Next-Generation Sequencing and Emerging Technologies

Kishore R. Kumar, MBBS, PhD, FRACP[1,2,3]   Mark J. Cowley, PhD[1,4]   Ryan L. Davis, PhD[1,2]

[1] Translational Genomics Group, Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, New South Wales, Australia

[2] Department of Neurogenetics, Kolling Institute, University of Sydney and Royal North Shore Hospital, St Leonards, New South Wales, Australia

[3] Molecular Medicine Laboratory, Concord Hospital, Sydney, Australia

[4] Computational Biology Group, Children's Cancer Institute, University of New South Wales, Randwick, New South Wales, Australia

Address for correspondence Ryan L. Davis, PhD, Department of Neurogenetics, Kolling Institute, University of Sydney and Royal North Shore Hospital, Level 11, Building 6, Reserve Road, St Leonards, NSW 2065, Australia (e-mail: ryan.davis@sydney.edu.au).

Abstract

Keywords

► next-generation sequencing
► whole-exome sequencing
► whole-genome sequencing
► nanopore
► PacBio
► Illumina
► Ion Torrent
► nanoball
► thrombosis
► hemostasis

Genetic sequencing technologies are evolving at a rapid pace with major implications for research and clinical practice. In this review, the authors provide an updated overview of next-generation sequencing (NGS) and emerging methodologies. NGS has tremendously improved sequencing output while being more time and cost-efficient in comparison to Sanger sequencing. The authors describe short-read sequencing approaches, such as sequencing by synthesis, ion semiconductor sequencing, and nanoball sequencing. Third-generation long-read sequencing now promises to overcome many of the limitations of short-read sequencing, such as the ability to reliably resolve repeat sequences and large genomic rearrangements. By combining complementary methods with massively parallel DNA sequencing, a greater insight into the biological context of disease mechanisms is now possible. Emerging methodologies, such as advances in nanopore technology, in situ nucleic acid sequencing, and microscopy-based sequencing, will continue the rapid evolution of this area. These new technologies hold many potential applications for hematological disorders, with the promise of precision and personalized medical care in the future.

Next-generation sequencing (NGS) is a rapidly advancing area of technology that has propelled the research space and is infiltrating clinical applications with enormous impact. It has revolutionized the ability to sequence nucleic acids by increasing the amount of sequence data that can be obtained in a rapid and cost-effective manner, considerably defying Moore's law.[1] As technologies are continually emerging and improving, it is necessary to understand the capabilities and drawbacks of each method and their clinical and research applicability. With the expansion in sequence output capabilities and reduction in costs, the challenge is not in producing sequence data, but in comprehensively analyzing and rationally interpreting it.

As informatics tools and analysis pipelines improve with time, the requirement of accurate patient phenotyping, variant pathogenicity assessment, functional validation of candidate findings, and informative reporting is critical to obtain a genetic diagnosis. In this review, we provide an overview of NGS techniques and emerging methodologies. Many of these technologies have had a remarkable impact on

Kishore R. Kumar's ORCID is https://orcid.org/0000-0003-3482-6962.

Mark J. Cowley's ORCID is https://orcid.org/0000-0002-9519-5714.

Ryan L. Davis's ORCID is https://orcid.org/0000-0003-0512-8989.

our understanding of disease biology and are serving as rapid, accurate, and economical tools for the diagnosis of inherited genetic disorders. Since the technology and chemistries behind NGS change so frequently, this review is not intended to be exhaustive but rather a snapshot of the past, present, and the foreseeable future.

The Genesis of Nucleic Acid Sequencing

Since the discovery of the structure of DNA in 1953[2] and the realization that the genetic code is the blueprint for life, a necessity to sequence DNA has existed. Despite this, it took nearly a quarter of a century before the first generation of DNA sequencing techniques were developed.[3] Sanger sequencing, as it became known, is a "sequencing by synthesis" approach where the nucleic acid sequence is determined one nucleotide at a time as a function of the DNA template length. Development of the method over a decade saw its automation by Applied Biosystems in 1987, paving the way for the Human Genome Project soon after. The Human Genome Project cost $2.7 billion, required international collaboration on an unprecedented scale, and took a decade to complete.[4–6] By the end of the project in 2001, first-generation Sanger sequencing had reached a theoretical technical threshold, and scaling the process to routinely perform genomic sequencing was unrealistic.[7] With an emphasis on performing routine genome sequencing, a new era of methods emerged, collectively known as NGS.

Basic Concepts
First-generation Sanger sequencing was technically limited in the amount and extent of DNA sequence that could be generated in a sequencing run and was generally restricted to regions of interest within genes of interest (►Fig. 1). With

the advent of NGS, larger regions of the genome could be more comprehensively sequenced, giving rise to multiple approaches. Targeted sequencing approaches (►Fig. 1) focus on specific regions within the genome and can include targets within individual or multiple genes (targeted amplicon sequencing) or specific genes of interest (targeted gene sequencing). Most clinical gene panels use a targeted gene sequencing approach to analyze genetic variation in genes associated with specific phenotypes. A more comprehensive approach to targeted gene sequencing is whole-exome sequencing (WES) (►Fig. 1), which targets the protein coding regions (i.e., exons, hence the name exome) of all genes, accounting for approximately 2% of the entire genome.[8] It is estimated that around 85% of currently known disease-causing variants are found within the exome, largely because these regions have been of primary focus until recently. The most comprehensive genome sequencing approach is whole-genome sequencing (WGS) (►Fig. 1), which produces a sequence of both the coding and noncoding regions. This allows for considered sub-analyses of the other approaches (targeted and WES) in silico, in addition to analysis of the non-coding regions. With developing bioinformatics tools and approaches, interrogating non-coding regions of the genome and transcriptome is becoming increasingly possible and is beginning to show a vast amount of small- and large-scale variation that could hold the key to unknown and unsolved disease (reviewed in[9–11]).

Briefly, DNA (or RNA) to be sequenced by NGS is fragmented into manageable sizes (see the Sample Preparation section), that when sequenced, provide a portion of nucleotide sequence known as a read. Different NGS platforms are generally compared by (1) the read size (35–600 base pair [bp] for second-generation short-read platforms and > 1 kilobase [kb] for third-generation long-read platforms,
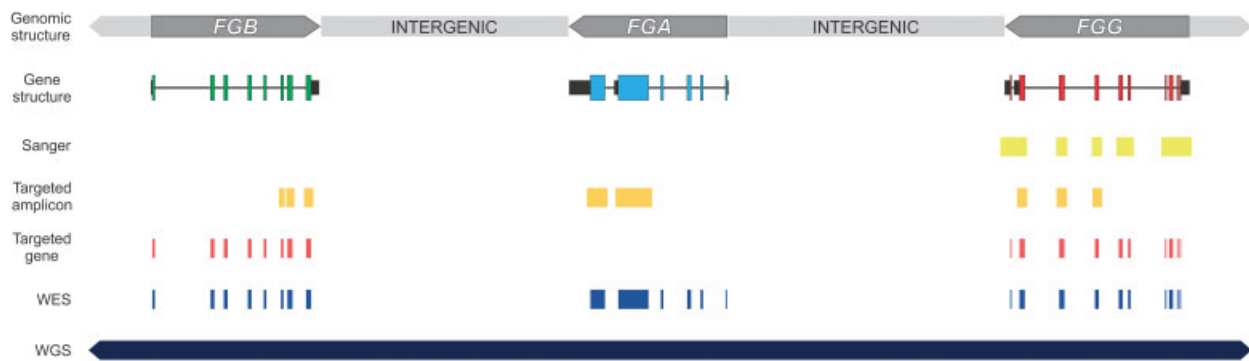

