The emergence of nanopores in next-generation sequencing

L J Steinbock and A Radenovic

STI IBI LBEN, Station 17, Lausanne, Switzerland

E-mail: Aleksandra.radenovic@epfl.ch

CrossMark

Abstract

Next-generation sequencing methods based on nanopore technology have recently gained considerable attention, mainly because they promise affordable and fast genome sequencing by providing long read lengths (5 kbp) and do not require additional DNA amplification or enzymatic incorporation of modified nucleotides. This permits health care providers and research facilities to decode a genome within hours for less than $1000. This review summarizes past, present, and future DNA sequencing techniques, which are realized by nanopore approaches such as those pursued by Oxford Nanopore Technologies.

Keywords: nanopore sequencing, next generation sequencing, oxford nanopore technologies, nanopore, sequencing

(Some figures may appear in colour only in the online journal)

DNA is the blueprint for all organisms. When a bacterium acquires resistance to an antibiotic, or when a healthy cell mutates into a tumor cell, it is due to changes in the DNA sequences. If it were possible to swiftly sequence human, bacterial, and viral genomes economically, health care would experience a quantum leap in the diagnosis, understanding, and treatment of diseases. For this reason, various techniques for DNA sequencing have been developed since the early 1970s. Today the largest market for sequencing technologies is the academic and corporate research environment, reaching almost $1.4 billion US dollars. In particular, clinical applications for DNA sequencing are showing the fastest growth (figure 1(a)) and could surpass the market for research applications in a couple of years.

First- and second-generation sequencing methods

First-generation sequencing is based on the Sanger method, which was presented in 1977 by the scientist Frederick Sanger [1]. In this method, natural and chain-terminating nucleotides are incorporated by a polymerase into a growing DNA chain during replication. The random incorporation of chain-terminating nucleotides, which are either fluorescently or radioactively labeled during the polymerase chain reaction (PCR), leads to a population of DNA strands with different lengths. These DNA strands are then separated according their size by capillary electrophoresis. A laser combined with a fluorescence detector detects the fluorescently labeled terminated DNA when the molecules pass through the capillary [2], which allows them to be sequenced. The maximum read length of such fragments is about 1000 base pairs (bp). Devices based on the Sanger sequencing method were the main instruments used in the Human Genome Project, which, in 2001, after a decade of multinational scientific cooperation and a cost of about $3 billion, sequenced the entire human genome [3].

