



Marius Eisele and Melanie Kappelmann-Fenzl

Contents

4.1 Introduction	48
4.2 Illumina	48
4.3 Ion Torrent	51
4.4 Pacific Bioscience	51
4.5 Oxford Nanopore	52
4.6 NGS Technologies: An Overview	54
References	58

What You Will Learn in This Chapter

There are a number of different companies that have developed and improved the NGS technology immensely in the last 10 years. In this chapter an overview of the most common technologies and their basic properties shall be given. Furthermore, it will be shown which technologies can be used for specific scientific or clinical questions and how they differ in their chemistry and output.

M. Eisele
Deggendorf Institute of Technology, Deggendorf, Germany

M. Kappelmann-Fenzl (✉)
Deggendorf Institute of Technology, Deggendorf, Germany

Institute of Biochemistry (Emil-Fischer Center), Friedrich–Alexander University Erlangen–Nürnberg, Erlangen, Germany
e-mail: melanie.kappelmann-fenzl@th-deg.de

4.1 Introduction

Since the completion of the human genome project in 2003, amazing progress has been made in sequencing technologies [1]. The cost per megabase decreased and the number and diversity of sequenced genomes increased dramatically. Some approaches maximize the number of bases sequenced in the least amount of time (short-read sequencing), generating big data enabling a better understanding of complex phenotypes and disease. Alternatively, other approaches now aim to sequence longer contiguous pieces of DNA (long-read sequencing), which are essential for resolving structurally complex regions. These and other strategies are providing researchers and clinicians a variety of tools to investigate genomes, exomes, transcriptomes, epigenomes in greater depth, leading to an enhanced understanding of how biological sequence variants lead to phenotypic alterations and thus the development of various disease patterns [2].

The yearly updates of the Travis Glenn's Field Guide to Next Generation DNA Sequencer [3] are a good summary of the state of instrumentation (<http://www.molecularecologist.com/next-gen-fieldguide-2016/>).

4.2 Illumina

The Illumina sequencing technologies support a wide range of genetic analysis research applications, such as:

- *Whole-Genome Sequencing*: A comprehensive method for analyzing entire genomes.
- *Genotyping*: Studying variation in genetic sequences.
- *Gene Expression and Transcriptome Profiling*: Analyzing which genes and transcripts are expressed in a given sample.
- *Epigenetics*: Studying heritable changes in gene regulation that occur without a change in the DNA sequence.

Therefore, Illumina developed the Sequencing by Synthesis (SBS) Technology and BeadArray Microarray Technology. In this textbook we will focus on SBS.

The NGS massively parallel sequencing technology has revolutionized the biological sciences. With its ultra-high throughput, scalability, and speed, NGS enables researchers to perform a wide variety of applications and study biological systems at a level never before possible.

Today's complex genomic research questions demand a depth of information beyond the capacity of traditional DNA sequencing technologies. NGS has filled that gap and

becomes an everyday research tool to address these questions [4]. Illumina NGS workflows include the following basic steps:

- Library Preparation (see Chap. 3)

Libraries for NGS applications can be generated for diverse methods. Which library preparation workflow to choose depends on your scientific or clinical question and its relation to the genome, transcriptome, or epigenome of any organism. An overview of the different Illumina Library Preparation Kits can be found at <https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits.html> (see Chap. 3).

- Cluster Generation

Sequencing templates are immobilized on a flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template and low non-specific binding of fluorescently labeled nucleotides. Solid-phase amplification creates up to 1000 identical copies of each single template molecule in close proximity.

- Sequencing

Illumina sequencing technology is also known as sequencing by synthesis (SBS) technology. Four fluorescently labeled nucleotides are used to sequence the tens of millions of clusters on the flow cell surface in parallel. During each sequencing cycle, a single labeled deoxynucleoside triphosphate (dNTP) is added to the nucleic acid chain and the nucleotide label serves as a reversible terminator for polymerization. After removing the fluorescence label of previously attached dNTP another labeled dNTP is added during a new sequencing cycle. Base calls are made directly from signal intensity measurements during each cycle.

- Data Analysis

The NextSeq 550/2000, NextSeq 2000, and NovaSeq 6000 Sequencing Systems generate raw data files in binary base call (BCL) format, requiring conversion to FASTQ format for use with user-developed or third-party data analysis tools. Illumina offers bcl2fastq Conversion Software to convert BCL files. bcl2fastq is an included, standalone conversion software that demultiplexes data and converts BCL files to standard FASTQ files, which are the starting format for data analysis.

