analytical chemistry

REVIEW

pubs.acs.org/ac

Landscape of Next-Generation Sequencing Technologies

Thomas P. Niedringhaus,[†] Denitsa Milanova,[‡] Matthew B. Kerby,[§] Michael P. Snyder,[⊥] and Annelise E. Barron*[,§]

[†]Department of Chemical Engineering, Stanford University, Palo Alto, California, United States

[‡]Department of Mechanical Engineering, Stanford University, Palo Alto, California, United States

[§]Department of Bioengineering, Stanford University, Palo Alto, California, United States

[⊥]Department of Genetics, Stanford University, Palo Alto, California, United States

CONTENTS

Second Generation Sequencing	4327
Second Generation Costs	4328
Third Generation Sequencing Technologies	4328
Single-Molecule Sequencing	4330
Pacific Biosciences	4330
Sequencing by Ligation	4330
Complete Genomics	4330
Sequencing by Synthesis	4333
Ion Torrent	4334
Nanopore Sequencing Technologies	4334
Protein Nanopore Sequencing	4336
Solid-State Nanopore Sequencing	4337
Long Read DNA Extension Methods	4338
Final Assembly by Optical Mapping	4338
Nonoptical, Stretched DNA Molecule Methods	4339
Concluding Remarks	4339
Author Information	4339
Biographies	4339
References	4340

DNA sequencing is in the throes of an enormous technological shift marked by dramatic throughput increases, a precipitously dropping per-base cost of raw sequence, and an accompanying requirement for substantial investment in large capital equipment in order to utilize the technology. Investigations that were, for most, unreachable luxuries just a few years ago (individual genome sequencing, metagenomics studies, and the sequencing of myriad organisms of interest) are being increasingly enabled, at a rapid pace. This Review concentrates on the technology behind the third- and fourth-generation sequencing methods: their challenges, current limitations, and tantalizing promise.

First-generation sequencing encompasses the chain termination method pioneered by Sanger and Coulson[1] in 1975 or the chemical method of Maxam and Gilbert in 1976−1977.[2] In 1977, Sanger sequenced the first genome, bacteriophage ΦX 174, which is 5375 bases in length.[3] These methods and their early history[4] have been reviewed in detail previously.[5] Four-color fluorescent Sanger sequencing, where each color corresponds to one of the four DNA bases, is the method used by the automated capillary electrophoresis (CE) systems marketed by Applied Biosystems Inc., now integrated into Life Technologies, and by Beckman Coulter Inc. (Table 1).[6] The first composite human genome sequence, reported in 2001, was obtained largely using CE, at great cost and with intense human effort over more than a decade.[7,8] While the genome reported in 2001 was a work in progress, the availability of an ever-improving "reference" genome is the basis of an ongoing transformation of biological science and remains fundamental to investigations of genotype−phenotype relationships. Considering reports that have appeared (and not appeared) in the literature to date, it could well be that medically meaningful (actionable) insights into complex diseases will require additional types of "personal" genomic data, for instance, tissue-specific mRNA expression profiling and mRNA sequencing, individualized analysis of gene regulatory regions, epigenetic profiling, and high-quality, long-range chromosome mapping to catalog significant deletions, insertions, rearrangements, etc. Correlation of such integrated genomic data sets with comprehensive medical histories for hundreds or thousands of individuals may be what it takes to reach an era of personalized medicine.[9−11] Large-scale sequencing centers are now completing the conversion to next-generation sequencers; the Joint Genome Institute (JGI) has retired all of their Sanger sequencing instruments.[12] At the other extreme, until small-scale next-generation sequencers can outperform CE on a cost per accurate base called as well as read length, CE systems will likely remain in heavy use for benchtop-scale, targeted sequencing for directed investigations such as quantitative gene expression, biomarker identification, and pathway analysis.

SECOND GENERATION SEQUENCING

Several reviews of what were first called "next-generation" or, more precisely, second-generation sequencing technologies have appeared.[4,13−15] We propose to classify the second-generation technologies as a combination of a synchronized reagent wash of nucleoside triphosphates (NTPs) with a synchronized optical detection method. However, this definition is not rigid, as several

Special Issue: Fundamental and Applied Reviews in Analytical Chemistry

Published: May 25, 2011

ACS Publications © 2011 American Chemical Society

EX2002
Guardant Health, Inc. v. Univ. of Washington
IPR2022-00817

real-time synthesis strategies, which comprise third-generation technologies, also rely on optical detection, with Pacific Biosciences' single DNA polymerase sequencing method being a prime example. Second-generation technologies rely upon sequencing by ligation or sequencing by synthesis, including pyrosequencing and reversible chain termination (Table 1). Commercially available instruments from Roche, Illumina, Helicos, and Life Technologies deliver several Gbp of DNA sequence per week in the form of short contiguous fragments or reads. A review of second-generation methods based on sequencing by synthesis, in which a polymerase or ligase controls the biochemistry, details the challenges and advantages to these types of enzymatic approaches.[16]

SECOND GENERATION COSTS

Over the past few years, companies marketing second-generation sequencers have competed to demonstrate their increasingly cost-effective approaches to generating an assembled complete human genome, relying on the known reference genome. Compared to the costs of generating the draft of J. Craig Venter's genome with ABI's Sanger-CE instruments,[17] Roche's 454 Genome Sequencer FLX,[18] Illumina's Genome Analyzer,[19−21] and Helicos' Heliscope[22] have decreased the cost of obtaining raw sequence by roughly 1, 2, and 3 orders of magnitude, respectively. In all of these reports, only the costs of consumables and reagents were taken into account, however. These new "massively parallel" sequencing instruments require a concomitantly massive investment in capital equipment, since many of these high-throughput instruments are priced between $500K and $1 M each. The labor costs to operate the equipment and the informatics cost for reassembly of the sequence should be factored into the overall sequencing cost. As of this writing, Illumina dominates the market with 60%[23] of the second-generation sequencer installations, while Life Technologies' SOLiD system and Roche split nearly all of the remaining market at 19% each. Illumina's whole-genome sequencing service will sequence a human genome for $19,500,[24] a great deal less than Illumina's reagent costs of $250,000 needed to sequence a complete human genome (or $0.0002 per sequenced base) in 2008[19] and orders of magnitude less than the cost back in 1996, when first-generation sequencing cost $1 per finished base. To reduce costs, Illumina, which uses reversible terminator-based sequencing by synthesis chemistry, recently launched the smaller, less expensive, MiSeq platform, which promises over 1 Gb of 150 bp reads in 27 h. This more compact system is specifically designed to challenge CE-based sequencing for common experiments such as clone verification, amplicon sequencing, and small genome sequencing. On the larger scale, Life Technologies 5500xl series instruments, which use sequencing by ligation chemistry, can collect up to 30 Gb per day over 7 days of operation. For the benchtop market, Ion Torrent, a division of Life Technologies, is developing a third-generation solution and has recently launched the Personal Gene Machine (PGM) and the Ion Express OneTouch template preparation system.[25] The Roche 454 relies on pyrosequencing to detect single base extensions from beads using a luciferase-based method, refined for synchronized DNA sequencing in 1996.[26] The light-emitting pyrosequencing method, which does not use multiple fluorophores, does not require lasers or expensive optical filters, greatly reducing the cost of the equipment. The Roche 454 GLX Flex Titanium series, a $500k instrument, reportedly can generate 400−600 million high-quality base calls per day. New development aims to raise the read length to 800+ base calls.[27] The $100K 454 GS Junior, launched in 2009 and also targeted for benchtop research, produces 35 Mb in 10 h, with 400 base pair reads. "Benchtop" NGS technology development, which squarely challenges first-generation Sanger CE sequencing,[27] seeks to achieve a drastic decrease in cost, physical size, and complexity while continuously increasing throughput, read length, and read accuracy.