The second generation of DNA sequencing instruments works by detecting the incorporation of the labeled nucleotides directly and prevents the necessity of separating the DNA in a gel [4–6]. However, since earlier optical sensors were not able to detect the incorporation of a single nucleotide, a PCR step is still needed to amplify the DNA molecules. This creates a large number of fluorescently labeled DNA molecules to generate enough photons to excite the optical detectors. Second-generation instruments are sold by companies such as 454 Life Sciences, Illumina, and Ion Torrent [7]. Illumina, which was one of the first next-generation device manufacturers, dominates today's market (figure 1(b)) with various machines such as MiSeq, GAIIx, and HiSeq that target different customer needs such as cost and device performance (figure 1(c)). The supremacy of
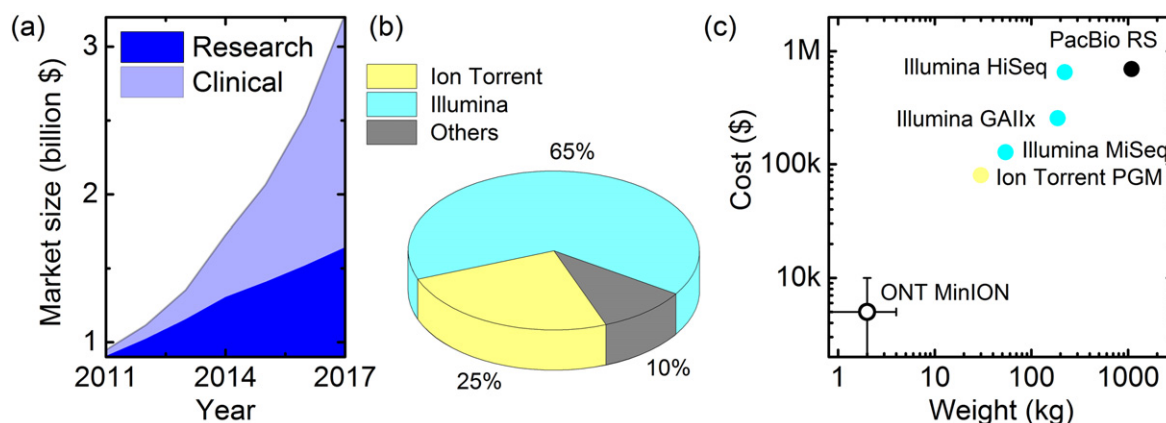
Figure 1. (a) Market size for DNA sequencing, divided into research and clinical applications. Clinical applications such as Down syndrome testing for prenatal diagnostics were a late entry into the market due to regulatory hurdles but are predicted to display stronger market growth to surpass the market for research applications in the DNA sequencing sector by 2017. (b) Currently the DNA sequencing device market is dominated by Illumina, which owns approximately 65% of the market. At the same time, Ion Torrent has about 25% of the market, whereas the rest is split between PacBio and 454 Life Sciences and also other, older, instrument manufacturers. (c) Cost as a function of the weight of current and future DNA sequencing devices such as the MinION from ONT. PacBio RS represents the latest next-generation sequencing instruments that are entering the market. Although devices from Illumina and Ion Torrent tend to occupy the middle range of this matrix, devices from Oxford Nanopore Technologies might represent a quantum leap. Since the MinION is not yet commercially available, the size and price presented here are estimates (hence the use of error bars). Although the error rate of the MinION might be high, its read length will probably reach up to 20 kbp and will require only a USB port for its power supply.
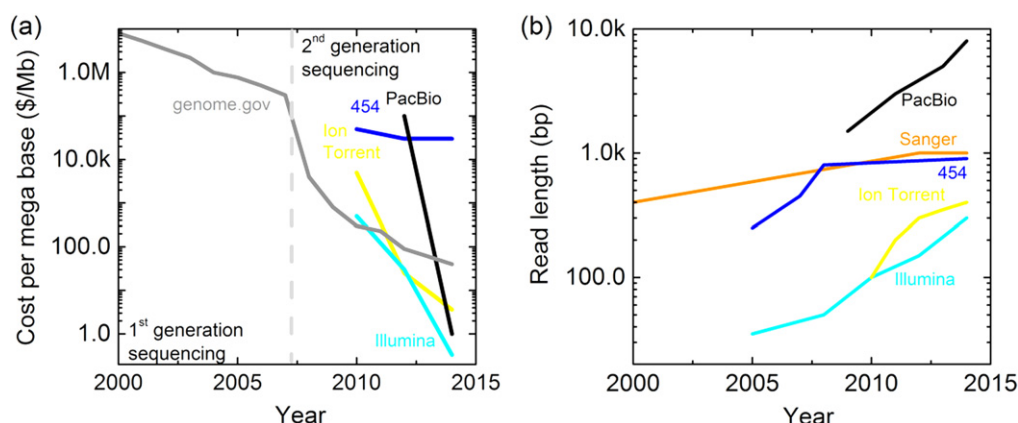

Figure 2. (a) Cost per base of the different sequencing techniques as a function of time. The gray curve shows data calculated by the National Human Genome Research Institute, United States, and represents the average costs, including reagents and instruments [38]. The introduction of next-generation sequencing devices in 2007 has increased the rate of cost improvement. The blue, black, green, and magenta curves show the decline in costs for the various NGS techniques such as 454 Life Sciences, Pacific Biosystems, Ion Torrent, and Illumina, respectively. (b) Read lengths plotted against time. The orange curve represents the original Sanger sequencing technique, which was ideal for *de novo* genome sequencing tasks but is gradually being replaced by instruments from Pacific Biosystems. Techniques from Ion Torrent and Illumina cannot perform long read lengths but are profitable due to their low cost-to-base ratio.

