

Chapter 6

New Horizons in Next-Generation Sequencing

Abstract In the previous chapters, we described the most common and well-established next-generation sequencing technologies and platforms. However, several methodologies and sequencers with outstanding features have also been released in the last few years. Furthermore, additional technologies demonstrating great promise are currently in development. In this chapter, we will briefly describe these recent and ongoing developments that may have a profound impact on the future of sequencing.

6.1 Third-Generation Sequencing Methods

Despite the advantages of next-generation sequencing methods, soaring expectations in the field have driven the demand for even better technologies (see Chap. 5). Therefore, a new staple of sequencing methods known as third-generation sequencing or next-generation sequencing is being developed in the hopes of elevating the platform to a whole new dimension [1]. Ideally, third-generation sequencing methodology should reduce or eliminate some or all of the three main challenges faced by the next-generation techniques, i.e., excessive machine costs, short read lengths, and significant error rate. To date, three methods have been introduced that can be considered as third-generation methods or in the transitional phase between the next-generation and third-generation tools.

6.1.1 Heliscope Single-Molecule Sequencing

Heliscope Single-Molecule Sequencing (or Helicos Single-Molecule Fluorescent Sequencing) is the first single-molecule sequencing (SMS) method that can directly identify the exact sequence of a given DNA stretch [2]. In this technique, the DNA

to be sequenced is sheared and the resulting fragments are then attached to Poly-A tails, which allow the fragments to be connected to a flow cell surface. A single type of fluorescently labeled nucleotide is added in cycles to extend the DNA by one nucleotide per cycle. After the addition of each nucleotide, the reaction is paused using a terminating nucleotide in order to capture an image of the fluorescent label. Subsequently, the flow cell surface is washed and the blocking is removed to repeat the cycle [3]. This technology was developed by Helicos Biosciences and was used in 2009 to sequence whole human genome (the genome of Stephen Quake, Professor of Stanford University, USA and a co-founder of Helicos BioSciences) for less than 50,000 dollars [2]. It was also used to sequence the genome of the M13 bacteriophage [4]. However, by the end of 2012, Helicos BioSciences shut its doors and filed for bankruptcy.

6.1.2 Single-Molecule Real-Time Sequencing

The single-molecule real-time (SMRT) sequencing technique is another SMS method that is based on the principle of sequencing by synthesis. It utilizes small well-like containers with a single DNA polymerase enzyme affixed at the bottom of a structure called the zero-mode waveguide (ZMW) [5]. Each ZMW contains a polymerase enzyme and a DNA fragment as a template, and creates an observation volume that is sufficiently illuminated to view a single nucleotide when being incorporated by DNA polymerase. This observation is accomplished through capturing the fluorescent label of the incorporated nucleotide by a detector [6]. The SMRT Sequencing technology was developed by Pacific Biosciences and is currently implemented in their commercial sequencing machines, where the actual sequencing is fulfilled on a chip that contains several ZMWs (see below).

6.1.3 Nanopore Sequencing

The Nanopore sequencing method was first introduced in the middle of the 1990s as a technique for determining the nucleotide order in a DNA sequence [7]. The technique is based on the utilization of a surface comprising of 1 nm diameter pores. The passage of DNA through a pore alters its ion current. This effect is indicative of the types of nucleotides present as current changes depend on the shape, size, and length of the DNA molecules being sequenced. Thus, each nucleotide can be identified based on its corresponding ion blockage time. Nanopore sequencing is a promising and low-cost method that does not require modified nucleotides, chemical labeling, or PCR amplification [8].

The major challenge of utilizing the nanopore method is the preparation involved in developing the nanopore surface, which can be either solid-state nanopore

surfaces or protein-based nanopore surfaces. Solid-state surfaces are used in solid-state nanopore sequencing techniques such as sequencing with fluorescent labels [9]. On the other hand, protein-based nanopore sequencing employs proteins such as *Alpha hemolysin* and *Mycobacterium smegmatis porin A* (MspA) as nanopore surfaces [10–12]. Nanopore sequencing is still in the developmental stages, and thus far have not been commercially available [13, 14].