In an effort to illustrate the true cost of complete genome sequencing, the National Human Genome Research Institute (NHGRI) has compiled data from their sequencing centers to appropriately estimate the overall costs of sequencing a human genome.[28] Their calculations take into account labor, three-year amortization of sequencing instruments, data processing, and sample preparation. Figure 1 illustrates the cost associated with sequencing a human-sized haploid genome (3,000 Mb) over time since the initial draft of the human genome was published in 2001. The dramatic drop in cost seen in 2008 is the result of transitioning from first-generation Sanger CE sequencing to second-generation platforms installed in sequencing centers (i.e., 454, Illumina, and SOLiD). The second-generation technologies yield lower contiguous read lengths and require greater genome coverage for assembly; however, their high throughput reduces consumable costs and the number of sequencing runs.

Technology development costs and data analysis costs are omitted from these sequencing cost calculations. In general, these costs are much higher for less established second- and third-generation sequencing technologies. For instance, the data depicted in Figure 1 produced by second-generation sequencing technologies (after 2008) are the result of resequencing efforts in which a reference human genome was used to guide the reassembly process. The practicality and cost associated with the sequencing and de novo assembly of a human genome using only second- or third-generation technologies is difficult to assess at this time, given that de novo sequencing has only been accomplished using Sanger-based CE.[29] It appears that the greatest cost barrier is the complex hardware required for the achievement of precisely aligned optical detection and downstream data processing.

THIRD GENERATION SEQUENCING TECHNOLOGIES

With the final goal of bringing the cost of a human genome to under $1000, NIH/NHGRI has funded several groups developing alternative approaches to improving second-generation technologies, as well as novel approaches to sequencing that include the use of scanning tunneling electron microscope (TEM), fluorescence resonance energy transfer (FRET), single-molecule detection, and protein nanopores. Two of the leading third-generation sequencing technologies (Pacific Biosciences and Complete Genomics) still rely on optical detection of fluorescent events but aim to increase sequencing speed and throughput (Table 2). Ion Torrent's technology, on the other hand, uses ion-sensitive field effect transitors (ISFETs) to eliminate the need for optical detection of sequencing events. Nanopore technologies, such as Oxford Nanopore, also aim to remove optics as well as the need for DNA amplification in their sequencing design by measuring changes in conductivity across a nanopore. Nonoptical TEM approaches used by Halcyon Molecular and ZS Genetics require million-dollar capital equipment and, to date, have limited throughput yet, in principle, could give the sequence

Table 1. First- and Second-Generation Sequencing Technologies

Generation	Company	Platform Name	Method of Sequencing	Method of Detection	Approx. Read Length (bases)	Advantages	Relative Limitations
First	ABI/Life technologies	3130xL-3730xL	CE-Sanger	Fluorescence/ Optical	600−1000	Long read lengths; high single-pass accuracy; good ability to call repeats and homopolymer regions	Low throughput; high cost of Sanger sample preparation makes massively parallel sequencing prohibitive
First	Beckman	GeXP Genetic Analysis System	CE-Sanger	Fluorescence/ Optical	600−1000	Long read lengths; high single-pass accuracy; can call repeats/ homopolymer regions; scales down well	Low throughput; relatively high per-sample cost of Sanger sample preparation
Second	Roche/454	Genome Sequencer FLX System	Pyrosequencing	Optical	230−400	Longest read lengths among second generation; high throughput compared to first-generation sequencing	Challenging sample prep; difficulty reading through repetitive/ homopolymer regions; sequential reagent washing gives steady accumulation of errors; expensive instrument ($500K)
Second	Illumina	HiSeq 2000/miSeq	Reversible terminator sequencing by synthesis	Fluorescence/ Optical	2 × 150	Very high throughput	Expensive instrument; significant cost of data reduction and analysis (∼$650K)
Second	ABI/SOLiD	5500xl SOLiD System	Sequencing by ligation	Fluorescence/ Optical	25−35	Very high throughput; lowest reagent cost needed to reassemble a human genome among the widely accepted second generation platforms (Illumina, 454, SOLiD)	Long sequencing runs (days); short reads increase cost and difficulty of data analysis and genome assembly; high instrument cost (∼$700K)
Second	Helicos	HeliScope	Single-molecule sequencing by synthesis	Fluorescence/ Optical	25−30	High throughput; single-molecule nature of technology unique among second-gen platforms	Short reads increasing the costs and reduce quality of genome assembly; very costly instrument (∼$1M)

dx.doi.org/10.1021/ac2010857 |Anal. Chem. 2011, 83, 4327–4341

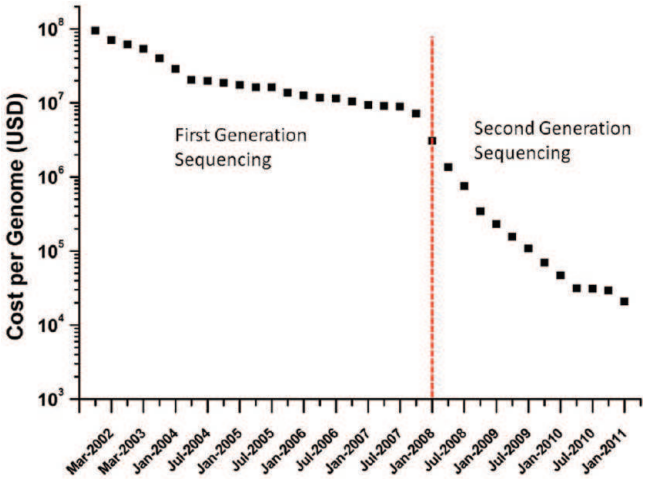

Figure 1. Estimated cost required to sequence a complete human genome based on data generated from NHGRI-funded large-scale DNA sequencing centers.[28]

of thousands of contiguous bases. Finally, new methods involving optical methods are being developed that allow for previously unattainable levels of long-range mapping, which is essential for accurate assembly of individual human genomes and cancer genomes. We now examine these third- and next-generation technologies in detail and outline the advantages and disadvantages of each technique.

SINGLE-MOLECULE SEQUENCING

Pacific Biosciences. Pacific Biosciences (PacBio) has led the charge to develop a reliable third-generation sequencing platform based on a real-time, single-molecule sequencing technology. Their process directly measures DNA polymerase incorporation of fluorescently labeled nucleotides onto a complementary sequencing template. At the heart of this technology is a dense array of zero-mode waveguide (ZMW) nanostructures that allow for optical interrogation of single fluorescent molecules. While ZMW structures have been used in the past to differentiate single fluorescent molecules from substantially large bulk concentrations,[30−34] they have not been used in a highly parallel fashion. To address this issue and increase throughput, PacBio developed a method to efficiently pack ZMW nanostructures onto a surface using electron beam lithography and ultraviolet photolithography[35] as well as a highly parallel confocal imaging system that permits high sensitivity and resolution of fluorescent molecules in each of the ZMW nanostructures.[36] A specialized heavy concrete foundation is employed to maintain optical confocality.