As already described in Chap. 3, the library, which was prepared by random fragmentation of the DNA or cDNA (in terms of RNA-Seq) sample, followed by 5' and 3' adapter ligation, PCR amplification, and gel purification (see Fig. 4.1a and Chap. 3). For cluster generation, the library is loaded onto a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing (Fig. 4.1b). The sequencing by synthesis (SBS) technology uses a proprietary reversible terminator-based method that

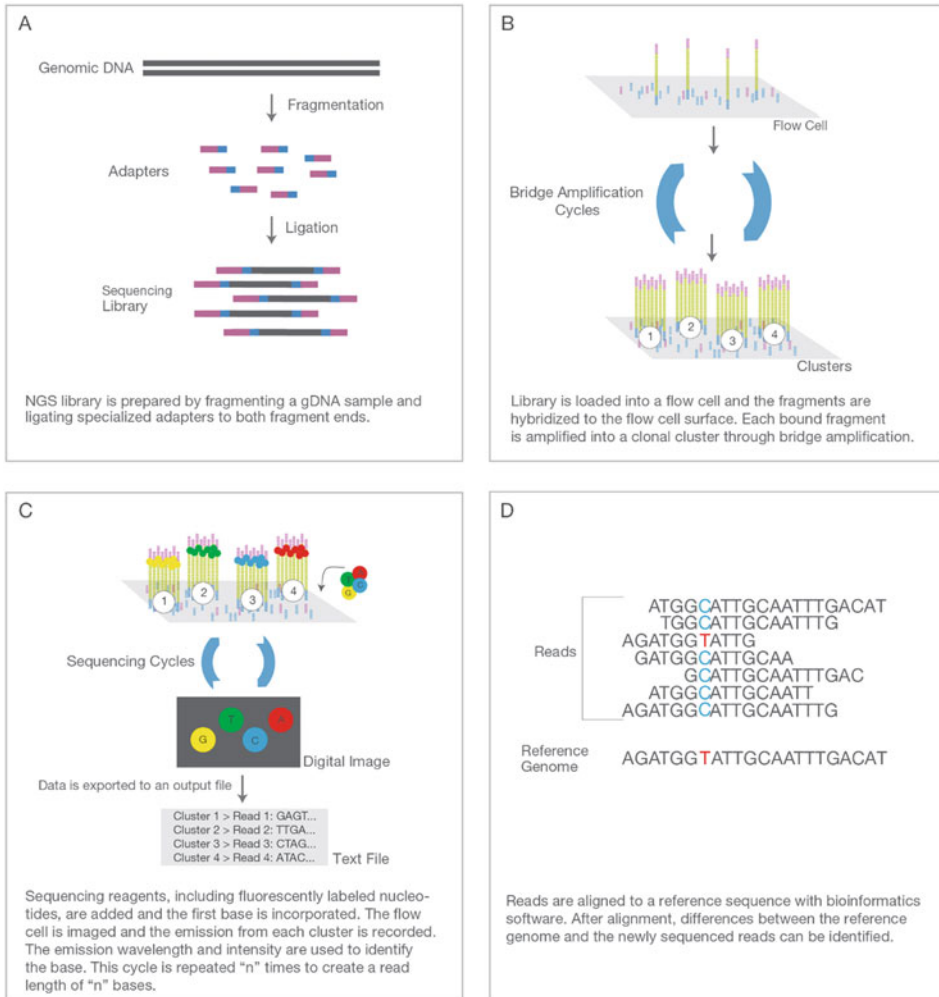


Fig. 4.1 Next Generation Sequencing Chemistry Overview—Illumina NGS includes four steps: (a) library preparation, (b) cluster generation, (c) sequencing, and (d) alignment and data analysis. (source: www.illumina.com)

detects single bases as they are incorporated into DNA template strands. As all four reversible terminator-bound dNTPs are present as single, separate molecules during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates. The result is highly accurate base-by-base sequencing that virtually eliminates sequence context-specific errors, even within repetitive sequence regions and homopolymers (see Fig. 4.1c). During data analysis and alignment, the newly identified

sequence reads are aligned to a reference genome. Following alignment, many variations of analysis are possible, such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for RNA methods, phylogenetic or metagenomic analysis, and more. A graphical overview of the NGS chemistry is depicted in Fig. 4.1.

4.3 Ion Torrent

Unlike Illumina Ion Torrent semiconductor sequencing from Thermo Fisher Scientific does not make use of optical signals. Instead, they exploit the fact that addition of a dNTP to a DNA polymer releases an H^+ ion.