Fig. 1 Sequence output capacity of different next-generation sequencing approaches. From top down: the genomic structure of the fibrinogen gene cluster involves three genes (fibrinogen β gene [*FGB*], fibrinogen α gene [*FGA*], and fibrinogen γ gene [*FGG*]) separated by intergenic regions, with *FGB* on the plus strand and *FGA* and *FGG* on the minus strand of chromosome 4. Structurally, the genes consist of untranslated regions (small black boxes), protein coding exons (colored boxes), and noncoding introns (horizontal line between exons). Due to the technical limitations of first-generation *Sanger sequencing*, it is usually restricted to a gene of interest and can involve amplicons spanning regions of interest up to approximately 800 bp. *Targeted amplicon sequencing* focuses on genomic regions of interest, with several targets being sequenced in parallel at high-read depth. The diagram indicates that 3′ exons are targeted as they encode C-terminal functional domains where mutations often cluster. *Targeted gene sequencing* focuses on genes of particular interest and is generally represented by gene panels associated with a particular disease. Here, *FGB* and *FGG* are the focus of targeted gene sequencing, highlighting that combinations of genes can be sequenced. *Whole-exome sequencing* (WES) provides sequence data enriched for all protein coding exons of the entire genome, offering a cost-effective means of considering pathogenic variants throughout the coding regions of the genome. Finally, *whole-genome sequencing* (WGS) provides comprehensive sequencing data across the entire genome, including the protein coding exome, intronic regions between exons, and intergenic regions between genes.

discussed later), (2) the number of reads that can be sequenced in a run, and (3) the total amount of data that is generated (►**Table 1**). After a sequencing run, raw sequencing reads are reassembled into the correct order by aligning to a known reference sequence or by de novo genome assembly. To do so, reads are stacked on top of each other as they overlap and align to the same reference sequence. The number of unique reads in which a given reference nucleotide appears is known as the read depth and determines the confidence with which a base is called. This also determines whether variant base calls that may be related to the disease can be considered genuine or an error.[7] For instance, a single-nucleotide variant (SNV) in the nuclear genome should occur in more than 10 unique reads (i.e., read depth of $10 \times$) to provide enough confidence that it is a genuine genomic variant and not a technical error from the sequencing process. This sequencing platform-specific parameter defines the type of genetic variants that can be identified and the confidence with which they can be determined.[12]

Next-Generation Sequencing

Next-generation sequencing collectively describes sequencing approaches that have reduced the time and cost and tremendously increased the sequence output when compared with Sanger sequencing. Sanger sequencing is only capable of producing a sequence for one template per reaction, whereas NGS is typified by performing millions to billions of individual sequencing reactions simultaneously in a process referred to as massively parallel sequencing. The approach has revolutionized DNA sequencing to the point that a whole human genome can now be sequenced within 3 days for around US $1,000. Most next-generation technologies currently available (►**Table 1**) are based on the sequencing by synthesis approach, with second- and third-generation sequencing methods defined as short-read and real-time long-read technologies, respectively. Despite the vast differences in NGS methods, the workflow largely follows three broad steps: sample preparation, nucleic acid sequencing, and data analysis. For a review of NGS with comparisons of platforms and their specifications, as well as schematic representations of sequencing methods, refer to Goodwin et al.[13]

Sample Preparation

Library preparation refers to the preparation of nucleic acid templates (DNA or RNA) for sequencing, which differs for each platform. For short-read applications, this is generally achieved in three main steps: (1) DNA fragmentation to application-specific template lengths, (2) ligation of adaptors to facilitate the attachment of fragments to solid surfaces (such as microchips, microbeads, or nanowells) or to be circularized, and (3) amplification of templates to provide enough copies of each template to allow the sequencer to detect them. The libraries can either be sequenced from one end only (known as single-end reads) or from both ends (known as paired-end reads). Library preparation can introduce several technical errors, predominantly related to polymerase chain reaction (PCR) amplification and sequencing

chemistry inefficiencies (i.e., inherent polymerase errors and inefficiencies associated with template guanine–cytosine [GC] content). Recently, PCR-free library preparation has been released for short-read platforms, increasing the fidelity of a template sequence.

Long-read platforms differ in their library preparation as an amplification step is not required, and in some, the original DNA or RNA molecule is directly sequenced. Various library preparation methods also exist for several other applications, including RNA sequencing and chromatin immunoprecipitation sequencing (discussed in the Complementary Methods section).

Nucleic Acid Sequencing

Second-Generation (Short-Read) Sequencing

Short-Read Sequencing by Synthesis (Illumina)
Illumina is the current market leader for short-read NGS, using sequencing by synthesis with optical base calling, reminiscent of Sanger sequencing but on an enormously increased scale.[14] Following library preparation, which affixes adaptors to DNA fragments of approximately 450 bp, the templates are hybridized to a glass slide that has patterned clusters of complementary adaptors. Once attached to the solid surface, the fragments are PCR amplified from one end only (single-end read) or from both ends (paired-end reads), producing millions to billions of clusters of clonal template DNA fragments that can be sequenced simultaneously.[14] One adaptation over Sanger sequencing is the use of reversible terminator nucleotides. This permits one nucleotide to be incorporated at a time and the representative fluorescence to be recorded as a base call by high-resolution optical imaging, followed by cleavage of the terminal chemical modification, thereby allowing the next complementary nucleotide to be incorporated. This process is repeated for the length of the read to generate the sequence output, where read lengths are now typically between 75 to 250 bp.[14,15]

Illumina currently has four benchtop and two production-scale platforms with variable applications, including in vitro diagnostics. When paired with different library preparation kits, each instrument offers a diverse range of throughput, output, cost per base and applications (►**Table 1**). The benchtop sequencers (iSeq, MiniSeq, MiSeq, and NextSeq) cater for smaller experiments, ranging from bacterial genomes to transcriptome and WES, with increasing capacity.[14] The production-scale sequencing platforms (HiSeq and recently released NovaSeq) are primarily used for large-scale analysis of WES and WGS. The patterned flow cell technology of the HiSeq X and 4000 Systems provides a very high level of throughput, with the capability to sequence tens of thousands of genomes per year and recently lowering the cost to below US$1,000 per genome.[16] The Novaseq 6000, the newest addition to the suite of Illumina sequencers, combines a patterned flow cell with a two-color nucleotide detection system that enables enhanced speed, flexibility, and throughput for ultra deep WES and WGS, as well as tumor normal profiling.[16]