Once the ZMW array fabrication and detection scheme was established, the major technical hurdles for this technology came in the form of immobilizing a single functioning DNA polymerase at the bottom of each ZMW, which can process fluorescently labeled nucleotide substrates. This was accomplished in two steps. First, a set of fluorescently labeled deoxyribonucleoside pentaphosphate (dN5Ps) substrates was synthesized so that each base is spectrally differentiable without decreasing the processivity of the DNA polymerase.[37] Second, surface treatment of the ZMW nanostructure was needed to selectively localize the DNA polymerase. The ZMW array is composed of a fused silica bottom layer and an aluminum top layer in which all ZMW nanostructures are defined. When the aluminum surface was derivatized with polyvinylphosphonic acid (PVPA), protein adsorption to the aluminum layer was significantly decreased without compromising protein adsorption to the bottom glass layer.[38] Combining these chemical modifications with the highly parallel ZMW array, PacBio was able to demonstrate a single-molecule real-time (SMRT) sequencing technique that generates long read lengths (on the order of 1000 bases) and a four-color sequencing trace.[39] A limit to throughput was imposed by the stochastic nature of immobilizing DNA polymerases at the bottom of each ZMW. In the published study, roughly one-third of the ZMWs in the array contained a single DNA polymerase and had the capacity to generate full-length sequencing reads. Figure 2 depicts the four-color SMRT sequencing strategy employed in this important article.

Following the proof-of-concept of SMRT sequencing study, PacBio streamlined the sequencing template construction by creating what they call a SMRTbell template.[40] The SMRTbell template allows consecutive sequencing of both the sense and antisense strand of a double-stranded DNA fragment by ligating universal hairpin loops to the ends of the fragment. Sample preparation time is decreased since template amplification is not needed, and DNA fragments over a broad size range can be used to generate SMRTbell templates. In the end, the use of the SMRTbell template increases the accuracy of sequencing and SNP detection.

PacBio now offers a commercially available sequencing instrument, the PacBio RS system. Consumables for this instrument include single-use ZMW arrays (called SMRT Cells) that contain 150,000 ZMWs and kits for SMRTbell template preparation. Recently, the PacBio RS Sequencer was used for the rapid genotyping of five *Vibrio cholerae* strains to determine the source of a cholera outbreak in Haiti.[41] Average read lengths for the five strains ranged from ∼700 to 1,000 bases, while the average depth of coverage ranged from 28 to 60, and the mean single-pass accuracy ranged from 81% to 83%. For three of the strains, read lengths approaching 3,000 bases were reported for a small percentage of the sequencing runs.

Besides sequencing, other applications are being developed using the SMRT detection technology. PacBio enhanced the robustness of genetic information generated by their single-molecule detection assay by correlating polymerase kinetic data to DNA methylation patterns during DNA sequencing.[42] The ability to sequence strands of mRNA at the level of codon resolution has been proven by simply substituting the DNA polymerase at the bottom of each ZMW with a ribosome and monitoring incorporation of fluorescently labeling tRNAs (tRNAs).[43]

SEQUENCING BY LIGATION

Complete Genomics. The basis for Complete Genomics' sequencing platform is centered on a hybridization and ligation method. While sequencing by hybridization and ligation has been around for some time,[44−49] the sample preparation and nanoarray platform developed by Complete Genomics is novel.[50] Sequencing fragments are prepared by sonication of genomic DNA followed by a series of repeated adapter site insertions, template circularization, and restriction enzyme scission. In the end, circularized sequencing fragments on the order of 400 bases are generated, each containing four distinct adapter sites. Circularized fragments are amplified by 2 orders of magnitude using

Table 2. Next-Generation Sequencing Technologies

Generation	Company	Platform Name	Method of Sequencing	Method of Detection	Read Length (bases)	Advantages	Limitations
Third	Pacific Biosciences	PacBio RS	Real-time, single-molecule DNA sequencing	Fluorescence/Optical	~1000	Long average read lengths; decreased sequencing time compared to first-gen platforms; no amplification of sequencing fragments; longest individual reads approach 3,000 bases	Inefficient loading of DNA polymerase in ZMWs; low single-pass accuracy (81−83%); degradation of the polymerase in ZMWs; overall, high cost per base (expensive instrument)
Third	Complete Genomics	In-house lab-built instrumentation	Combinatorial probe-anchor hybridization and ligation (cPAL)	Fluorescence/Optical	10	Highest (claimed) throughput of third-gen platforms; lowest reagent cost for reassembling a human genome of all sequencing technologies; each sequencing step is independent, minimizing accumulation of errors	Short read lengths; template preparation prevents sequencing through long repetitive regions; labor intensive sample preparation; no commercially available instrument
Third	Ion Torrent/Life Technologies	Personal Genome Machine (PGM) sequencer	Sequencing by synthesis	Change in pH detected by Ion-Sensitive Field Effect Transistors (ISFETs)	100−200	Direct measurement of nucleobase incorporation events; DNA synthesis reaction operates under natural conditions (no need for modified DNA bases)	Sequential washing steps can lead to accumulation of errors; potential difficulties in reading through highly repetitive or homopolymer regions of the genome
Fourth	Oxford Nanopore	gridION	Nanopore exonuclease sequencing	Current	Not yet quantified	Potential for long read lengths; low cost of αHL nanopore production; no fluorescent labeling or optics necessary	Cleaved nucleotides may be read in the wrong order; difficult to fabricate a device with multiple parallel pores

dx.doi.org/10.1021/ac2010857 |Anal. Chem. 2011, 83, 4327–4341

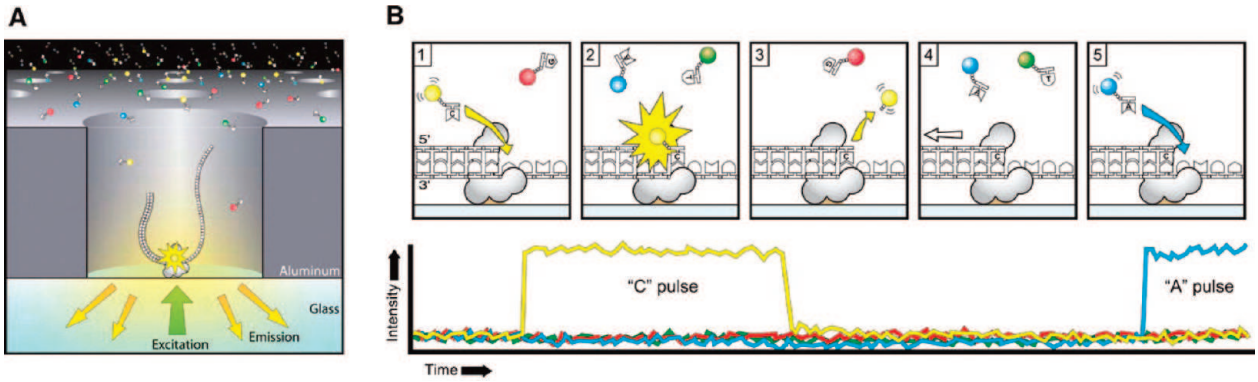

Figure 2. Schematic of PacBio's real-time single molecule sequencing. (A) The side view of a single ZMW nanostructure containing a single DNA polymerase (Φ29) bound to the bottom glass surface. The ZMW and the confocal imaging system allow fluorescence detection only at the bottom surface of each ZMW. (B) Representation of fluorescently labeled nucleotide substrate incorporation on to a sequencing template. The corresponding temporal fluorescence detection with respect to each of the five incorporation steps is shown below. Reprinted with permission from ref 39. Copyright 2009 American Association for the Advancement of Science.