6.2 Third-Generation Sequencing Platforms

6.2.1 *HeliScope Single-Molecule Sequencer*

The Heliscope Single-Molecule Sequencer was the first commercialized SMS developed by Helicos Biosciences in 2009. It implements the Heliscope SMS technology that was developed by the same company and represents a revolutionary sequencing paradigm that allows the sequencing of about one billion molecules in about 7 days, a rate 1,000-fold over the technology available when first released [2]. It uses novel reagents that allow digital measurement of homopolymer sequences as well as a new alignment algorithm to perform whole genome assembly (reference-based assembly). The sequencer reads are between 24 and 70 bp, which are very short based on previous expectations from a third-generation product. However, the higher speed of sequencing and lower associated costs are the significant strengths of the platform.

The Heliscope Single-Molecule Sequencer was used to sequence the genome of one of the co-founders of Helicos Biosciences (referred to as Patient Zero or P0 in the published article), with promising results [2]. Four sequencers were used to sequence the whole human genome and the results were mapped to ~90 % of the reference genome with a coverage depth near a Poisson distribution [2]. However, Helicos Biosciences closed down at the end of 2012 and, therefore, the Heliscope Single-Molecule Sequencer was excluded from comparisons in this chapter.

6.2.2 *PacBio RS II*

PacBio RS is a DNA sequencing system developed by Pacific Biosciences. The PacBio RS systems (PacBio RS and PacBio RS II) are single-molecule sequencers that implement the SMRT sequencing technology developed by the same company. These can be considered as genuine third-generation sequencers with a read length that is >3,000 bp, which is one of the longest available read lengths to date. The sequencer is compact with a short run time (~10 h). However, it is very expensive and still suffers from high error rates and a low total number of reads per run (Tables 6.1, 6.2, 6.3, and 6.4) [13, 14].

Table 6.1 Comparison of major third-generation sequencers advantages and disadvantages^a

Instrument	Primary advantages	Primary disadvantages
Oxford Nanopore GridION 2000	Extremely long reads are feasible; low-cost instrument (node); nodes can be placed in standard computer racks; error rate doesn't increase along the length of the read; hairpin on one end allows reading of the complementary strand	Not yet available; no data publicly available; 4 % error rates; errors are likely to be biased (thus multiple reads will lead to higher confidence in the wrong answer)
Oxford Nanopore GridION 8000	Same as GridION 2000, but more reads per unit time; lower cost per GB	Same as GridION 2000
Oxford Nanopore minION	No Instrument; IT IS A USB DEVICE; can load "raw" samples	Not yet available; no data publicly available; high cost per MB relative to other Nanopore sequencers
PacBio	Single-molecule real-time sequencing; longest available read length; ability to detect base modifications; short instrument run time; random error profile; modest cost per sample; many methods being developed	High error rates; low total number of reads per run; high cost per MB; high capital cost; many methods still in development; weak company performance

^aThe data is derived from 2013 NGS Field Guide update [13, 33]

Table 6.2 Comparison of major third-generation sequencers run time, read length, and output data^a

Instrument	Run time (h)	Millions of reads/run	Read length	Yield MB/run	Data file sizes (GB)
Oxford Nanopore GridION 2000	[NA]	[4]	[10,000]	[40,000]	[Variable]
Oxford Nanopore GridION 8000	[5]	[10]	[10,000]	[100,000]	[Variable]
Oxford Nanopore minion	≤6	[0.1]	[9,000]	1,000	[Small]
PacBio RS	≤2	0.03	>3,000	100–150	2 (basecalls, QV, kinetics)

^aThe data is derived from 2013 NGS Field Guide update [13, 33]

Table 6.3 Comparison of major third-generation sequencers purchase and operation costs^a

Instrument	Reagent cost/run	Reagent cost/MB	Minimum unit cost (% run)	Purchase cost	Service contract	Computational resources
Oxford Nanopore GridION 2000	Varies	[\$0.04]	–	–	–	–
Oxford Nanopore GridION 8000	Varies	\$0.02	–	–	–	–
Oxford Nanopore minion	≤\$900	\$1	~\$1,100 (10 %)	–	–	Laptop
PacBio RS	≥\$300	\$2–17	\$500 (100 %)	\$695	85	\$65 cluster

^aThe data is derived from 2013 NGS Field Guide update [13], some values are from 2011's guide [33]. All costs are in thousands of US dollars

Table 6.4 Comparison of major next-generation sequencers errors and error rates^a