Table 1 Comparison of next-generation sequencing instrument parameters

	Instrument	Maximum reads per run	Maximum output range (gigabases)	Read length	Sequencing run time (hours)	Main applications
Illumina	Iseq 100	4 M	0.144	36-bp SE	9	Small genome sequencing, targeted gene sequencing, long-range PCR, amplicon sequencing
			0.2	50-bp SE	9	
			0.3	75-bp SE	10	
		8 M	0.6	75-bp PE	13	
			1.2	150-bp PE	17.5	
	MiniSeq[a]	25 B	1.65–1.875	75-bp SE	7	Small genome sequencing, targeted gene sequencing, targeted gene expression profiling, 16S metagenomic sequencing
		50 B	3.3–3.75	75-bp PE	13	
			6.6–7.5	150-bp PE	24	
	MiSeq[b,c]	25 B	3.3–3.8	75-bp PE	21	Small genome sequencing, targeted gene sequencing, 16S metagenomic sequencing
		50 B	13.2–15	300-bp PE	56	
	NextSeq 550[a,b]	0.4 B	25–30	75-bp PE	11	Small genome sequencing, targeted gene sequencing, mRNA-Seq gene expression profiling, miRNA and small RNA analysis
		0.8 B	50–60	75-bp PE	18	
			100–120	150-bp PE	29	
	HiSeq 3000	2.5 B	105–125	50-bp PE	24–84	Exome Sequencing, whole transcriptome sequencing
		5 B	325–375	75-bp PE		
			650–750	150-bp PE		
	HiSeq 4000	5 B	210–250	50-bp SE	24–84	Exome sequencing, whole transcriptome sequencing
		10 B	650–750	75-bp PE		
			1,300–1,500	150-bp PE		
	HiSeqX[d]	6 B[e]	1,600–1,800[e]	150-bp PE	72	Whole-genome sequencing
	NovaSeq 6000	2.6–3.2 B	134–500	50-bp PE	13–16	Whole-genome sequencing, exome sequencing, whole transcriptome sequencing, methylation sequencing
		6.6–8.2 B	333–1,250	100-bp PE	19–36	
		16–20 B	1,600–3,000	150-bp PE	25–44	
Ion Torrent[f]	Ion 510 Chip	2–3 M	0.3–0.5	200 bp	2.5–4	Targeted, exome, transcriptome, small genome
			0.6–1	400 bp		
	Ion 520 Chip	4–6 M	0.6–1	200 bp		
			1.2–2	400 bp		
		3–4 M	0.5–1.5	600 bp		
	Ion 530 Chip	15–20 M	3–4	200 bp		
			6–8	400 bp		
		9–20 M	1. 5–4.5	600 bp		

Table 1 (*Continued*)

	Instrument	Maximum reads per run	Maximum output range (gigabases)	Read length	Sequencing run time (hours)	Main applications
	Ion 540 Chip	60–80 M	10–15	200 bp		
	Ion 550 Chip	100–130 M	20–25	200 bp		
BGI	MGISEQ-200	300 M	60	50-bp SE/PE, 100-bp SE/PE	48	Small genome sequencing, targeted sequencing
	BGISEQ-500	1,300 M	520	50-bp SE/PE, 100-bp SE/PE	45–213	Whole-genome sequencing, whole-exome sequencing, targeted sequencing, RNA sequencing
	MGISEQ-2000	1,800 M	1,080	50-bp SE/PE, 100-bp SE/PE, 150-bp PE, 300-bp SE	48	Whole-genome sequencing, whole-exome sequencing, transcriptome sequencing
	MGISEQ-T7[g]	–	6,000	150 bp PE	24	Whole-genome sequencing, deep exome sequencing, transcriptome sequencing, targeted panel sequencing
PacBio	Sequel[h]	0.5 M	20	10–30 kb on average	10–20	Whole-genome sequencing, RNA sequencing, Targeted sequencing, complex population sequencing, epigenetics
Nanopore	MinION Mk 1B	0.5 M	50	Hundreds to thousands of kb	Up to 72	Whole-genome sequencing, targeted sequencing, gene expression and RNA sequencing, epigenetic interrogation
	GridION X5	2.5 M	250		Up to 72	
	PromethION (Beta)	375 M	15,000		Up to 64	

Abbreviations: B, billion; bp, base pairs; kb, kilobase pairs; M, million; PCR, polymerase chain reaction; PE, paired-end reads; SE, single-end reads.
Note: all information was accessed from company Web sites on November 12, 2018: www.illumina.com; www.thermofisher.com; http://en.mgitech.cn/; www.pacb.com; www.nanoporetech.com.
[a]High-Output Kit specifications provided (Mid-Output Kit also available for 150 base pair paired-end reads).
[b]Diagnostic instrument version also available.
[c]MiSeq Reagent Kit v3 specifications provided (v2, Micro and Nano Kits also available).
[d]Single HiSeqX instrument specifications provided. This instrument is only available as clusters of 5 or 10 instruments.
[e]Specifications provided for a dual flow cell.
[f]Specifications provided refer to a chip rather than an instrument: Ion Gene Studio S5 instrument compatible with Ion 510 and Ion 540 chips, Ion Gene Studio S5 Plus and Prime instruments compatible with Ion 510 and Ion 550 chips.
[g]Reported release date in 2019.
[h]Specifications provided for one SMRT Cell. Up to 12 SMRT Cells can be run simultaneously.

The advantages of Illumina sequencers include high output (1.2–6,000 Gb), high accuracy, relatively low cost per base, diversity of applications, and diversity of library preparation configurations. Disadvantages include those inherent to short-read sequencing platforms, such as the inability to resolve repetitive regions of the genome, making some types of genetic variants challenging to identify, including repeat expansion disorders, and structural variants (SVs) if the breakpoints are in repetitive regions.

Ion Semiconductor Sequencing (ThermoFisher Scientific)
The ion semiconductor method essentially pairs sequencing by synthesis with a massively parallel array of microwells, each with a dedicated sensor. Library preparation first

isolates template DNA to the surface of microbeads, which are then amplified by emulsion PCR. The microbeads are then dispersed into individual microwells where sequencing is paired to a dedicated sensor that essentially acts as a highly sensitive pH meter. This nonoptical sequencing by synthesis method does not rely on labeled nucleotides but rather measures the electrical changes associated with hydrogen ion release as nucleotides are incorporated into the growing complementary strand.[17] Individual nucleotides are primed through the flow cell and, if incorporated, an electrical response is measured by the semiconductor sensor, and the relevant base call is registered to the sequence output. If a homopolymer of the same nucleotide is present in the template sequence, an electrical signal proportional to the number of nucleotides is registered and converted to a multibase call. However, this can be inconsistent and is generally limited to homopolymers of less than six nucleotides.

The current iteration of ion semiconductor sequencer, the Ion Torrent Gene Studio S5, has three instruments that use a range of chips producing read lengths of 200 to 600 bp with more than 99% accuracy and outputs up to 50 Gb (►**Table 1**). The main applications for these instruments include targeted resequencing, exome sequencing, transcriptome sequencing, small genome sequencing, liquid biopsy, and large cancer gene panels. Advantages of the system are the relatively low cost per base, owing to the absence of labeled nucleotides and the requirement for expensive optical imaging detection. Drawbacks to the Ion Torrent method are its relatively low throughput and challenges with resolving homopolymers, which can affect indel performance.

Nanoball Sequencing (Beijing Genomics Institute)
Complete Genomics (a subsidiary of Beijing Genomics Institute [BGI]) developed a genome sequencing platform that achieves efficient imaging and low reagent consumption with combinatorial probe-anchor ligation chemistry,[18] currently the only sequencing by ligation method. It independently assays each base from patterned arrays of self-assembling DNA nanoballs.[18] The BGISEQ-500 sequencer was initially announced by BGI in October 2015.[19] BGI now produces four versions of its unique sequencing by ligation platform (►**Table 1**), which are capable of multiple research and clinical applications.