Φ29 polymerase. Each amplified product of a circularized fragment is called a DNA nanoball (DNB). DNBs are selectively attached to a hexamethyldisilizane (HMDS) coated silicon chip that is photolithographically patterned with aminosilane active sites. Figure 3A illustrates the DNB array design.

The use of the DNBs coupled with the highly patterned array offers several advantages. The production of DNBs increases signal intensity by simply increasing the number of hybridization sites available for probing. Also, the size of the DNB is on the same length scale as the active site or "sticky" spot patterned on the chip, which results in attachment of one DNB per site. Since the active sites are spaced roughly 1 μm apart, up to three billion DNB can be fixed to a 1 in. by 3 in. silicon chip.[51] In addition to increasing the number of sequencing fragments per chip, the length scales of the size and spacing of the DNBs maximizes the pixel use in the detector. This highly efficient approach to generating a hybridization array results in decreased reagent costs and increased throughput compared to other second generation DNA sequencing arrays that have been used to sequence complete human genomes.[19,22,52]

Once the DNB array chip is generated, a library of forty common probes is used in combination with standard anchors and extended anchors to perform an unchained hybridization and ligation assay. The forty common probes consist of two subsets: probes that interrogate 5′ of the distinct adapter site in the DNB and probes that interrogate 3′ of the distinct adapter site in the DNB. In each subset, there are five sets of four common probes; each probe is 9 bases in length. Each set corresponds to positions 1 to 5 bases away from the distinct adapter sites in the sequencing substrate, and within each set, there are four distinct markers corresponding to each base. The standard anchors bind directly to the 5′ or 3′ end of the adapter site on the DNB and allow for hybridization and ligation of the common probes. The extended anchor scheme consists of ligation of a pair of oligo anchors (degenerate and standard) to expand the hybridized anchor region 5 bases beyond the adapter sites in the DNB and into the sequencing region. This combinatorial probe-anchor ligation (cPAL) method developed by Complete Genomics extends read lengths from 5 bases to 10 bases and results in a total of 62 to 70 bases sequenced per DNB. A schematic demonstrating both the standard anchor scheme and the extended anchor scheme is shown in Figure 3B.

Each hybridization and ligation cycle is followed by fluorescent imaging of the DNB spotted chip and subsequently regeneration of the DNBs with a formamide solution. This cycle is repeated until the entire combinatorial library of probes and anchors is examined. This formula of the use of unchained reads and regeneration of the sequencing fragment reduces reagent consumption and eliminates potential accumulation errors that can arise in other sequencing technologies that require close to completion of each sequencing reaction.[19,52,53]

Complete Genomics showcased their DNB array and cPAL technology by resequencing three human genomes and reported an average reagent cost of $4,400 per genome.[50] The three genome samples sequenced in this study (NA07022, NA19240, and NA20431) were then compared to previous sequence data.[54,55] The average coverage of these samples ranged from 45X to 87X, and the percent of genome identified ranged from 86% to 95%. While this technology greatly increases throughput compared to Sanger/CE and second-generation sequencing technologies, there are several drawbacks to Complete Genomics' approach. First, the construction of circular sequencing fragments results in an underrepresentation of certain genome regions, which leads to partial genome assembly downstream. Also, the size of the circular sequencing fragments (∼400 bases) as well as the very short read lengths (∼10 bases) prevents complete and accurate genome assembly, given that these fragments are shorter than a number of the long repetitive regions.

Just five months after Complete Genomics' proof-of-concept study was published, the first externally published application of Complete Genomic's sequencing technology was released. A group at the Institute for Systems Biology in Seattle, Washington (USA), studied the genetic differences in a human family of four.[56] In the study, whole-genome sequencing was used to determine four candidate genes responsible for two rare Mendelian disorders: Miller syndrome and primary ciliary dyskinesia. The subjects were a set of parents and their two children who both suffer from the disorders. This study highlighted the benefits of whole genome sequencing within a family when determining Mendelian traits. The ability to recognize inheritance patterns greatly reduced the genetic search space for recessive disorders and increased the sequencing accuracy. In the end, sequencing the entire family instead of just the two siblings affected by Miller syndrome and primary ciliary

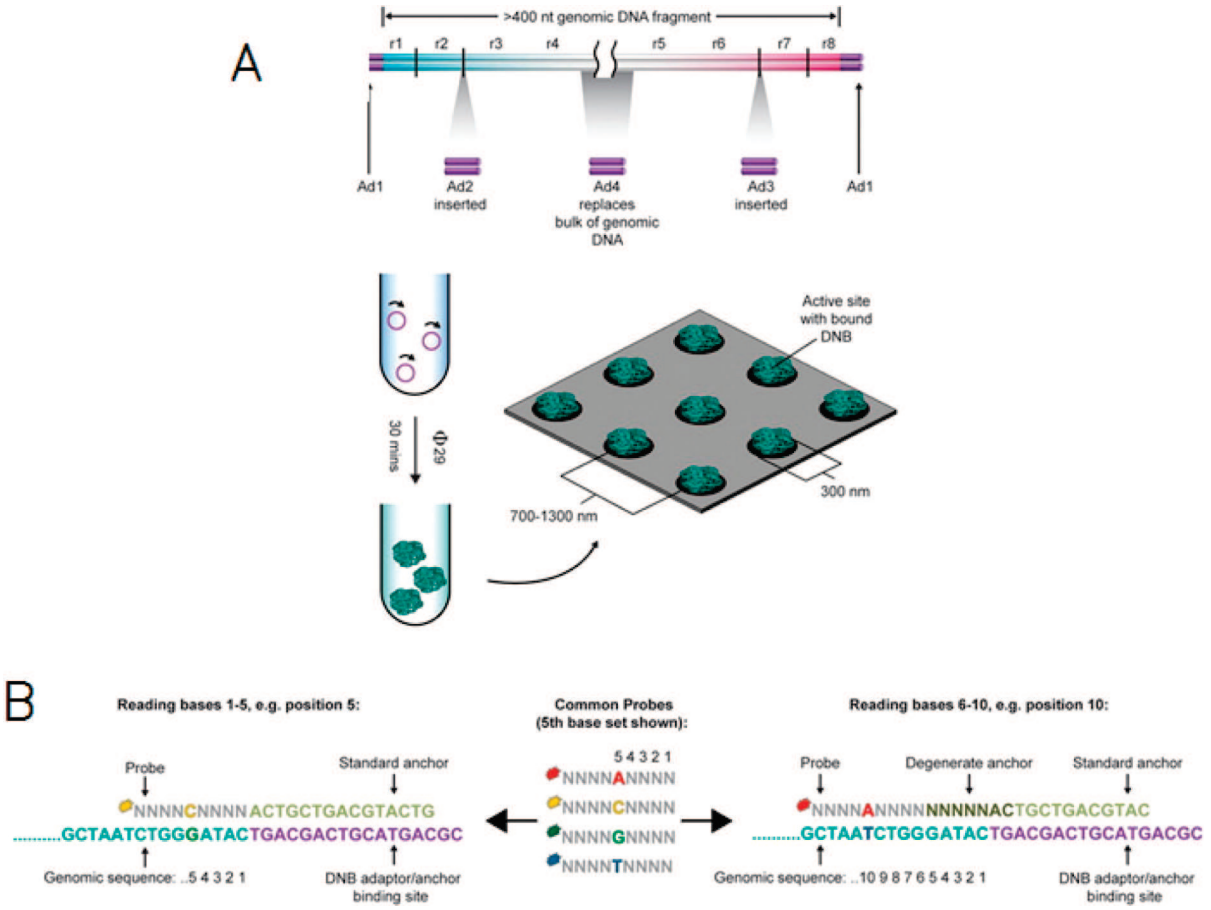

Figure 3. Schematic of Complete Genomics' DNB array generation and cPAL technology. (A) Design of sequencing fragments, subsequent DNB synthesis, and dimensions of the patterned nanoarray used to localize DNBs illustrate the DNB array formation. (B) Illustration of sequencing with a set of common probes corresponding to 5 bases from the distinct adapter site. Both standard and extended anchor schemes are shown. Reprinted with permission from ref 50. Copyright XXXX American Association for the Advancement of Science.

dyskinesia greatly decreased the number of false positive gene candidates which ultimately reduced the number gene candidates from 34 to just four.