As in other kinds of NGS, the input DNA or RNA is fragmented to approximately 200bp, adapters are added, and one molecule is placed onto a bead. The molecules are amplified on the bead by emulsion PCR resulting in millions of different beads with millions of different fragments. These beads then flow across the semiconductor chip depositing each bead into a single well. Next the slide is flooded with a single species of dNTP (one NTP at a time), along with buffers and polymerase. The pH is detected, as each H^+ ion released will decrease the pH. The changes in pH allow to determine if that base, and how many thereof, was added to the sequence read. The dNTPs are washed away and the process is repeated cycling through the different dNTP species. The pH change (if any) is utilized to determine how many bases (if any) were added with each cycle.

If a nucleotide, for example, a C, is added to a DNA template and is then incorporated into a strand of DNA, a hydrogen ion will be released. The charge from that ion will change the pH of the solution in the well, which can be detected by a specific ion sensor. This process happens simultaneously in millions of wells, that is why this technology is often described as massively parallel sequencing.

The Ion Torrent NGS instruments Genexus, Ion GeneStudio S5, ION PGM Dx, Ion Chef, and Ion OneTouch2 are essentially the world's smallest solid-state pH meters, calling the base, going directly from chemical information to digital information.

4.4 Pacific Bioscience

Single-molecule, real-time (SMRT) sequencing developed by Pacific BioSciences (PacBio) offers longer read lengths than the second-generation sequencing (SGS) technologies, making it well-suited for unsolved problems in genome, transcriptome, and epigenetics research [5].

Introducing the PacBio Sequel II system powered by SMRT sequencing technology the first step is to isolate DNA or RNA from any sample type. Next a SMRTbell library is created by ligating hairpin adapters to double stranded DNA creating a circular template. Primer and polymerase are added to the library that is placed on the instrument for sequencing. The smart cell contains millions of small, tiny wells called zero-mode waveguides (ZMWs). A single molecule of DNA is immobilized in a ZMW sequencing

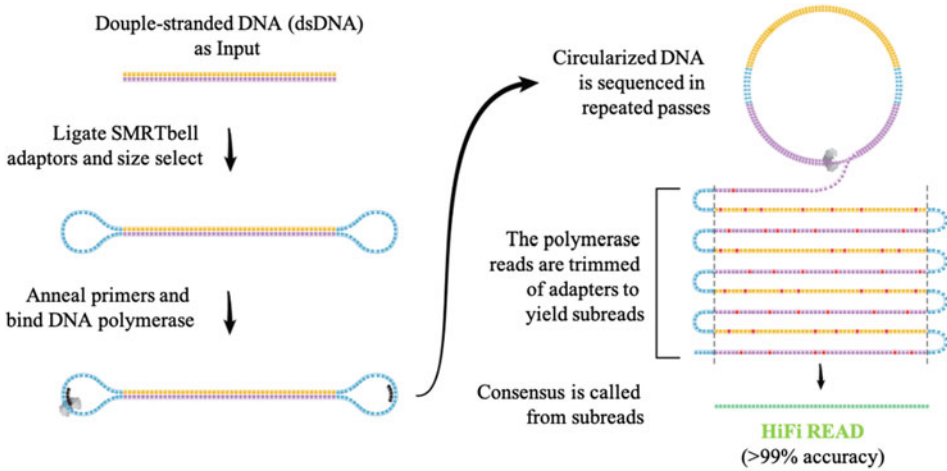


Fig. 4.2 Using the circular consensus sequencing (CCS) mode for HiFi READ production to provide base-level resolution with >99% single-molecule read accuracy for the detection of all variant types from single nucleotide to structural variants (source: modified according to <https://www.pacb.com>)

unit, which provides the smallest available volume for light detection and as the polymerase incorporates fluorescently labeled deoxyribonucleoside triphosphates (dNTPs) light is emitted. The order of their enzymatic incorporation into a growing DNA strand is detected via ZMW nanostructure arrays, which allow the simultaneous detection of thousands of single-molecule sequencing reactions. The replication processes in all ZMWs of a SMRT cell are recorded by a “movie” of light pulses, and the pulses corresponding to each ZMW can be interpreted to be a sequence of bases. With this approach nucleotide incorporation is measured in real time. With the Sequel II system you can optimize your results with two sequencing modes. You can use the circular consensus sequencing (CCS) mode to produce highly accurate long reads, known as HiFi reads (Fig. 4.2), or use the continuous long-read sequencing (CLR) mode to generate the longest possible reads (Fig. 4.3). The average read length from the PacBio instrument is approximately 2 kb, and some reads may be over 20 kb. Longer reads are especially useful for de novo assemblies of novel genomes as they can span many more repeats and bases.