Instrument	Data file sizes (GB)	Primary errors	Single-pass error rate (%)	Final error rate (%)
Oxford Nanopore GridION 2000	[Variable]	Deletions	$\geq 4^a$	4^a
Oxford Nanopore GridION 8000	[Variable]	Deletions	$\geq 4^a$	4^a
Oxford Nanopore minion	[Small]	Deletions	$\geq 4^a$	4^a
PacBio RS	2 (basecalls, QV, kinetics)	Indel	~ 13	≤ 1

^aThe data is derived from 2013 NGS Field Guide update [13, 33]

6.2.3 Oxford Nanopore GridION

The Oxford Nanopore GridION sequencers are sequencing machines that implement the Nanopore sequencing methodology. The sequencers are being developed by Oxford Nanopore Technologies Ltd. (UK), which had originally announced that their first commercialized instrument would be available by the end of 2013 [15]. However, at the time of manuscript preparation, it had not yet launched.

The Oxford Nanopore GridION systems promise small, inexpensive and high-throughput sequencers with an unprecedented long read length of $\sim 10,000$ bp. According to the product page on the company website [15], the Oxford Nanopore GridION can be used as a single desktop machine or stacked in racks in a similar manner to computer servers. Furthermore, it is stated that the instrument does not require a dedicated server and utilizes a single-use disposable cartridge that contains all the reagents necessary for the experiment. The available information on the performance of the Oxford Nanopore GridION systems shows a relatively high error rate ($\sim 4\%$), though this rate does not rise upon increasing the read length [13, 14].

6.3 Sequencing Methods Under Development

We have previously discussed the rapid rate at which methodology has been developed in the DNA sequencing field, and how this fact has helped alleviate prior technical challenges. Moreover, several additional methods are currently in development and hold the promise of making DNA sequencing cheaper, easier, faster, and more accurate. The ultimate goal of these developments is to make whole human genome DNA sequencing as simple and affordable as other standard laboratory procedures. This would allow its widespread utilization towards innumerable clinical applications such as personalized medicine, and would augment research to unprecedented levels [16]. In this section, we will discuss methodologies that are presently in the developmental phase as well as their expected outcomes.

6.3.1 Solution-Based Hybridization Sequencing

The idea behind sequencing by hybridization is not a new one and has been previously presented [17]. Sequencing by hybridization involves a nonenzymatic approach based on the creation of a hybrid between the DNA molecule of interest and another molecule of known sequence. When one short strand of DNA binds to its complementary strand, the binding becomes very sensitive to mismatches, even at the level of a single-base. Thus, the sequence of the complementary strand can be inferred from the sequence of its hybrid. The method requires a library of DNA probes (short single-stranded DNA sequences) based on the organism of interest, its variants or its single-base variations, and can be accomplished using DNA chips or microarrays [17]. The technique has several advantages including homogenous coverage, though the preliminary requirement of DNA and the need for a significant amount of chemicals limit its overall utility. However, the recent introduction of solution-based hybridization has drastically reduced the dependency on chemicals and expensive equipment [18, 19].

6.3.2 Tunneling Current DNA Sequencing

The novel approach of identifying a DNA sequence and differentiating between the four types of nucleotides through the use of electrical signals was first presented via nanopore sequencing [8]. Based on these findings, the Tunneling Current DNA Sequencing method identifies specific nucleotides through tunneling current conducted by single-base molecules as they pass through a channel comprising of a pair of nanoelectrodes [10, 20, 21]. The differing structures of the nucleotides have varied effects on the current during this process. Thus, differentiating between them is possible through the identification of the characteristic changes in the current influenced by each nucleotide. A recent report also presented a hybrid method that combined single-base electrical identification and random sequencing to allow successful sequence reads from nine different DNA oligomers and microRNA [21]. The method promises an elevated sequencing speed in comparison to those currently available.

6.3.3 Microscopy-Based DNA Sequencing

Microscopy-based DNA Sequencing utilizes an electron microscope to directly visualize the nucleotide sequence of intact DNA molecules. In this approach, nucleotides are enzymatically modified to contain atoms with higher atomic number that can be directly visualized and identified by the electron microscope. Using this technique, an intact synthetic molecule of length >3,200 bp and an intact viral DNA of length >7,000 bp were sequenced successfully, proving the potential of this methodology in the sequencing of long intact DNA molecules [22].