Just one month later, the second externally published application of Complete Genomics's sequencing technology was released by a research group at Genetech.[57] The study compared the genome of primary lung tumor cells to that of adjacent normal tissue obtained from a 51-year-old Caucasian male who reported a heavy 15-year smoking history. When the complete genomes of different tissue samples were compared, over 50,000 single-base variations were identified and 530 previously reported single-base mutations were confirmed. Consequently, the importance of complete cancer genome analysis in understanding cancer evolution and treatment was brought to light due to the large number of single nucleotide mutations located outside of oncogenes as well as chromosomal structural variations found in the primary lung tumor.

A third application of the high-throughput cPAL method developed by Complete Genomics was published by a research group from the University of Texas Southwestern Medical Center in Dallas, Texas (USA).[58] This group used whole genome sequencing to diagnose a hypercholesterolemic 11-month-old girl with sitosterolemia after a series of blood tests and selective genetic sequencing were unable to confer a reasonable diagnosis. The gene and the subsequent mutations responsible for the sitosterolemia phenotype were determined after comparison of the patient's genome to a collection of reference genomes. Ultimately, it was determined that the patient failed the standard blood test due to low levels of plant sterols that were the result of a heavy diet of breast milk. This study illustrated the importance of whole genome sequencing in effectively diagnosing a disease in the presence of complex environmental factors that can influence standard assays.

SEQUENCING BY SYNTHESIS

The idea of sequencing by synthesis has been around for some time and is the basis for several second-generation sequencing technologies including Roche's 454 sequencing platform and Illumina's line of sequencing systems. 454's pyrosequencing method, which uses an enzyme cascade to produce light from a pyrophosphate released during nucleotide incorporation, was first piloted in the late 1980s and developed for DNA sequencing in the mid-1990s.[26,59−63] Illumina's fluorescently labeled sequencing by synthesis technique employs fluorescently labeled nucleotides with reversible termination chemistry and modified polymerases for improved incorporation of nucleotide analogues.[19] These sequencing by synthesis methods increased throughput compared to first-generation sequencing methods; however, optical imaging is needed to detect each sequencing

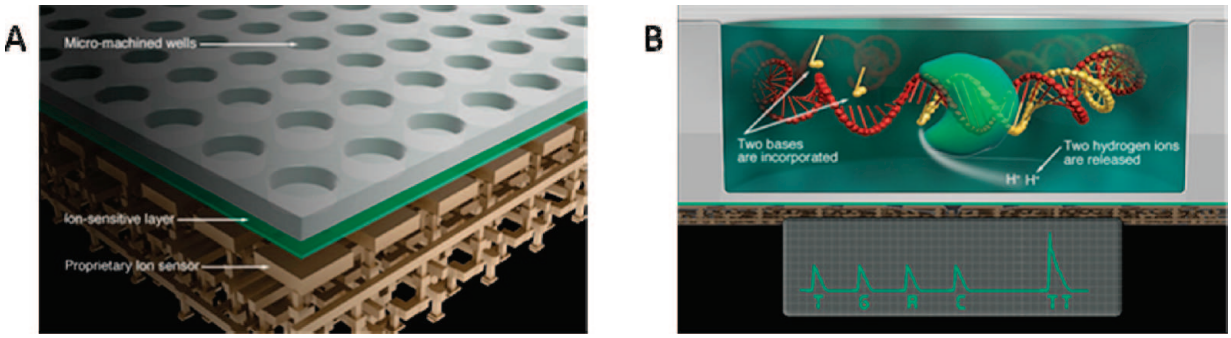

Figure 4. Layout of Ion Torrent's semiconductor sequencing chip technology. (A) A layer-by-layer view of the chip revealing the structural design. The top layer contains the individual DNA polymerization reaction wells, and the bottom two layers comprise the FET ion sensor. Each well has a corresponding FET detector that identifies a change in pH. (B) A side view of an individual reaction well depicting DNA polymerase incorporation of a repeat of two TTP nucleotides on a sequencing fragment. The hydrogen ions released during this process are detected by the FET below. Reprinted with permission from Ion Torrent (Wes Conrad).

step. Since an intricate optics system can increase the overall cost of a sequencing system, the next logical advancement in the sequencing field has been to abandon the use of optics for a less expensive approach to detection.

When this is taken into account, research in the Pease and Davis laboratories at Stanford University evolved from earlier pyrosequencing technology by proposing a new method of detection to measure temperature or pH change in microstructures.[64−66] Since both changes are byproducts of nucleotide incorporation in a DNA polymerization reaction, the need for optical detection of light produced by the luciferase enzyme was eliminated. Like pyrosequencing, this thermosequencing method requires sequential cycles in which one of the four nucleotides is introduced to the system, followed by measurement of nucleotide incorporation by heat detection. Between each cycle, the system is regenerated by thorough washing of reaction wells to minimize residual NTPs and, therefore, reduce error accumulation. This innovative detection scheme led to the startup of the company Genapsys from the Stanford Genome Technology Center. Their thermal detection method has an inherent advantage over pH detection, in that temperature can be reset quickly by conduction from a cooling block, while hydrogen ions must be washed away. Ion Torrent, a startup recently acquired by Life Technologies,[67] has made significant progress in bringing to market a next-generation sequencing system that utilizes pH changes to detect base incorporation events.

Ion Torrent. According to Ion Torrent's patent applications,[68,69] field-effect transistors (FETs) are used to measure a change in pH in a microwell structure (see Figure 4). To increase throughput, the Ion Torrent sequencing chip makes use of a highly dense microwell array in which each well acts as an individual DNA polymerization reaction chamber containing a DNA polymerase and a sequencing fragment. Just below this layer of microwells is an ion-sensitive layer followed by a sublayer composed of a highly dense FET array aligned with the microwell array. Following the pyrosequencing scheme, sequential cycling of the four nucleotides into the microwells enables primary sequence resolution since the FET detector senses the change in pH created during nucleotide incorporation and converts this signal to a recordable voltage change. Since the change in voltage scales with the number of nucleotides incorporated at each step, Ion Torrent's sequencing chip has an inherent capacity to call repeats.