Illumina is mainly due to its competitive cost per base performance (figure 2(a)). Introduction of the second-generation devices in 2007 considerably accelerated the reduction in the cost-to-base ratio (dashed line in figure 2(a)).

Although these new techniques drove down costs and time, their read length was lower (around 200 bp) than the pioneering Sanger method, which reached up to 1000 bp (figure 2(b)). Since most sequence targets are much longer than 200, the DNA has to be over-sequenced to generate enough overlapping regions to merge the pieces for the entire sequence. Because many genomes have already been decoded, many projects require only sequencing of short segments, rendering Illumina's devices ideal for these kinds of tasks.

In 2008, a second-generation device from 454 Life Sciences with an improved read length of about 800 bp managed to sequence an entire human genome in a few weeks. This improvement reduced the speed and costs for sequencing a genome to about $1 million [8]. The last addition to the market of second-generation sequencing instruments is manufactured by Ion Torrent; this device is based on a similar technique as 454 Life Sciences, but it senses the incorporation of a nucleotide electrically by using a small, sensitive CMOS-compatible pH meter [9]. Instead of detecting the incorporation of a nucleotide through a fluorescent dye, the generation of a hydrogen atom, which is released each time a nucleotide is added into the growing strand, is probed. The pH meter

chip replaces the camera and fluorescent molecules in the 454 device and permits cheaper sequencing through lower reagent and device costs. This advantage and an aggressive prize strategy has enabled fast market penetration by Ion Torrent, representing approximately 25% of all purchased instruments (figure 1(b)).

Next-generation sequencing methods

The current next-generation sequencing (NGS) technologies are defined by several characteristics. First, they are able to detect a single unmodified nucleotide by relying on new optical and electrical single-molecule techniques. This avoids the need for amplification by PCR, thereby reducing time and expenses for reagents. Another advantage lies in the extended read length, which surpasses the 1000 bp limit, reaching up to 50 kbp; this reduces the need for over-sequencing to obtain enough overlapping DNA sequences. Similar to the first feature, the increased read length also reduces the amount of reagents and time needed to oversequence the DNA, thus further driving costs down.

The first next-generation sequencing instrument, developed by Pacific Biosystems, is referred to as single-molecule real-time sequencing (SMRT) and has been available since 2011. SMRT is based on a chip pioneered by Levene *et al* [10] that contains an array of zero-mode waveguides (ZMWs). A single DNA polymerase is attached to the bottom of the well, and the ZMWs create an illuminated volume that is small enough to observe the incorporation of a single nucleotide [10]. Each time a nucleotide is added to the DNA at the bottom of the well, the dye is detected before being cleaved off and diffusing away. Although the devices are very expensive (figure 1(c)), they are ideal for *de novo* sequencing of unknown genomes due to their long read lengths of 8 kbp (figure 2(b)). They even permit the quantification of methylation, DNA damage, and other epigenetic information [11].

Another NGS technique that is gaining much attention recently is based on nanopores [12, 13]. The first detection of DNA using the biological nanopore α-hemolysin was accomplished in 1996 by John Kasianowitz *et al* [14]. The nanopore was incorporated into a phospholipid bilayer, which separated two reservoirs filled with a KCl solution. Using two electrodes placed on opposite sides of the bilayer, an electrical potential can be applied, causing an ionic current. DNA, which is negatively charged, can therefore be forced to translocate through the nanopore by applying a positive potential to the electrode on the opposite side of the membrane. The translocation velocity varies, depending on parameters such as the electrical potential, the type of nanopore, and whether the DNA is single-stranded or double-stranded [15]. The optimal velocity is around 2 nucleotides per millisecond, which will enable sequencing of a human genome in 8 h in a 10×10 array. During the translocation, the ionic current is partially blocked, leading to a reduction in the current and allowing differentiation between the four different nucleotides [16, 17]. The amplitude and duration of these blockades depend on the length and width of the translocating
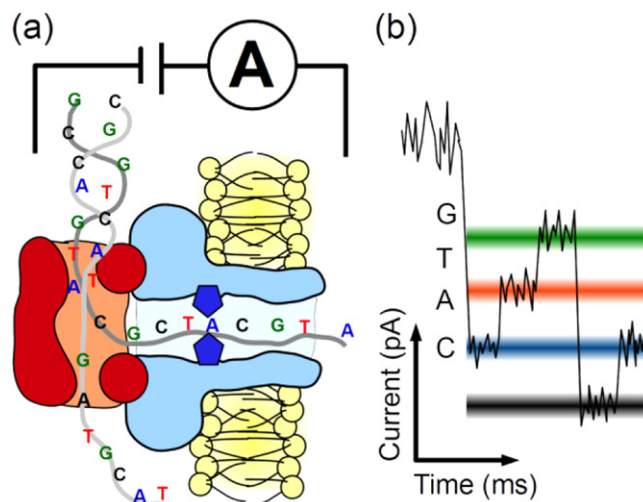