4.5 Oxford Nanopore

In essence, Oxford Nanopore is a real-time, high-throughput technology and is specialized on long-read and single-molecule sequencing. Oxford Nanopore technology consists of millions of nanoscale pores spanned across an impermeable thin membrane, allowing massively parallel sequencing. The membrane separates two chambers, both contain an electrolyte and a single connection to each other via a single nanopore. The applied voltage

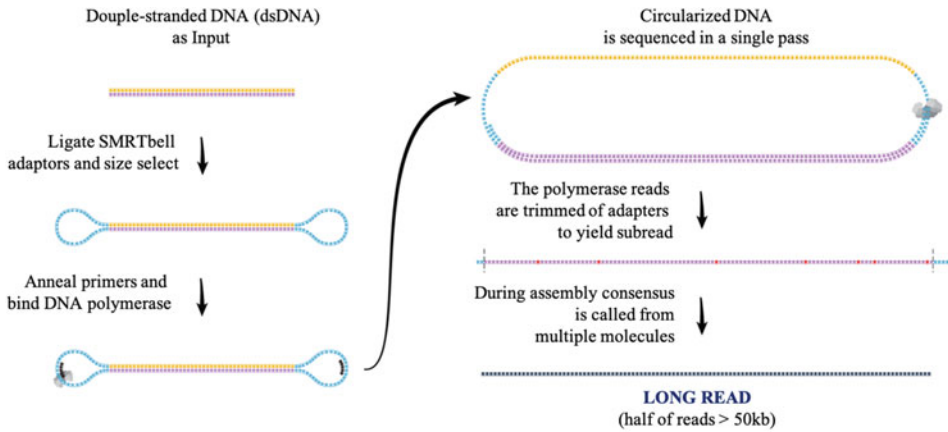


Fig. 4.3 Using the continuous long-read (CLR) sequencing mode for sequence read lengths in the tens of kilobases to enable high-quality assembly of even the most complex genomes. With SMRT sequencing you can expect half the data in reads >50 kb and the longest reads up to 175 kb (source: modified according to <https://www.pacb.com>)

by two electrodes generates an ion flow from one chamber, through the pore and into the other chamber.

This way, ions and charged biomolecules like the nucleic acid molecules with their negative charge can be driven through the pore. The ions act as a motor, allowing the molecules to be passed through the channel. Consequently, structural features, such as the bases or the epigenetic modification of the sequences, can be identified by tracing the ionic current, which is partially blocked by the molecule. Compared to other sequencing technologies, Oxford Nanopore with its fascinating simple biophysical approach has resulted in overwhelming academic, industrial, and national interest (Fig. 4.4, [6]).

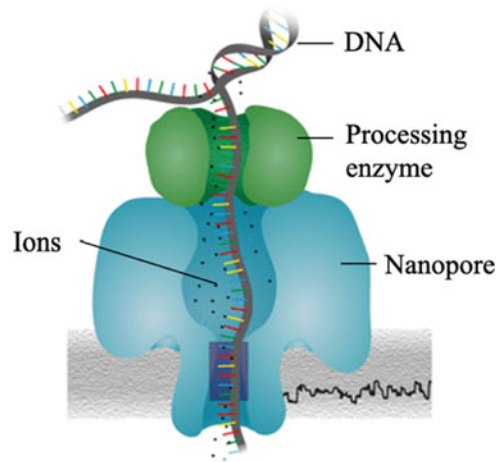
Historically, the pioneer technology giving rise to the Oxford Nanopore was invented by Wallace H. Coulter in the late 1940s. Coulter's technology was using essentially the same basic chemo-physical principle as Oxford Nanopore, but was used for counting and sizing blood cells. An automated version of Coulter's counters is still used in hospitals today. However, the true reincarnation of the Coulter's counters was in the 1990s, when the pore was not of millimeter but of nanometer dimensions, allowing the analysis of ions and biomolecules instead of whole cells [7].

Properties that an analyte should have: Every analyte molecule consists of multiple ions that allow it to pass the pore. Furthermore, the pore has to be wide enough (i.e., around 2 nm) and must permit the transport of ions. Ultimately, the flow of ions across the pore should be able to report on subtle differences between the analytes.