At present, Ion Torrent offers the one-time-use Ion 314 sequencing chip (Figure 4); however, within the next year, they are scheduled to release their second- and third-generation chips: the Ion 316 sequencing chip and the Ion 318 sequencing chip (Ion Torrent Application Note, Spring 2011). The 1.2 million microwells on the Ion 314 chip generates roughly 10 Mb of sequence information with average read lengths on the order of a 100 bases. To further increase throughput, the Ion 316 chip and the Ion 318 chip are being built with 6.2 million and 11.1 million microwells, respectively. The expectations for the Ion 318 chip are to produce 1 Gb of sequencing data with average read lengths of 200 bases or higher. Ultimately, Ion Torrent seeks to "democratize" sequencing by offering the first reasonably priced (∼$50K) benchtop-scale, high-throughput sequencing machine.[70]

While this newly developed method of ion sensing-based sequencing by synthesis offers great potential to reduce the cost of sequencing, there are several limitations with regards to sequencing complete genomes. Currently, the short read lengths place a large burden on the reassembly process and limit the assembly of de novo sequencing projects due to an inability to read through long repetitive regions in the genome. Also, due to the sequential nature of this sequencing by synthesis method, error accumulation can occur if reaction wells are not properly purged between reaction steps. Finally, as for pyrosequencing in the previous generation, sequencing through smaller repetitive regions of the same nucleotide (homopolymer regions) on the order of 5 to 10 bases can prove challenging. Ion Torrent has reported sequencing accuracy data in which an *E. coli* DH10B sample was sequenced and homopolymer regions were analyzed (Ion Torrent Application Note, Spring 2011). The sequencing accuracy for a 5-mer homopolymer region was shown to be around 97.5%; however, it was difficult to tell the size of the sample set from which these data were generated. Also, accuracy data for homopolymer lengths greater than 5 bases were not reported.

■ NANOPORE SEQUENCING TECHNOLOGIES

A fundamentally different class of sequencing technology is based on nanopore structures, described in prior reviews by Branton[71] and Bayley.[72] Individual base detection was envisioned to be possible through the measurement of conductivity either across or through a membrane, via a nanoscale pore. These nanopores consist of an orifice slightly larger than the width a double-stranded DNA molecule, which is 4 nm, where DNA is threaded through the pore. The chemical differences of each base

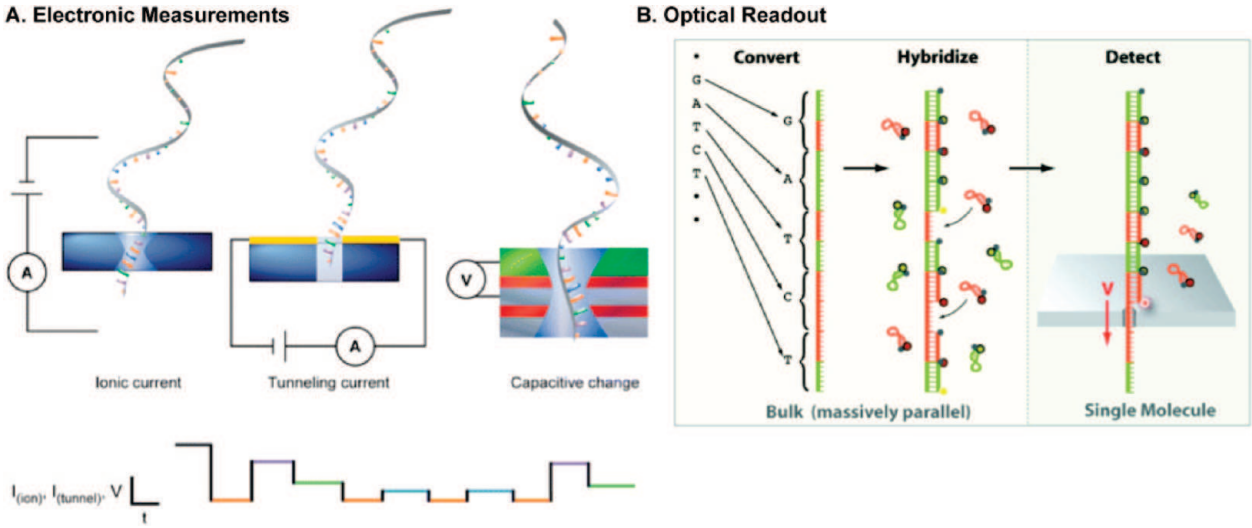

Figure 5. Nanopore DNA sequencing using electronic measurements and optical readout as detection methods. (A) In electronic nanopore schemes, signal is obtained through ionic current,[73] tunneling current,[78] and voltage difference[79] measurements. Each method must produce a characteristic signal to differentiate the four DNA bases. Reprinted with permission from ref 83. Copyright 2008 Annual Reviews. (B) In the optical readout nanopore design, each nucleotide is converted to a preset oligonucleotide sequence and hybridized with labeled markers that are detected during translocation of the DNA fragment through the nanopore. Reprinted from ref 82. Copyright 2010 American Chemical Society.

would result, in theory, in detectably altered current flow through the pore. Theoretically, nanopores could also be designed to measure tunneling current across the pore as bases, each with a distinct tunneling potential, could be read. The nanopore approach, while still in development, remains an interesting potential fourth-generation technology. This "fourth-generation" moniker is suggested, since optical detection is eliminated along with the requirement for synchronous reagent wash steps.

Nanopore technologies may be broadly categorized into two types, biological and solid-state. The protein alpha hemolysin, which natively bridges cellular membranes causing lysis, was first used as a model biological nanopore. The protein was inserted into a biolayer membrane separating two chambers while sensitive electronics measured the blockade current, which changed as DNA molecules moved through the pore. However, chemical and physical similarities between the four nucleotides made the sequence much more difficult to read than envisioned. Further, sufficient reduction of electronic noise remains a constant challenge, which is achievable in part by slowing the rate of DNA translocation. Recently, Oxford Nanopore and several other academic groups working on this concept have made progress toward addressing these challenges.

The second class is based on the use of nanopores fabricated mechanically in silicon or other derivative. The use of these synthetic nanopores alleviates the difficulties of membrane stability and protein positioning that accompany the biological nanopore system Oxford Nanopore has established. For example, Nabsys created a system using a silicon wafer drilled with nanopores using a focused ion beam (FIB), which detects differences in blockade current as single-stranded DNA bound with specific primers passes through the pore. IBM created a more complex device that aims to actively pause DNA translocation and interrogate each base for tunneling current during the pause step. The technology of both of these nanopore types are presented in detail below.

John Kasianowicz and colleagues[73] were the first to show the translocation of polynucleotides (poly[U]) through a *Staphylococcal* α-hemolysin (αHL) biological nanopore suspended in a lipid bilayer, using ionic current blockage method. The authors predicted that single nucleotides could be discriminated as long as: (1) each nucleotide produces a unique signal signature; (2) the nanopore possesses proper aperture geometry to accommodate one nucleotide at a time; (3) the current measurements have sufficient resolution to detect the rate of strand translocation; (4) the fragment should translocate in a single direction when potential is applied; and (5) the nanopore/supporting membrane assembly should be sufficiently robust. All of the biological and synthetic nanopores have barrels of ∼5 nm (which is considerably longer than the base-to-base distance of 3.4 Å) in thickness and accommodate ∼10−15 nucleotides at a time. It is, therefore, impossible to achieve single-base resolution using blockage current measurements. In addition, the average rate at which a polymer typically translocates through a nanopore is on the order of 1 nucleotide/μs (i.e., on the order of MHz detection), which is too fast to resolve. The nucleotide strand should be slowed down to ∼1 nucleotide/ms to allow for a pA-current signal at 120−150 mV applied potential.[74] Furthermore, the translocation of a polymer strand should be uniform between two events. The time distribution of two processes (capture, entry, and translocation) is nonPoisson and often differs by an order of magnitude. This means that two molecules pass through a nanopore at considerably different rates and the slower one could be missed or misinterpreted. Andre Marziali and coworkers at UBC[75,76] used force spectroscopy to study these events through single-molecule bond characterization. The nonuniform kinetics of DNA passage through an αHL nanopore is attributed to weak binding of DNA to amino acid residues of the protein nanopore.[77]