Figure 3. Scheme depicting the sequencing technique of Oxford Nanopore Technologies. (a) Double-stranded DNA is separated into single-stranded DNA by a polymerase such as phi29. This decelerates the translocation velocity of the ssDNA through the nanopore. The nanopore possesses a constriction inside the channel (dark blue diamond), which enables reading of the ssDNA sequence. (b) Simplified scheme depicting the decoding method. The ionic current trace is altered by the ssDNA sequence translocating through the nanopore, as shown in figure 3(a). Each level represents one nucleotide residing inside the nanopore at a specific point in time. By detecting these levels, the sequence of the DNA can be decoded. In the current experiment, the nanopore is not sensitive enough to detect one single nucleotide, but four nucleotides can increase the number of levels to 256.

polymer. Classical biological nanopores such as α-hemolysin are only big enough to allow single-stranded DNA to pass through, and they probe 10 to 12 nucleotides at a time (figure 3(a)) [18]. Recent advances, such as the shorter and narrower nanopore MspA, permit the identification of four nucleotides residing inside the nanopore by determining the amount of blocked current (figure 3(b)). Combining the MspA with the phi29 DNA polymerase can slow the translocation of DNA through the nanopore and subsequently sequence 50-nt-long DNA fragments (figure 3(a)) [19]. Recent advances in signal processing have extended this approach to 5-kilobase-long DNA strands [20]. Another critical parameter in this technique is the translocation velocity [21], which has to be optimized to allow sufficient time for the acquisition of enough data points to determine the level and hence the nucleotide. However, the velocity cannot be too slow because it will result in a small throughput performance.

Oxford Nanopore Technologies (ONT) is developing a device based on an array of biological nanopores as described in figures 3(a) and (b) and launched a beta test at the beginning of this year [22]. A commercial launch has not yet been disclosed, but if the technology is coupled with a device that enables reliable decoding of long sequences with an acceptable error rate, it could change the current landscape of DNA sequencing. In particular, the low cost and footprint (weight and volume) could make these devices ideal for private users, field scientists in remote areas, and food processing industries (figure 1(c)).

Other nanopore developments concentrate largely on solid-sate nanopores. Promising approaches are 2D materials made out of graphene and molybdenum disulfide (MoS2) made by exfoliation and chemical vapor deposition. Single-layered 2D materials can reach a thinness of only 0.2 nm, making them ideal candidates for interrogating single-stranded DNA, whose bases are 0.34 nm apart. Although DNA translocation through graphene nanopores was demonstrated first, the disadvantage of the technology was that the DNA adhered to the nanopore and resulted in high background noise [23–25]. A new 2D material is MoS2, which, can also be manufactured in atom-thick layers and fitted with a nanopore. Previous work by Liu *et al* demonstrated that these MoS2 nanopores have a much better signal-to-noise ratio (amplitude of the current change caused by the DNA translocation divided by the noise of the current) of 15, compared with 3.3 for graphene [26, 27]. However, to allow sequencing of single-stranded DNA with these 2D materials, further noise reduction has to be achieved by using materials such as quartz substrates as supporting material [28, 29]. Moreover, since the translocation velocity of dsDNA through solid-state nanopores is relatively high, strategies have to be tested to decrease the translocation velocity. One approach could be to increase viscosity by using high-viscosity media such as glycerol or by using nanopore materials that slow translocation [30–32]. Other avenues envisage the combination of solid-state nanopores with zero-mode waveguides to increase the loading of the ZMWs with the DNA-protein complexes or the integration of electrodes inside the nanopores to sequence the DNA [33–35].