A recent study demonstrated that the sequencing throughput and turnaround time, single-base quality, read quality, and variant calling were similar to Illumina HiSeq2500 data.[19] The main advantage of BGI nanoball sequencing is the reduced cost relative to Illumina sequencing, and recent updates have led to equivalent read lengths (i.e., 150-bp paired-end reads). Novel analysis methods, including Google's DeepVariant,[20] have also improved the variant calling accuracy. Limitations of the system are similar to other short-read technologies, namely the inability to resolve repeat sequences and some SVs.[18] In October 2018, BGI launched a new high throughput sequence platform called the MGISEQ-T7, which is capable of sequencing a whole human genome in less than 24 hours with 150-bp

paired-end reads.[21] The instrument is expected to rival Illumina's NovaSeq when it is released in mid-2019.

Third-Generation (Long-Read and Real-Time) Sequencing
The complexity of short-read second-generation sequencing data assembly and the inability to reliably resolve repeat sequences or large genomic rearrangements[22] were overcome by the third generation of sequencing, typified by long reads (> 1 kb to 2 Mb) and real-time sequencing.

Single-Molecule Real-Time Sequencing (Pacific Biosciences)
Single-molecule real-time (SMRT) sequencing from Pacific Biosciences (PacBio) was the first popular third-generation sequencer.[23] SMRT sequencing introduced the capability of real-time sequence acquisition for read lengths >1 kb using sequencing by synthesis and optical detection. The predominant library preparation method provides a template known as SMRTbell, a single-stranded circular DNA template formed by ligation of hairpin adaptors to both ends of a double-stranded DNA fragment.[22,24] This format only allows for relatively short reads with very high accuracy, whereas an alternative linear template approach enables longer read lengths but lower accuracy. The SMRTbell molecules are then individually dispersed into nanowells, known as zero-mode waveguides,[25] that have a single DNA polymerase molecule anchored to the bottom. The polymerase performs rolling amplification of the circularized SMRTbell template incorporating fluorescently tagged nucleotides. The fluorescent tag is cleaved by the polymerase as the nucleotide is incorporated, and the release of the fluorophore is recorded as a base call by real-time optical recognition before the next nucleotide is incorporated and its fluorophore is released. By continually amplifying the circular template, sequence from both the template and complementary strands of DNA can be captured multiple times to increase read depth and improve base call accuracy. As the sequencing rate is a function of nucleotide addition by the polymerase, PacBio sequencing can detect modified bases in the template as a variable delay in nucleotide incorporation.[26] Furthermore, the limiting factor in PacBio sequencing runs is the finite functional life of a polymerase molecule, resulting in single reads for ultra long templates and multiple contiguous reads of both strands for shorter templates.[27]

The chief advantage of PacBio sequencing is the long-read length that allows for greater certainty in read overlap and assembly, thus providing better resolution of repetitive regions and SVs.[22] Furthermore, PacBio can sequence DNA with unusual GC content, which is generally challenging to amplify and thus poorly resolved by short-read sequencing. The main shortcomings for earlier PacBio platforms were lower throughput, higher error rate (~13%), and greater cost per base.[22] However, with the release of the Sequel platform and advances in both flow cell design (increased from 150,000 to over 1 million zero-mode waveguides per chip) and sequencing chemistry, error rates as low as 3% can now be achieved. The main applications of PacBio sequencing include WGS, targeted sequencing, full-length mRNA sequencing, complex population sequencing, and detection of epigenetic modifications.[28]

At the time of writing, Pacific Biosciences reported the highest quality and most complete (2.89 Gb = 96% of the human genome) human genome to date, also showing unprecedented insight into the extent of structural variation across the genome.[29] In late 2018, Illumina announced that they had acquired Pacific Biosciences to add true long-read sequencing to their suite of next-generation sequencers.[30]

Nanopore Sequencing (Oxford Nanopore Technologies)

In 2014, Oxford Nanopore Technologies released nanopore sequencing in the form of the MinION, a handheld sequencer that uses a grid of membrane-embedded biological nanopores.[31] The membrane in which the nanopores sit provides separation of two ionic solutions allowing an electrical current to flow through the nanopores. Long DNA molecules are prepared by adding a hairpin adaptor to one end of the double-stranded molecule before a helicase, and motor protein attached to the template unwind and thread single-stranded DNA through the nanopore channel. The DNA can be ratcheted through the pore one base at a time, with the nucleotide bases inducing characteristic changes in the electrical current running through the nanopore that are translated to base calls.[31] As the sequencing process uses very few depletable reagents, the run can effectively continue until a satisfactory result is achieved.

Currently, Nanopore sequencing has much higher error rates (~15%) than short-read sequencing, but there is a great deal of research and development occurring to continually improve nanopore structure and function. One such way to improve read accuracy is termed $1D^2$ sequencing.[31] This method attaches a special adaptor to one end of the double-stranded DNA template molecule and allows the nanopore to sequence both the template and complementary strands contiguously, thereby providing higher sequence accuracy.

The MinION has the distinct advantage of being highly portable and capable of sequencing when plugged into a laptop.[31] It proved particularly advantageous in the field when there was an Ebola outbreak in 2015 and it was used to conduct the first sequencing experiments in space.[32,33] In addition to the MinION, Oxford Nanopore Technologies have released the GridION X5 and PromethION platforms (5 and 42 flow cell configurations, respectively) allowing vast throughput and scaling to many whole genomes per run (►**Table 1**). A single-use flow cell for the MinION, called a Flongle, is set be released and should enable smaller sequencing experiments to be performed, such as targeted sequencing, quality analysis, small genomes, and targeted diagnostic testing.

Nanopore sequencing has many potential benefits over other platforms, including label-free sequence determination of native DNA and RNA molecules without the need for amplification, and is capable of producing extremely long-read lengths. The read lengths from nanopore sequencing are rapidly increasing, with some groups reporting read lengths of > 2 Mb,[34] making it an excellent approach for de novo genome assembly.[31] Furthermore, as no amplification steps are necessary for nanopore sequencing, nucleotide modifi-

cations (such as methylation) are preserved on the template, with the capability to discriminate at least 25 DNA modifications and many more RNA modifications.[26,35,36]

Synthetic Long Reads (Illumina and 10X Genomics)

With respect to reduced quality and the cost associated with third-generation sequencing, 10X Genomics created the Chromium synthetic linked-read technology, an alternative short-read sequencing method that effectively produces long-read outputs. High molecular weight DNA templates (> 100 kb) are partitioned into gel beads, where they are fragmented into short-read libraries, with all library fragments from the same bead being tagged with the same unique molecular barcode. After traditional short-read Illumina sequencing, the unique barcodes can be read and used to phase the short reads back onto their original long template molecule, resulting in synthetic linked-reads. The linked-reads provide long-range information, which can improve read alignment, phasing of variants, and SV detection. This approach has the same accuracy and throughput as standard Illumina sequencing, but the library preparation is more expensive.