Because of these challenges with ionic current measurements (the current created by the flow of ions through the nanopore), researchers have looked at other measurement schemes such as the detection of tunneling current and capacitance changes (Figure 5A). In transverse tunneling current scheme, electrodes are positioned at the pore opening and the signal is detected from subnanometer probes.[78] In capacitance measurements, voltage is

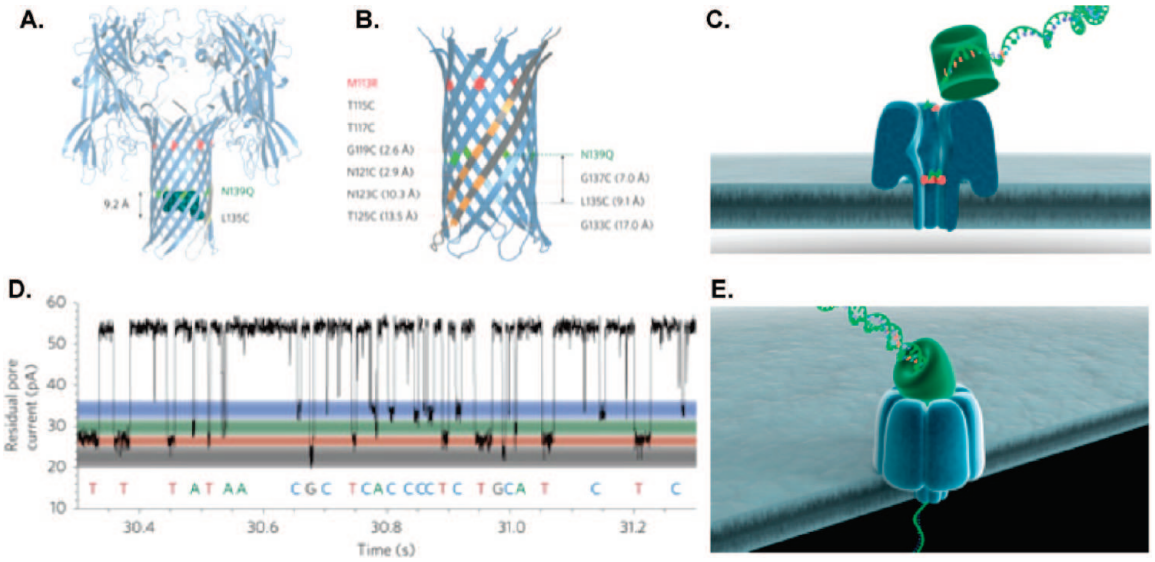

Figure 6. Biological nanopore scheme employed by Oxford Nanopore. (A) Schematic of αHL protein nanopore mutant depicting the positions of the cyclodextrin (at residue 135) and glutamines (at residue 139). (B) A detailed view of the β barrel of the mutant nanopore shows the locations of the arginines (at residue 113) and the cysteines. (C) Exonuclease sequencing: A processive enzyme is attached to the top of the nanopore to cleave single nucleotides from the target DNA strand and pass them through the nanopore. (D) A residual current-vs-time signal trace from an αHL protein nanopore that shows a clear discrimination between single bases (dGMP, dTMP, dAMP, and dCMP). (E) Strand sequencing: ssDNA is threaded through a protein nanopore and individual bases are identified, as the strand remains intact. Panels A, B, and D reprinted with permission from ref 91. Copyright 2009 Nature Publishing Group. Panels C and E reprinted with permission from Oxford Nanopore Technologies (Zoe McDougall).

detected across a metal oxide-silicon layered structure. The voltage signal is induced across the capacitor by the passage of charged nucleotides in longitudinal direction.[79] A different readout approach is optical detection (Figure 5B). A typical optical recognition of nucleotides is essentially executed in two steps. First, each base (A, C, G, or T) in the target sequence is converted into a sequence of oligonucleotides, which are then hybridized to two-color molecular beacons (with fluorophores attached).[80] Because the four nucleotides (A, C, G, or T) have to be determined, the two fluorescent probes are coupled in pairs to uniquely define each base. For example, if the two probes are A and B, the four unique permutations will be AA, AB, BA, and BB. As the hybridized DNA strand is threaded through the nanopore, the fluorescent tag is stripped off from its quencher and an optical signal is detected. Both protein[81] and solid-state nanopores can be used.[80,82] Detailed electronic measurement schemes and optical readout methods have been reviewed in more detail in previously published papers.[71,83,84]

In a 2008 review article,[71] Daniel Branton and colleagues discussed the nanopore sequencing development and the prospect of low sample preparation cost at high throughput. They estimated that purified genomic DNA sufficient for sequencing ($\sim 10^8$ copies or 700 μg) could be extracted and purified from blood at a cost of less than $40/sample using commercial kits. All existing sequencing techniques require breaking the DNA into small fragments of ~100 bps and sequencing those chunks multiple times to find overlapping regions, so that they can be reassembled together. Because one of most appealing advantages of nanopores is achieving long read lengths, the genomic assembly process should be considerably simplified. In practice, the read length may be limited only by the DNA shearing that occurs during pipetting in the sample preparation step. For example, Meller and Branton[85] demonstrated that 25 kb ssDNA could be threaded through a biological nanopore and 5.4 kb ssDNA translocated through a solid-state nanopore. In addition, several groups have shown very high throughput of small oligonucleotides (~5.8 oligomers (s μM)$^{-1}$)[85] and native ssDNA and dsDNA (~3−10 kb at 10−20 nM concentration).[86,87]

Protein Nanopore Sequencing. Oxford Nanopore Technologies, formerly Oxford Nanolabs, together with leading academic collaborators, has addressed some of the aforementioned technological challenges and implemented the nanopore technology in a commercial product (GridION system). Oxford Nanopore, founded by Prof. Hagan Bayley at University of Oxford, aimed at commercializing the research work on biological nanopores coming out of his laboratory. The company works in collaboration with Professors Daniel Branton, George Church, and Jene Golovchenko at Harvard; David Deamer and Mark Akeson at UCSC; and John Kasianowicz at NIST.

Recently, Gordon Sanghera, CEO of Oxford Nanopore, announced that the company is preparing to launch the GridION system[88] for direct single-molecule analysis, which would adopt exonuclease sequencing. The system is based on "lab on a chip" technology and integrates multiple electronic cartridges into a rack-like device. A single protein nanopore is integrated in a lipid bilayer across the top of a microwell, equipped with electrodes. Multiple microwells are incorporated onto an array chip, and each cartridge holds a single chip with integrated fluidics and electronics for sample preparation, detection, and analysis. The sample is introduced into the cartridge, which is then inserted in an instrument called a GridION node. Each node can be used separately or in a cluster, and all nodes communicate with each other and with the user's network and storage system in real time. Although the main application of the platform is sequencing of DNA, it can be adapted (by proper modification of the αHL nanopore) for the detection of proteins and small molecules.