6.3.4 Mass Spectrometry-Based DNA Sequencing

Mass Spectrometry is well known as the technology of choice in the study of proteins and the identification of amino acid sequences [23]. Additionally, it is utilized in the study of metabolites via the capillary electrophoresis mass spectrometry (CE-MS) approach [24]. For the purposes of DNA sequencing, electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) were used to determine the nucleotide sequence of DNA through the examination of nucleotide mass. This contrasted with previous methodology that employed the study of nucleotide size, structure, fluorescent labeling, or radioactive labeling [25, 26]. Since each type of nucleotide has its own unique chemical structure, each of them also possesses a unique mass. Therefore, spectrometry can be used to identify the nucleotide sequences accurately and in high resolution. This method was found to be more effective with RNA, so the DNA is converted to RNA prior to the sequencing process. An early attempt to use MS for DNA sequencing showed that the longest read in the procedure could be 100 bp [27]. In more recent studies, MS-based DNA sequencing has been used to identify SNPs in pathogens [26] and the comparison of human mitochondrial DNA with DNA from the bones of dead soldiers during a forensic investigation [28].

6.3.5 RNA Polymerase Sequencing

RNA polymerase (RNAP) Sequencing involves the utilization of an RNAP enzyme that is attached to a polystyrene bead while the DNA molecule to be sequenced is attached to another bead, following which the two beads are placed in optical traps. The sequencing information is obtained from the movement of the nucleic acid enzyme and the sensitivity of the optical trap. During transcription, the motion of the RNAP brings the two beads closer, which can be recorded in single nucleotide resolution (in Angstrom range). The differentiation between the four types of nucleotides is then accomplished using a Sanger approach-like method. The concentration displacement of the four types of nucleotides over the transcription time is compared and used to pinpoint the specific types of the nucleotides in the sequence [29, 30].

In addition to the above, several other sequencing methods and instruments are currently either in the research phase or at the initial stages of commercialization. These include in vitro virus high-throughput sequencing [31] and microfluidic Sanger sequencing [32], for instance. However, due to text limitations, it is not possible to discuss them all within the confines of this book. Reports that survey or compare upcoming methods and platforms are readily available [13, 30], though the rapid pace of the field necessitates sources that are frequently updated such as the NGS Field Guide [33].

References

1. Rusk N (2009) Cheap third-generation sequencing. *Nature Methods* 6 (4):244-245. doi:[10.1038/nmeth0409-244a](https://doi.org/10.1038/nmeth0409-244a)
2. Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. *Nature Biotechnology* 27 (9):847-850. doi:[10.1038/Nbt.1561](https://doi.org/10.1038/Nbt.1561)
3. Thompson JF, Steinmann KE (2010) Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol Chapter 7:Unit7 10*. doi:[10.1002/0471142727.mb0710s92](https://doi.org/10.1002/0471142727.mb0710s92)
4. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J et al. (2008) Single-molecule DNA sequencing of a viral genome. *Science* 320 (5872):106-109. doi:[10.1126/science.1150427](https://doi.org/10.1126/science.1150427)
5. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG et al. (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299 (5607):682-686. doi:[10.1126/science.1079700](https://doi.org/10.1126/science.1079700)
6. Eid J, Fehr A, Gray J, Luong K, Lyle J et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323 (5910):133-138. doi:[10.1126/science.1162986](https://doi.org/10.1126/science.1162986)
7. Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* 93 (24):13770-13773
8. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. *Hum Mol Genet* 19 (R2):R227-240. doi:[10.1093/hmg/ddq416](https://doi.org/10.1093/hmg/ddq416)
9. McNally B, Singer A, Yu Z, Sun Y, Weng Z et al. (2010) Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett* 10 (6):2237-2244. doi:[10.1021/nl1012147](https://doi.org/10.1021/nl1012147)
10. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H (2009) Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A* 106 (19):7702-7707. doi:[10.1073/pnas.0901054106](https://doi.org/10.1073/pnas.0901054106)
11. Purnell RF, Mehta KK, Schmidt JJ (2008) Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano Lett* 8 (9):3029-3034. doi:[10.1021/nl802312f](https://doi.org/10.1021/nl802312f)
12. Stoddart D, Maglia G, Mikhailova E, Heron AJ, Bayley H (2010) Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew Chem Int Ed Engl* 49 (3):556-559. doi:[10.1002/anie.200905483](https://doi.org/10.1002/anie.200905483)
13. Glenn TC (2011) Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 11 (5):759-769. doi:[10.1111/j.1755-0998.2011.03024.x](https://doi.org/10.1111/j.1755-0998.2011.03024.x)
14. Glenn TC (2013) Field guide to next-generation DNA sequencers-Update. <http://www.molecular ecologist.com/next-gen-fieldguide-2013/>. Accessed 10-01-2014
15. Oxford Nanopore Technologies Ltd. (2014) The GridION System. <https://www.nanoporetech.com/technology/the-gridion-system/the-gridion-system>. Accessed 10-01-2014
16. Collins FS, Hamburg MA (2013) First FDA authorization for next-generation sequencer. *N Engl J Med* 369 (25):2369-2371. doi:[10.1056/NEJMp1314561](https://doi.org/10.1056/NEJMp1314561)
17. Hanna GJ, Johnson VA, Kuritzkes DR, Richman DD, Martinez-Picado J et al. (2000) Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase. *J Clin Microbiol* 38 (7):2715-2721
18. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML et al. (2013) A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* 110 (1-2):3-24. doi:[10.1016/j.ymgme.2013.04.024](https://doi.org/10.1016/j.ymgme.2013.04.024)
19. Qin Y, Schneider TM, Brenner MP (2012) Sequencing by hybridization of long targets. *PLoS One* 7 (5):e35819. doi:[10.1371/journal.pone.0035819](https://doi.org/10.1371/journal.pone.0035819)
20. Di Ventra M (2013) Fast DNA sequencing by electrical means inches closer. *Nanotechnology* 24 (34):342501. doi:[10.1088/0957-4484/24/34/342501](https://doi.org/10.1088/0957-4484/24/34/342501)
21. Ohshiro T, Matsubara K, Tsutsui M, Furuhashi M, Taniguchi M et al. (2012) Single-molecule electrical random resequencing of DNA and RNA. *Sci Rep* 2:501. doi:[10.1038/srep00501](https://doi.org/10.1038/srep00501)