Data Output and Prebioinformatics Processing

Following a sequencing run, raw sequencing data must be computationally processed to a useable format, usually following established protocols, such as the Genome Analysis Toolkit (GATK) Best-Practices Pipeline.[37] Briefly, raw sequence reads are generally configured in FASTQ format, a text-based representation of every nucleotide base sequenced and an associated base quality score.[38] Sequencing reads are then aligned to a reference genome[39] using an aligning tool, which differ by manufacturer, DNA or RNA template, and read length. This produces a human-readable sequence alignment map (SAM) file and a binary version (BAM), which has a smaller file size.[40] These files enable visualization and interrogation of the sequence (read assembly and base sequence) using programs such as the Integrative Genomics Viewer.[41,42] Variant detection algorithms identify single nucleotide, short indel, or copy number variants (CNVs) and SVs based on reads that differ from the reference sequence. These variations are typically stored in a variant call format (VCF) file,[43] and are then annotated to determine the consequence, if any, of each genetic variant. A typical whole genome sequence will yield in the order of 5 million SNVs and 250,000 short indels. By disregarding common single-nucleotide polymorphisms and short insertions or deletions (indels), there are generally around 300,000 unique variants left to consider.[44] Downstream bioinformatics analysis of NGS data is covered elsewhere in this issue.

Complementary Methods

Several methods exist that can complement high-throughput next-generation DNA sequencing to provide greater confidence in DNA sequence findings and also to support pathological and biological inference obtained from genomic sequencing.

Optical Mapping (Bionano Genomics Irys and Saphyr)

Next-generation mapping (NGM), or optical mapping, allows for the interrogation of megabase (Mb) length DNA molecules that are beyond the detection range of single-base resolution NGS.[45] The Irys platform from Bionano Genomics enables very long DNA molecules to be optically mapped using a 7-bp recognition sequence that becomes fluorescently labeled. The fluorescent signal provides a genome-wide, sequence-specific pattern that can be directly visualized by the Irys instrument and aligned to a reference sequence to highlight regions of structural variation > 1.5 kb with high accuracy or for de novo genome assembly and phasing by hybrid sequencing (see the next section).[46–48] The Irys platform recently identified deletions between 45 and 250 kb, an insertion of 13 kb, and a 5.1-Mb inversion in a Duchenne muscular dystrophy cohort.[49] NGM using the Irys platform was also able to distinguish 85 large somatic structural rearrangements, 89% of which were undetectable using high-coverage, short-read NGS.[45] By inspecting short reads associated with large structural rearrangements identified by NGM, 94% were able to be confirmed. A new platform, named Saphyr,[50] was recently released by Bionano Genomics, promising a higher throughput, faster turnaround, and simplified workflow, which were major limitations of the Irys platform.

Hybrid Sequencing (Combining Second- and Third-Generation Sequencing)

The shortcomings of second- and third-generation sequencing (shorts reads and relatively higher error rates, respectively) can be compensated for by using them in combination. This is particularly beneficial for de novo genome assembly. This approach, known as hybrid sequencing, uses long-read sequencing or mapping to define a high-level genome assembly, including its repetitive regions that cannot be resolved by short-read sequencing, and then uses short reads to resolve the underlying sequence with high quality.[51,52] Conversely, genomic features unresolved by short-read sequencing can be confidently determined by long-read sequencing. This hybrid sequencing may present as a costly approach but can considerably reduce the amount of long-read data required to provide accurate sequence.[22] For instance, Illumina reads were recently used to improve the base call accuracy of ultra long nanopore reads covering the centromere of the Y chromosome.[53] In addition, the Vertebrate Genomes Project from the UK Genome 10k consortium[54] is now using a combination of technologies, including PacBio, 10X Genomics, Bionano Genomics, and HiC-proximity ligation data to generate high-quality sequence assemblies, combined with Illumina short-read sequencing data.

NanoString

The NanoString Technologies nCounter platform is a relatively new technology that can directly quantitate gene expression for a variety of different applications including molecular diagnostics.[55] The automated nCounter platform directly hybridizes fluorescent barcodes to RNA, allowing for non-amplified measurement of up to 800 gene targets within a sample using commercial or custom target gene panels.[56] The NanoString platform can achieve quantitative RNA expression levels with RNA as small as 100 nucleotides in length, thereby lending itself to the analysis of potentially degraded RNA from clinical samples, such as formalin-fixed paraffin-embedded tissue.[56] In 2018, NanoString announced the release of a quantitative spatial profiling application for protein and RNA in tissues, allowing for the accurate quantification of protein and gene expression spatially for deep characterization of a heterogeneous tissue sample, as well as an enzyme, amplification and library preparation-free hybridization and sequencing application that can sequence DNA and RNA concurrently.[57]

RNA-Seq

RNA-Seq provides a detailed view of the transcriptome, allowing for the detection of novel RNA transcript variation.[58,59] RNA-Seq has numerous potential advantages over gene expression microarrays, such as an increased dynamic range of expression, measurement of specific changes (including SNVs and indels), detection of transcript isoforms, splice variants, and chimeric gene fusions.[58] It has even been shown that PacBio and Nanopore sequencing can detect RNA nucleotide base modifications by monitoring reverse transcription in real time,[36,60] whereas other approaches require RNA to be converted to cDNA before sequencing, thereby losing any base modifications.

Potential applications for RNA-Seq include the diagnosis of infectious diseases, particularly RNA viruses.[58] As an example, nontargeted metagenomic RNA-Seq was used to directly detect influenza virus RNA in respiratory cases, with additional viral pathogens found in some of these cases.[58,61] In clinical studies, integrating DNA and RNA analysis has provided further evidence of altered regulation of mutated genes, with the potential to detect gene fusions and splicing variants, allowing for accurate prioritization of variants.[58] A recent study investigated the utility of RNA-Seq as a complementary diagnostic tool in 50 patients with genetically undiagnosed rare muscle disorders.[62] RNA-Seq was able to validate candidate splice-disrupting mutations and to identify splice-altering variants in exonic and deep intronic regions, resulting in an overall diagnosis rate of 35%.[62] RNA-Seq may have a broad range of clinical applications; however, test proficiency and validation need to be elucidated before it can be introduced into clinical practice.[58]

Chromatin Immunoprecipitation with Seq

Chromatin immunoprecipitation combined with NGS (ChIP-Seq) is an important tool for research into gene regulation.[63] ChIP-Seq permits mapping of in vivo DNA–protein interactions at high resolution and on a genome-wide scale.[63] Cross-linking of DNA-bound proteins is followed by immunoprecipitation, degradation of exposed DNA, unlinking of proteins, and recovery of DNA recognition sequences by NGS.[63] This has been widely applied to determine regulatory elements, including transcription factor binding sites, promoters, and enhancers. ChIP-Seq has a low cost and high

throughput and is highly amenable to short-read sequencing approaches, making data analysis the main bottleneck for this technique.[63]

Single-Cell RNA/DNA Sequencing

A promising application of NGS is the ability to consider nucleic acids from individual cells. There are several reasons why single-cell sequencing may be necessary, including (1) some cells that are important to analyze are found in lower numbers (e.g., human oocytes and circulating tumor cells), (2) every cell has a unique genome (e.g., immune cells), (3) the genomes of individual cells change over time, (4) there may be heterogeneity for individual cells in the same sample (e.g., primary cancer tissue), and (5) single-cell genomes may be altered by environment, lifestyle, disease, or therapeutic treatment.[64]

Usually, NGS requires nanogram amounts of DNA to construct a library for sequencing[65,66]; however, a single cell only contains 6 to 7 picograms of genomic DNA.[65] Therefore, an important step for single-cell sequencing is whole-genome amplification to generate an adequate amount of DNA for library construction.[65] Given the trace amounts of DNA from a single human cell, strict measures must be adhered to in order to avoid contamination from the environment or operators.[64]

There are several exciting potential applications of single-cell sequencing, including isolation and WGS of fetal cells from maternal blood as a step toward comprehensive, non-invasive, prenatal testing of genetic disorders.[67] Single-cell WGS has also been used to map and quantify karyotype heterogeneity in primary mouse lymphoma and human leukemia samples.[68] This revealed copy number heterogeneity, consistent with ongoing chromosomal instability that other platforms failed to detect.[68]

Methods under Development: The Next "Next-Generation Sequencing"

With the potential impact on rapid and personalized health care that NGS holds, there is a great deal of interest in developing new technologies and applications that will increase throughput, speed, and accuracy of sequencing while reducing the associated cost and error. A few of the many methods under development are presented here.