There are two types of pores that are currently being used: protein and solid-state channels.

Examples of protein channels are toxin α -hemolysin, which is secreted by *Staphylococcus aureus*, and MspA from *Mycobacterium smegmatis*. A promising approach for solid-state channels is the use of TEM (Transmission electron microscopy), combined with a

Fig. 4.4 Graphic representation of DNA sequencing using a MinION. A processive enzyme (green) ratchets DNA into the pore (blue), causing a change in ionic current (ions are shown as black dots) that is determined by the 6-mer in the central channel (purple box). The current is recorded over time (black trace, bottom right). (modified according to Muller et al. [6])



single-layer graphene membrane, however, up to this point (2020) the protein channels are superior to the solid-state channels [8].

As mentioned earlier, biomolecules that pass through the pore generate the signal by partially blocking the flow of ions, which can then be translated into the sequence and epigenetic modifications. Nevertheless, ions lining up at the membrane, together with the counterions on the opposite site of the membrane, also contribute to the signal, generating noise. These noise fluctuations increase with bandwidths, which limits the time resolution in experiments. Apart from shorter measurement times, a common way to compensate the noise is achieved by using analog or digital low-pass filters. Still, the generation of noise, introducing error-prone data, might be the biggest struggle Oxford Nanopore Technology has yet to overcome.

Big advantages of this sequencing technology compared to others on the market are its portability and price tag. Oxford Nanopores' smallest device, the MinION is controlled and powered by a USB cable and is just slightly bigger than a regular USB stick. Depending on the experiment (DNA or RNA), Oxford Nanopore devices do not need an amplification step (PCR) prior to the sequencing. Theoretically, the only limitation in sequencing length is the time and therefore the induced noise. So far, the maximum of usable read length is around 100 kilobases [9]. However, longer reads result in less accurate data [10].

4.6 NGS Technologies: An Overview

A more detailed overview of the major NGS platforms and their general properties [10] are listed in the Table in the Appendix section (Table 13.1: Major NGS platforms and their general properties.)

Take Home Message

- The term NGS is used to summarize second- and third-generation sequencing methods.
- Basically, a distinction can be made between first-generation sequencing (Sanger sequencing), second-generation sequencing (massively parallel sequencing), and third-generation sequencing (single-molecule sequencing).
- With regard to sequencing technologies, one also differentiates between short- and long-read sequencing.
- The different NGS technologies entail different preprocessing and analysis steps and are applied depending on the scientific or clinical question to the resulting data.
- Each NGS technology can be characterized by its input template, read length, error rate, sequencing scheme, visualization method, sequencing principle, and amount of data output.

Further Reading

- See Table 4.1: Sequencing technologies of some companies and their products.

Table 4.1 Sequencing technologies of some companies and their products

	Products	Technology	Further Information
Illumina	iSeq 100, MiniSeq, MiSeq, NextSeq 550, NextSeq 2000, NovaSeq 6000	SBS, short-read	https://emea.illumina.com/
10xGenomics	Chromium Controller	Linked reads	https://www.10xgenomics.com/
Thermo Fisher	Ion GeneStudio S5 series Ion PGM Dx System	SBS, short-read	https://www.thermofisher.com/
Pacific Bioscience	PacBio Sequel Systems	Long-read sequencing (Up to 160Gb), Single-Molecule, Real-Time (SMRT) sequencing	https://www.pacb.com/
Qiagen	GeneReaderPlatform	End-to-end NGS workflow, clinical testing	https://www.qiagen.com/
Oxford Nanopore	MinION (Up to 30Gb) GridION X5 (Up to 150Gb) PromethION (Up to 9,600 Gb)	Long-read (Up to 2Mb), real-time nanopore sequencing	https://nanoporetech.com/

