, J. McCluskey, The role of HLA genes in pharmacogenomics: unravelling HLA associated adverse drug reactions, Immunogenetics 69 (2017) 617.

[102] V. Matzaraki, V. Kumar, C. Wijmenga, A. Zhernakova, The MHC locus and genetic susceptibility to autoimmune and infectious diseases, Genome Biol. 18 (2017) 76.
[103] P. Vogiatzi, Some considerations on the current debate about typing resolution in solid organ transplantation, Transplantation Research 5 (2016) 3.
[104] S. Boegel, M. Löwer, M. Schäfer, T. Bukur, J. de Graaf, V. Boisguérin, et al., HLA typing from RNA-Seq sequence reads, Genome Med. 4 (2012) 102.
[105] F. Yamamoto, S. Suzuki, A. Mizutani, A. Shigenari, S. Ito, Y. Kametani, et al., Capturing Differential Allele-Level Expression and Genotypes of All Classical HLA Loci and Haplotypes by a New Capture RNA-Seq Method, Front. Immunol. 11 (2020).
[106] C.C. Chien, S. Shekar, D.J. Niedzwiecki, K.L. Shepard, M. Drndić, Single-Stranded DNA Translocation Recordings through Solid-State Nanopores on Glass Chips at 10 MHz Measurement Bandwidth, ACS Nano 13 (2019) 10545.
[107] Y.C. Chou, P. Masih Das, D.S. Monos, M. Drndić, Lifetime and Stability of Silicon Nitride Nanopores and Nanopore Arrays for Ionic Measurements, ACS Nano 14 (2020) 6715.