Future next-generation approaches for biological or solid-state nanopores are being pursued by companies such as Genia Technologies, Quantapore, Quantum Biosystems, Base4, and Noblegen Biosciences [36]. Genia Technologies was acquired by Roche in a deal surpassing $300 million, attesting to the potential of the technique and interests from larger pharmaceutical companies in the NGS field. They aim to combine biological nanopores with an optical detection method. Quantum Biosystems is commercializing research by Prof. T Kawai, who combined a tunneling electron detector with a nanopore to sequence DNA [37]. Base4, on the other hand, aims to cleave single nucleotides into droplets in a water–oil emulsion and detect their presence by a chemical cascade reaction. A different approach is being studied by the group of Professor D Stein, which places a mass spectroscopy instrument at the outlet of a nanopore to identify the cleaved nucleotides by their size and charge.

Conclusion and outlook

Since the introduction of the Sanger sequencing method, many new sequencing technologies have advanced read length, increased bandwidth, and reduced costs. Companies such as Pacific Biosystems and 454 Life Sciences differentiate themselves by allowing long read lengths and also information about nucleotide modifications such as methylation. In contrast, companies like Illumina and Ion Torrent attract customers with their low instrument price, user-friendly sample preparation protocols, and low reagent costs, which reduce the cost per base. Illumina even demonstrated an instrument assembly that can sequence a human genome for $1000. This, however, requires the acquisition of several instruments, which can be used solely for sequencing large genomes. New instruments such as those from Oxford Nanopore Technologies offer a disruptive technology with relatively high error rates but very low costs, lower power requirements, and portability (figure 1(c)). This novel technology could tap into a brand-new sector of customers who may require genetic fingerprinting for fast identification of cancer types, pathogen and food safety.

References

[1] Sanger F, Nicklen S and Coulson A R 1997 DNA sequencing with chain-terminating inhibitors *Proc. Natl. Acad. Sci. USA* **12** 5463–7
[2] Kheterpal I and Mathies R A 1999 Peer reviewed: capillary array electrophoresis DNA sequencing *Anal. Chem.* **71** 31A–7A
[3] Lander E S *et al* 2001 Initial sequencing and analysis of the human genome *Nature* **409** 860–921
[4] Ansorge W J 2009 Next-generation DNA sequencing techniques *Nat. Biotechnol.* **25** 195–203
[5] Shendure J and Ji H 2008 Next-generation DNA sequencing *Nat. Biotechnol.* **26** 1135–45
[6] Mardis E R 2008 Next-generation DNA sequencing methods *Annu. Rev. Genomics Hum. Genet.* **9** 387–402
[7] Schuster S C 2008 Next-generation sequencing transforms today's biology *Nat. Methods* **5** 16–8
[8] Wheeler D A *et al* 2008 The complete genome of an individual by massively parallel DNA sequencing *Nature* **452** 872–6
[9] Rothberg J M *et al* 2011 An integrated semiconductor device enabling non-optical genome sequencing *Nature* **475** 348–52
[10] Levene M J *et al* 2003 Zero-mode waveguides for single-molecule analysis at high concentrations *Science* **299** 682–6
[11] Song C-X *et al* 2012 Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine *Nat. Methods* **9** 75–7
[12] Branton D *et al* 2008 The potential and challenges of nanopore sequencing *Nat. Biotechnol.* **26** 1146–53
[13] Wanunu M 2012 Nanopores: a journey towards DNA sequencing *Phys. Life Rev.* **9** 125–58
[14] Kasianowicz J J, Brandin E, Branton D and Deamer D W 1996 Characterization of individual polynucleotide molecules using a membrane channel *Proc. Natl. Acad. Sci. USA* **93** 13770–3
[15] Li J, Gershow M, Stein D, Brandin E and Golovchenko J A 2003 DNA molecules and configurations in a solid-state nanopore microscope *Nat. Mater.* **2** 611–5