Review Questions

- Which of the following statements regarding the quantity of template for a sequencing reaction is correct?
 - Excess template reduces the length of a read.
 - Too little template will result in very little readable sequences.
 - Excess template reduces the quality of a read.
 - All of the above.
- What will heterozygous single nucleotide substitution look like on your chromatogram?
 - Two peaks of equal height at the same position.
 - One peak twice the height of those around it.
 - Two peaks in the same position, one twice the height of the other.
 - Three peaks of equal height at the same position.
- Which of the following is important for preparing templates for Next Generation Sequencing?
 - Isolating DNA from tissue.
 - Breaking DNA up into smaller fragments.
 - Checking the quality and quantity of the fragment library.
 - All of the above.
- Which of the below sequencing techniques require DNA amplification during the library preparation step (is considered a 2nd generation sequencing technique)?
 - PacBio AND Oxford Nanopore.
 - Illumina AND Ion Torrent.
 - Illumina AND Oxford Nanopore.
 - PacBio AND Ion Torrent.
- Which of the below sequencing techniques use(s) fluorescently labeled nucleotides for identifying the nucleotide sequence of the template DNA strand?
 - Illumina AND PacBio.
 - Illumina AND Oxford Nanopore.
 - PacBio AND Ion Torrent.
 - Only Illumina.
 - All sequencing methods use fluorescently labeled nucleotides for identifying the nucleotide sequence of the template DNA strand.
- The below figures illustrate five cycles of Illumina sequencing. The colored spots represent the clusters on the flow cell. What is the sequence of the DNA template (cluster) in the top, left corner according to the figure?
 - Yellow
 - Red
 - Blue
 - Green



7. What is the main enzyme component of Sanger sequencing?
8. Which of the following best describes three cyclic steps in PCR, in the correct order?
 - A. Denaturing DNA to make it single stranded, primer annealing, synthesis of a new DNA strand.
 - B. Primer annealing, denaturing DNA to make it single stranded, synthesis of a new DNA strand.
 - C. Synthesis of a new DNA strand, primer annealing, denaturing DNA to make it single stranded.
9. Which of the following “omes” relates to the DNA sequence of expressable genes?
 - A. Genome.
 - B. Exome.
 - C. Proteome.
 - D. Metabolome.
10. Targeted sequencing:
 - A. Allows to focus on specific areas of interest and thus enables sequencing at much higher coverage levels.
 - B. Is a sequencing method used within molecular diagnostics.
 - C. Is equivalent to sequencing tumor panels.
 - D. This method refers to sequence a novel genome.
11. Library preparation involves generating a collection of DNA fragments for sequencing. NGS libraries are typically prepared by fragmenting a DNA or RNA sample and ligating specialized adapters to both fragments ends. Bring the following working steps in terms of library preparation in the right order:
 - A. Perform End Repair and size selection via AMPure XP Beads.
 - B. Quantification and profile samples.
 - C. Purify Ligation Products.
 - D. Validate Library.
 - E. Adenylate 3'-Ends.
 - F. Normalize and pool libraries.
 - G. Enrich DNA Fragments and size selection via AMPure XP Beads.
 - H. Adapter Ligation and size selection via AMPure XP Beads.
12. Indicate whether each of the following descriptions better applies to Illumina[®] sequencing (I), Ion Torrent[™] sequencing (T), or both sequencing technologies (B).
 - It uses fluorescently labeled nucleotides.
 - It uses PCR-generated copies of DNA.

- It is a second-generation sequencing technology and employs a cyclic wash-and-measure paradigm.
- It relies on the fidelity of DNA polymerase for its accuracy.
- pH change indicates how many bases were added.

Answers to Review Questions

1D; 2A; 3D; 4B; 5A; 6 GAGAC; 7 Polymerase; 8A; 9B; 10A, B, C; 11 BAEHCGDF; 12 IBBBT;

Acknowledgements We are grateful to Melinda Hutcheon, (Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, US) for reviewing this chapter. We would also like to thank *Illumina* for the material and illustrations provided to help us write this chapter.

References

1. Hood L, Rowen L. The human genome project: big science transforms biology and medicine. *Genome Med.* 2013;5(9):79.
2. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genetics.* 2016;17(6):333–51.
3. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour.* 2011;11(5):759–69.
4. Morganti S, Tarantino P, Ferraro E, D’Amico P, Viale G, Trapani D, et al. Complexity of genome sequencing and reporting: next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Crit Rev Oncol Hematol.* 2019;133:171–82.
5. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics.* 2015;13(5):278–89.
6. Muller CA, Boemo MA, Spingardi P, Kessler BM, Kriaucionis S, Simpson JT, et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods.* 2019;16(5):429–36.
7. Wanunu M. Nanopores: a journey towards DNA sequencing. *Phys Life Rev.* 2012;9(2):125–58.
8. Jain M, Olsen HE, Paten B, Akeson M. The Oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016;17(1):239.
9. De Coster W, Strazisar M, De Rijk P. Critical length in long-read resequencing. *NAR Genomics Bioinform.* 2020;2(1):lqz027.
10. McCombie WR, McPherson JD, Mardis ER. Next-generation sequencing technologies. *Cold Spring Harb Perspect Med.* 2019;9(11):a033027.