Nanopore Advances

Current nanopore technology uses protein-based nanopores that are structurally precise at an atomic level, easily modified, and capable of being engineered and produced on a large scale using bacteria. To increase the resolution of nanopore DNA sequencing, several improvements have been proposed, including integrated sensors, solid-state nanopores, and hybrid nanopores (protein nanopore in a solid-state membrane).[69] One such method under development uses the concept of electron tunneling currents, a quantum physical property of atoms. Briefly, a pair of electrodes separated on the nanometer scale across the nanopore aperture could resolve electrical differences in nucleotide

bases more effectively than the existing ionic current method.[70] This would not only improve the resolution of DNA sequence but could also speed up the process of base calling.[71] Theoretically, this could be further improved by solid-state nanopores engineered from synthetic materials, such as graphene or carbon nanotubes. Indeed, in a 2010 study by Garaj et al, a graphene nanopore showed electrical variations as a single molecule of DNA was translocated through the pore.[72] Some advances have been made, but a great deal more development is required to realize solid-state nanodevice DNA sequencing (reviewed by Heerema and Dekker[73]). Challenges associated with nanopore sequencing development and several other nanopore sequencing approaches are summarized by Wang et al.[74]

In addition to these research-based developments, Oxford Nanopore Technologies has indicated that an even smaller platform than the handheld, highly portable MinION will be available in the near future. The SmidgION is being developed to attach to a smartphone for analysis and will be complemented by rapid library preparation and simplified analysis pipelines.[50]

In Situ Nucleic Acid Sequencing

In a progression from single-cell RNA sequencing methods, in situ sequencing is performed intracellularly within intact tissues, thereby preserving the spatial context of gene expression within and between cell types.[75,76] In a recent methodological development, more than 1,000 genes were simultaneously resolved to show three-dimensional expression patterns between neurons in mouse brain on a cubic millimeter scale.[77] The hybrid technique used pairs of DNA probes binding to target RNA sequences, which were amplified to create DNA nanoballs. The DNA nanoballs were then bound to an in situ hydrogel matrix and the specific RNAs determined from fluorescent barcodes by means of in situ sequencing.[78] Furthermore, a truly innovative approach to mutation detection was recently reported where in situ sequencing analysis was achieved using a mobile phone camera paired with a small portable fluorescent and bright-field microscope unit.[79]

Microscopy-Based Sequencing

Interest in nucleic acid sequencing by electron microscopy has been around since the 1960s, but fell out of favor due to DNA damage from the high levels of radiation used and the inability to resolve nucleotides without heavy atom labeling. More recently, advanced low-energy electron microscopy and spectroscopy platforms have been used to discriminate individual nucleotides.[80,81] Further development may realize the original aim of providing high throughput label-free DNA sequencing capable of discriminating all manner of modified nucleotides at an atomic level.

Another theoretical approach proposed in 2005 is the use of tunneling currents to sequence DNA.[82,83] Aside from the theoretical use of tunneling currents in nanopores (mentioned previously), efforts to develop this method for microscopy-based DNA sequencing have been slow. However, over the past decade, Tanaka et al have managed to use scanning

electron tunneling microscopy to distinguish between the purine nucleotides adenine and guanine.[84,85] This heralds a significant advance for DNA sequencing by scanning electron tunneling microscopy, as a complete sequence could be inferred through the complementarity of pyrimidines with purines when sequencing both strands of a DNA molecule.

Clinical Applicability of Next-Generation Sequencing Methods

Comparison of Sequencing Approaches

Although the limited approach of Sanger sequencing is costly and inefficient for large-scale genomic sequencing, it remains the most widely accepted means to validate variants identified by NGS methodologies, particularly for clinically reportable findings.[86]

Targeted gene panels typically focus on sequencing genes that are known to be causative of or associated with a given phenotype. While they can be convenient and are commonly used, the gene lists are often too restrictive, can rapidly become out of date, and can miss unanticipated findings.[87,88] In this regard, more comprehensive sequencing approaches, such as WES and WGS, are clinically more appealing.

Whole-exome sequencing is performed by targeted enrichment of exonic regions, inclusive of canonical splice sites to 20 bp either side of an exon, followed by short-read sequencing. Although the exome is said to contain the vast majority (~85%) of disease-causing mutations,[89] WGS and third-generation long-read sequencing analyses are beginning to show the extent of noncoding and SVs that have the potential to cause disease. The chief advantage of WES is that it is cheaper than WGS and focuses the analysis on regions of the genome most likely to be implicated in disease.

Whole-genome sequencing covers the entire genome, including noncoding regions, that have previously been considered nonfunctional, although it is now apparent that these regions produce RNA molecules essential for cell development and biology.[89] In fact, genetic variation that controls protein regulation lies outside of the coding genome and plays a critical role in complex traits and diseases.[89,90] WGS is less susceptible to technical artifacts that relate to capture bias inherent in targeted gene panels and WES, with newer PCR-free library protocols showing reduced GC-bias and improved resolution of short repeats. Consistently, it has been estimated that with incomplete sequencing of GC-rich regions by WES, up to 400 disease-causing variants may be missed in an exome.[91] In comparison, WGS offers a more uniform coverage and greater breadth of coverage, as well as being better able to detect SNVs, indels, and SVs, including short and large CNVs.[89,92,93]

As a result of the apparent advantages that WGS has over WES (better breadth of coverage and better variant identification), WGS has been referred to as "a better exome."[91,94] This is reflected in the diagnostic rate, where WES generally achieves in the range of 25 to 35%, as opposed to WGS, which is in the range of 40 to 60%.[89] Although these rates are highly dependent on the level of cohort selection, it is suggestive of an inherently higher diagnostic yield for WGS compared with

WES.[89] Furthermore, a recent study emphasized the utility of WGS for detecting pathogenic variants in deep intronic regions,[95] and long-read WGS has disclosed the extent of SVs in the genome.[96]

Whole-genome sequencing data represents a repository of comprehensive genetic information for an individual that can be interrogated and re-interrogated in the future as confidence in analysis and disease association improves.[89] We suggest that WGS will likely be the preferred method for clinical genetic sequencing as technology improves and the cost falls over time. However presently, it is often the view that WES is a more cost-effective alternative.