[16] Astier Y, Braha O and Bayley H 2006 Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5′-monophosphates by using an engineered protein nanopore equipped with a molecular adapter *J. Am. Chem. Soc.* **128** 1705–10

[17] Stoddart D, Heron A J, Mikhailova E, Maglia G and Bayley H 2009 Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore *Proc. Natl. Acad. Sci. USA* **106** 7702–7

[18] Rusk N 2014 Genomics: nanopores read long genomic DNA *Nat. Methods* **11** 887

[19] Manrao E A *et al* 2012 Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase *Nat. Biotechnol.* **30** 349–53

[20] Laszlo A H *et al* 2014 Decoding long nanopore sequencing reads of natural DNA *Nat. Biotechnol.* **32** 829–33

[21] Venkatesan B M and Bashir R 2011 Nanopore sensors for nucleic acid analysis *Nat. Nanotechnol.* **6** 615–24

[22] Mikheyev A S and Tin M M Y 2014 A first look at the Oxford nanopore MinION sequencer *Mol. Ecol. Resour.* **14** 1097–102

[23] Schneider G F *et al* 2010 DNA translocation through graphene nanopores *Nano Lett.* **10** 3163–7

[24] Garaj S *et al* 2010 Graphene as a subnanometre trans-electrode membrane *Nature* **467** 190–3

[25] Merchant C A *et al* 2010 DNA translocation through graphene nanopores *Nano Lett.* **10** 2915–21

[26] Liu K, Feng J, Kis A and Radenovic A 2014 Atomically thin molybdenum disulfide nanopores with high sensitivity for DNA translocation *ACS Nano* **8** 2504–11

[27] Farimani A B, Min K and Aluru N R 2014 DNA base detection using a single-layer MoS2 *ACS Nano* **8** 7914–22

[28] Steinbock L J, Otto O, Chimerel C, Gornall J and Keyser U F 2010 Detecting DNA folding with nanocapillaries *Nano Lett.* **10** 2493–7

[29] Steinbock L J, Bulushev R D, Krishnan S, Raillon C and Radenovic A 2013 DNA translocation through low-noise glass nanopores *ACS Nano* **7** 11255–62

[30] Kawano R, Schibel A E P, Cauley C and White H S 2009 Controlling the translocation of single-stranded DNA through alpha-hemolysin ion channels using viscosity *Langmuir* **25** 1233–7

[31] Fologea D, Uplinger J, Thomas B, McNabb D S D S and Li J 2005 Slowing DNA translocation in a solid-state nanopore *Nano Lett.* **5** 1734–7

[32] Larkin J *et al* 2013 Slow DNA transport through nanopores in hafnium oxide membranes *ACS Nano* **7** 10121–8

[33] Larkin J, Foquet M, Turner S W, Korlach J and Wanunu M 2014 Reversible positioning of single molecules inside zero-mode waveguides *Nano Lett.* **14** 6023–9

[34] Traversi F *et al* 2013 Detecting the translocation of DNA through a nanopore using graphene nanoribbons *Nat. Nanotechnol.* **8** 939–45

[35] Rutkowska A, Edel J B and Albrecht T 2013 Mapping the ion current distribution in nanopore/electrode devices *ACS Nano* **7** 547–55

[36] McNally B *et al* 2010 Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays *Nano Lett.* **10** 2237–44

[37] Ohshiro T *et al* 2012 Single-molecule electrical random resequencing of DNA and RNA *Sci. Rep.* **2** 501

[38] Wetterstrand K 2014 DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) at (www. genome.gov/sequencingcosts) (website accessed 1/9/2014)