22. Bell DC, Thomas WK, Murtagh KM, Dionne CA, Graham AC et al. (2012) DNA base identification by electron microscopy. *Microsc Microanal* 18 (5):1049-1053. doi:[10.1017/S1431927612012615](https://doi.org/10.1017/S1431927612012615)
23. Helmy M, Tomita M, Ishihama Y (2012) Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatics methods in proteogenomics. *Genes Genome Genomics* 6:76-85
24. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316 (5824):593-597. doi:[10.1126/science.1132067](https://doi.org/10.1126/science.1132067)
25. Edwards JR, Ruparel H, Ju J (2005) Mass-spectrometry DNA sequencing. *Mutat Res* 573 (1-2):3-12. doi:[S0027-5107\(05\)00023-0](https://doi.org/S0027-5107(05)00023-0)
26. Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC et al. (2010) Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A* 107 (9):4371-4376. doi:[10.1073/pnas.0911295107](https://doi.org/10.1073/pnas.0911295107)
27. Monforte JA, Becker CH (1997) High-throughput DNA analysis by time-of-flight mass spectrometry. *Nat Med* 3 (3):360-362
28. Howard R, Encheva V, Thomson J, Bache K, Chan YT et al. (2013) Comparative analysis of human mitochondrial DNA from World War I bone samples by DNA sequencing and ESI-TOF mass spectrometry. *Forensic Sci Int Genet* 7 (1):1-9. doi:[10.1016/j.fsigen.2011.05.009](https://doi.org/10.1016/j.fsigen.2011.05.009)
29. Greenleaf WJ, Block SM (2006) Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science* 313 (5788):801. doi:[313/5788/801](https://doi.org/313/5788/801)
30. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. *J Appl Genet* 52 (4):413-435. doi:[10.1007/s13353-011-0057-x](https://doi.org/10.1007/s13353-011-0057-x)
31. Fujimori S, Hirai N, Ohashi H, Masuoka K, Nishikimi A et al. (2012) Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Sci Rep* 2:691. doi:[10.1038/srep00691](https://doi.org/10.1038/srep00691)
32. Chen YJ, Roller EE, Huang X (2010) DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device. *Lab Chip* 10 (9):1153-1159. doi:[10.1039/b921417h](https://doi.org/10.1039/b921417h)
33. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8):1072-1075. doi:[10.1093/bioinformatics/btt086](https://doi.org/10.1093/bioinformatics/btt086)