Next-Generation Sequencing in Thrombosis and Hemostasis

Next-generation sequencing is having a large impact upon the hematology field, with numerous reports of NGS being used to elucidate the genetic etiology for inherited forms of bleeding, thrombosis, and platelet disorders. There is marked genetic heterogeneity with a rapidly expanding list of genes implicated in these disorders,[97] making NGS highly suitable for sequencing them. NGS promises to increase the diagnostic yield, reduce time to diagnosis, provide insights into genotype–phenotype correlations, and facilitate the discovery of novel genes.[98] In fact, the diagnostic yield may be above 90% when the phenotype corresponds to a known disorder but falls to 10% when the phenotype is consistent with a novel disorder.[97]

As an example, one study developed a targeted sequencing platform covering 63 genes (the ThromboGenomics platform) to identify previously detected variants in 100% of samples tested ($n = 159$), as well as 56 (91.8%) of 61 cases with clinical and laboratory phenotypes suggesting a particular molecular etiology.[99] In contrast, in a group classified as "affected with uncertain cause," a molecular diagnosis was reached in only 8 (10.5%) of 76 cases.[99]

Another recent study focused on the detection of SVs as a cause of hemoglobinopathies.[100] The authors developed an automated analytical pipeline (Snappy) to detect the first duplication of the entire β globin cluster. The analysis method incorporated the assessment of sequence coverage and detection of split and discordant reads in an approach that could be applied to a routine diagnostic pathway.[100]

Finally, an ongoing study that shows great promise is the My Life Our Future (MLOF) program, which was initiated in 2012 to provide genetic analysis and expand hemophilia research by creating a research repository.[101] DNA from 5,141 MLOF subjects has undergone WGS through the National Heart, Lung, and Blood Institute's Trans-Omics for Precision Medicine program of the National Institutes of Health and will provide a large, comprehensive database that will serve as a valuable resource to study this disease in the future.[101]

Conclusion

Sequencing technology is evolving at a rapid pace, and there is now a vast array of options available for different

applications of massively parallel sequencing. As the analysis and interpretation of sequence variants improve and catch up to data generation, the diagnosis of previously unresolved disorders will become commonplace. We believe that the benefit of these new technologies has not yet been fully exploited and will provide a means to offer precision and personalized medical care in the future.

Conflicts of Interest
The authors do not have any conflicts of interest to declare.

References

1 Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: https://www.genome.gov/sequencingcostsdata. Accessed December 12, 2018

2 Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953;171(4356): 737–738

3 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74(12): 5463–5467

4 Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11(01):31–46

5 Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010;19(R2):R227–R240

6 Schloss JA. How to get genomes at one ten-thousandth the cost. Nat Biotechnol 2008;26(10):1113–1115

7 Muzzey D, Evans EA, Lieber C. Understanding the basics of NGS: from mechanism to variant calling. Curr Genet Med Rep 2015; 3(04):158–165

8 van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014;30(09): 418–426

9 Zappala Z, Montgomery SB. Non-coding loss-of-function variation in human genomes. Hum Hered 2016;81(02):78–87

10 Deveson IW, Hardwick SA, Mercer TR, Mattick JS. The dimensions, dynamics, and relevance of the mammalian noncoding transcriptome. Trends Genet 2017;33(07):464–478

11 Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in clinical genomics. Exp Mol Med 2018;50(08):97

12 Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010;11(10):685–696

13 Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17(06):333–351

14 Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012;2012:251364

15 Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008;24(03):133–141

16 Raine A, Liljedahl U, Nordlund J. Data quality of whole genome bisulfite sequencing on Illumina platforms. PLoS One 2018; 13(04):e0195972

17 Merriman B, Rothberg JM, Rothberg JM; Ion Torrent R&D Team. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 2012;33(23):3397–3417

18 Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010;327(5961):78–81

19 Huang J, Liang X, Xuan Y, et al. A reference human genome dataset of the BGISEQ-500 sequencer. Gigascience 2017; 6(05):1–9

20 Poplin R, Chang PC, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol 2018;36(10):983–987

21 Heger M. BGI Launches New Sequencer as Customers Report Data From Earlier Instruments. 2018. Available at: https://www.genomeweb.com/sequencing/bgi-launches-new-sequencer-customers-report-data-earlier-instruments#.W-jL9-JxWUk. Accessed November 2, 2018

22 Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015;13(05):278–289

23 Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res 2018; 46(05):2159–2168

24 Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 2010;38(15):e159

25 Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science 2003;299(5607):682–686

26 Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 2010;7(06):461–465

27 Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323(5910):133–138

28 Nakano K, Shiroma A, Shimoji M, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell 2017;30(03):149–161

29 Pacific Biosciences Releases Highest-Quality, Most Contiguous Individual Human Genome Assembly to Date. 2018. Available at: https://www.pacb.com/press_releases/pacific-biosciences-releases-highest-quality-most-contiguous-individual-human-genome-assembly-to-date/. Accessed October 10, 2018

30 Illumina to Acquire Pacific Biosciences for $1.2 Billion. 2018. Available at: https://www.genomeweb.com/sequencing/illumina-acquire-pacific-biosciences-12-billion#.W9_EWuJxWUk. Accessed November 2, 2018

31 Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 2016;14(05):265–279

32 Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature 2016;530(7589): 228–232

33 Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA sequencing and genome assembly on the international space station. Sci Rep 2017;7(01):18022

34 Payne A, Holmes N, Rakyan V, Loose M. Whale watching with BulkVis: a graphical viewer for Oxford Nanopore bulk fast5 files. bioRxiv 2018 (e-pub ahead of print). doi:10.1101/312256

35 Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. RNA 2017;23(12):1754–1769

36 Garalde DR, Snell EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 2018; 15(03):201–206

37 Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43: 1–33

38 Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 2010;38(06):1767–1771

39 Auton A, Brooks LD, Durbin RM, et al; 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526(7571):68–74

40 Sequence Alignment/Map Format Specification. 2018. Available at: https://samtools.github.io/hts-specs/SAMv1.pdf. Accessed October 10, 2018

41 Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(01):24–26

42 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14(02):178–192

43 The Variant Call Format (VCF) Version 4.2 Specification. 2018. Available at: https://samtools.github.io/hts-specs/VCFv4.2.pdf. Accessed November 7, 2018

44 Pickrell WO, Rees MI, Chung SK. Next generation sequencing methodologies–an overview. Adv Protein Chem Struct Biol 2012; 89:1–26

45 Jaratlerdsiri W, Chan EKF, Petersen DC, et al. Next generation mapping reveals novel large genomic rearrangements in prostate cancer. Oncotarget 2017;8(14):23588–23602

46 Mak AC, Lai YY, Lam ET, et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics 2016;202(01):351–362

47 Mostovoy Y, Levy-Sakin M, Lam J, et al. A hybrid approach for de novo human genome sequence assembly and phasing. Nat Methods 2016;13(07):587–590

48 Hastie AR, Lam ET, Pang AW, et al. Rapid automated large structural variation detection in a diploid genome by NanoChannel based next-generation mapping. bioRxiv 2017 (e-pub ahead of print). doi:10.1101/102764

49 Barseghyan H, Tang W, Wang RT, et al. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. Genome Med 2017;9(01):90

50 Oxford Nanopore Technologies. SmidgION. Available at: https://nanoporetech.com/products/smidgion. Accessed November 7, 2018

51 Miller JR, Zhou P, Mudge J, et al. Hybrid assembly with long and short reads improves discovery of gene family expansions. BMC Genomics 2017;18(01):541

52 Minei R, Hoshina R, Ogura A. De novo assembly of middle-sized genome using MinION and Illumina sequencers. BMC Genomics 2018;19(01):700

53 Jain M, Olsen HE, Turner DJ, et al. Linear assembly of a human centromere on the Y chromosome. Nat Biotechnol 2018; 36(04):321–323

54 Koepfli KP, Paten B, O'Brien SJ; Genome 10K Community of Scientists. The Genome 10K Project: a way forward. Annu Rev Anim Biosci 2015;3:57–111

55 Tsang HF, Xue VW, Koh SP, Chiu YM, Ng LP, Wong SC. NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics. Expert Rev Mol Diagn 2017;17(01):95–103

56 Veldman-Jones MH, Brant R, Rooney C, et al. Evaluating robustness and sensitivity of the NanoString Technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. Cancer Res 2015;75(13):2587–2593

57 NanoString. NanoString Showcases Groundbreaking New Platforms at 2018 Advances in Genome Biology and Technology (AGBT) Conference. 2018. Available at: https://globenewswire.com/news-release/2018/02/12/1338884/0/en/NanoString-Showcases-Groundbreaking-New-Platforms-at-2018-Advances-in-Genome-Biology-and-Technology-AGBT-Conference.html. Accessed November 12, 2018

58 Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet 2016;17(05): 257–271

59 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5(07):621–628

60 Vilfan ID, Tsai YC, Clark TA, et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. J Nanobiotechnology 2013;11:8

61 Fischer N, Indenbirken D, Meyer T, et al. Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples. J Clin Microbiol 2015;53(07):2238–2250

62 Cummings BB, Marshall JL, Tukiainen T, et al; Genotype-Tissue Expression Consortium. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med 2017;9(386):eaal5209

63 Ambrosini G, Dreos R, Kumar S, Bucher P. The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. BMC Genomics 2016;17(01):938

64 Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-cell whole-genome amplification and sequencing: methodology and applications. Annu Rev Genomics Hum Genet 2015;16:79–102

65 Chen D, Zhen H, Qiu Y, et al. Comparison of single cell sequencing data between two whole genome amplification methods on two sequencing platforms. Sci Rep 2018;8(01):4963

66 Sanchez-Cespedes M, Cairns P, Jen J, Sidransky D. Degenerate oligonucleotide-primed PCR (DOP-PCR): evaluation of its reliability for screening of genetic alterations in neoplasia. Biotechniques 1998;25(06):1036–1038

67 Chen F, Liu P, Gu Y, et al. Isolation and whole genome sequencing of fetal cells from maternal blood towards the ultimate non-invasive prenatal testing. Prenat Diagn 2017;37(13):1311–1321

68 Bakker B, Taudt A, Belderbos ME, et al. Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. Genome Biol 2016;17(01):115

69 Goto Y, Yanagi I, Matsui K, Yokoi T, Takeda K. Integrated solid-state nanopore platform for nanopore fabrication via dielectric breakdown, DNA-speed deceleration and noise reduction. Sci Rep 2016;6:31324

70 Fanget A, Traversi F, Khlybov S, et al. Nanopore integrated nanogaps for DNA detection. Nano Lett 2014;14(01):244–249

71 Morikawa T, Yokota K, Tanimoto S, Tsutsui M, Taniguchi M. Detecting single-nucleotides by tunneling current measurements at sub-MHz temporal resolution. Sensors (Basel) 2017; 17(04):E885

72 Garaj S, Hubbard W, Reina A, Kong J, Branton D, Golovchenko JA. Graphene as a subnanometre trans-electrode membrane. Nature 2010;467(7312):190–193

73 Heerema SJ, Dekker C. Graphene nanodevices for DNA sequencing. Nat Nanotechnol 2016;11(02):127–136

74 Wang Y, Yang Q, Wang Z. The evolution of nanopore sequencing. Front Genet 2015;5:449

75 Ke R, Mignardi M, Pacureanu A, et al. In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods 2013; 10(09):857–860

76 Lee JH, Daugharthy ER, Scheiman J, et al. Highly multiplexed subcellular RNA sequencing in situ. Science 2014;343(6177): 1360–1363

77 Wang X, Allen WE, Wright MA, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science 2018;361(6400):5691

78 Chen X, Sun YC, Church GM, Lee JH, Zador AM. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. Nucleic Acids Res 2018;46(04):e22

79 Kühnemund M, Wei Q, Darai E, et al. Targeted DNA sequencing and in situ mutation analysis using mobile phone microscopy. Nat Commun 2017;8:13913

80 Mankos M, Shadman K, Persson HH, N'Diaye AT, Schmid AK, Davis RW. A novel low energy electron microscope for DNA sequencing and surface analysis. Ultramicroscopy 2014;145:36–49

81 Mankos M, Persson HH, N'Diaye AT, Shadman K, Schmid AK, Davis RW. Nucleotide-specific contrast for DNA sequencing by electron spectroscopy. PLoS One 2016;11(05):e0154707

82 Quantum steps to better sequencing. Nat Nanotechnol 2010; 5(12):823

83 Di Ventra M, Taniguchi M. Decoding DNA, RNA and peptides with quantum tunnelling. Nat Nanotechnol 2016;11(02):117–126

84 Tanaka H, Kawai T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. Nat Nanotechnol 2009; 4(08):518–522

85 Tanaka H, Taniguchi M. Sequencing of adenine in DNA by scanning tunneling microscopy. Jpn J Appl Phys 2017;56(8)

86 Alekseyev YO, Fazeli R, Yang S, et al. A next-generation sequencing primer—how does it work and what can it do? Acad Pathol 2018;5:2374289518766521

87 Caspar SM, Dubacher N, Kopps AM, Meienberg J, Henggeler C, Matyas G. Clinical sequencing: from raw data to diagnosis with lifetime value. Clin Genet 2018;93(03):508–519

88 Kumar KR, Wali GM, Kamate M, et al. Defining the genetic basis of early onset hereditary spastic paraplegia using whole genome sequencing. Neurogenetics 2016;17(04):265–270

89 Mattick JS, Dinger M, Schonrock N, Cowley M. Whole genome sequencing provides better diagnostic yield and future value than whole exome sequencing. Med J Aust 2018;209(05):197–199

90 Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet 2011;43(06):513–518

91 Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet 2016;135(03):359–362

92 Mallawaarachchi AC, Hort Y, Cowley MJ, et al. Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. Eur J Hum Genet 2016;24(11):1584–1590

93 Minoche AE, Horvat C, Johnson R, et al. Genome sequencing as a first-line genetic test in familial dilated cardiomyopathy. Genet Med 2019;21(03):650–662

94 Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A 2015;112(17): 5473–5478

95 Bagnall RD, Ingles J, Dinger ME, et al. Whole genome sequencing improves outcomes of genetic testing in patients with hypertrophic cardiomyopathy. J Am Coll Cardiol 2018;72(04):419–429

96 Chaisson MJ, Huddleston J, Dennis MY, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature 2015;517(7536):608–611

97 Freson K, Turro E. High-throughput sequencing approaches for diagnosing hereditary bleeding and platelet disorders. J Thromb Haemost 2017;15(07):1262–1272

98 Heremans J, Freson K. High-throughput sequencing for diagnosing platelet disorders: lessons learned from exploring the causes of bleeding disorders. Int J Lab Hematol 2018;40(Suppl 1):89–96

99 Simeoni I, Stephens JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. Blood 2016;127(23):2791–2803

100 Clark BE, Shooter C, Smith F, Brawand D, Thein SL. Next-generation sequencing as a tool for breakpoint analysis in rearrangements of the globin gene clusters. Int J Lab Hematol 2017; 39(Suppl 1):111–120

101 Konkle BA, Johnsen JM, Wheeler M, Watson C, Skinner M, Pierce GF; My Life Our Future programme. Genotypes, phenotypes and whole genome sequence: approaches from the My Life Our Future haemophilia project. Haemophilia 2018;24(Suppl 6):87–94