Oxford Nanopore's first-generation systems utilize the heptameric protein α-hemolysin (αHL) (Figure 6A).[72,89−91] αHL is

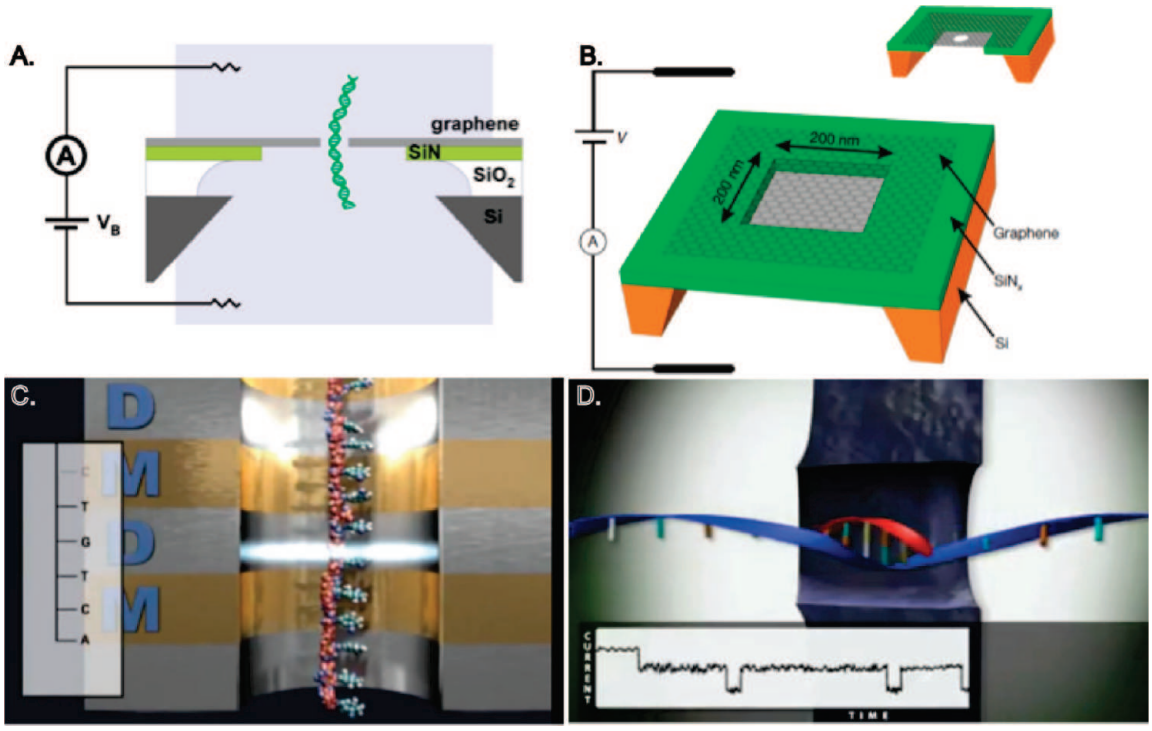

Figure 7. Several synthetic nanopore sequencing device designs. (A) The device consists of 1−5 nm thick graphene membrane which is suspended in a Si chip coated with 5 μm SiO_2 layer. It is placed in a PDMS cell with microfluidic channels on both sides of the chip. Reprinted from ref 96. Copyright 2010 American Chemical Society. (B) A nanopore (shown in the inset to the figure) is drilled through a graphene membrane, which is suspended in SiN_x across a Si frame. The graphene membrane separates two ionic solutions and is in contact with Ag/AgCl electrodes. Reprinted with permission from ref 97. Copyright 2010 Nature Publishing Group. (C) IBM DNA transistor setup. A nanometer sized pore is fabricated using an electron beam. Electric field is created between the gated regions allowing for charge trapping. The substrate is composed of metal and dielectric regions, labeled with "M" and "D", respectively. Reprinted with permission from IBM (Gustavo Stolovitzky). (D) HANS method adopted by NABsys for electronic readout of DNA fragments through solid-state nanopores. 6-mer oligonucleotide probes are hybridized to ssDNA fragments, and current-verses-time trace is detected. Reprinted with permission from NABsys (John Oliver).

secreted from bacteria, providing low-cost production of these robust bionanopores. Oxford Nanopore is working toward commercialization of two types of sequencing methodologies: exonuclease sequencing and strand sequencing. In the exonuclease method,[92] a cyclodextrin adapter molecule (Figure 6B) is bound to the inside of a protein nanopore and serves as a DNA binding site. The nanopore is additionally coupled with an exonuclease,[93] a processive enzyme that cleaves individual bases from the DNA strand and allows for accurate detection as DNA bases pass through and interact with the cyclodextrin (Figure 6C). Progressive enzymes positioned on the top of the nanopore regulate the translocation rate of the DNA strand by slowing down (to the order of ms) the intrinsic electrophoretic motion (to the order of μs).[93] Essentially, one nucleotide passes through the nanopore approximately every 20 ms, which is slow enough for accurate detection. The four nucleotides produce different magnitudes of current disruption (Figure 6D) and, therefore, the determination of DNA sequence is possible. Assuming a steady 1 ms per base sequencing rate, a single pore would require 69 days to process 6 billion bases. 100,000 pores operating perfectly at that rate could theoretically sequence a genome with 30× coverage in 30 min.

Oxford Nanopore is also working toward the development of a strand sequencing technology, in which a single-stranded DNA fragment is passed through the pore and identification of single bases is achieved as they pass through the pore (Figure 6E).[94] This method is potentially faster and more accurate than exonuclease sequencing. Because all nucleotides are attached to each other, there is no chance of reading them in the wrong order; however, the challenge lies in achieving the accurate identification of the individual bases as they pass through the nanopore.

Solid-State Nanopore Sequencing. Although the αHL heptamer pores are rather robust,[95] the lipid bilayers on which biopores are suspended are unstable and hard to manipulate. Solid-state or man-made nanopores are considered to be next-generation nanopore technology, because they bypass the use of organic supports and are thus, in principle, more stable. Also solid-state nanopores could be multiplexed to work in parallel on a single device, which is a challenge for biological nanopores. Artificial pores are fabricated in solid-state materials such as silicon nitride, silicon or metal oxides, and more recently graphene. Graphene is a new, single-atom thick material which is known to be the thinnest possible membrane. A group at the University of Pennsylvania led by Marija Drndic[96] presented translocation measurements of DNA through graphene nanopores, which comprised 1−5 nm thick membranes with 5−10 nm diameter pores (Figure 7A). In another publication,[97] researchers in the Golovchenko lab at Harvard showed that a graphene sheet can be used as a membrane material that holds a solid-state nanopore and separates two chambers of ionic solution (Figure 7B).

A novel approach to DNA sequencing through artificial nanopores in solid-state material (specifically a metal-dielectric layered structure) is currently being developed by IBM, in

collaboration with 454 Life Sciences, which is part of Roche. The idea originated with systems biologist Gustavo Stolovitzky and electrical engineer Stanislav Polonsky at IBM in 2006. Three nanometer artificial nanopores are fabricated by e-beam drilling in 10 nm thick membranes made of titanium nitride, separated by insulating layers of silica. As the DNA strand is drawn through the nanopore, the electrical field across the metal layers can be flipped, also referred to as the ratchet effect, resulting in immobilization and, subsequently, in principle, controlled motion of the DNA strand (Figure 7C). Alternation of the electric field can be potentially beneficial for improving sequencing accuracy. Two possibilities for detecting the signal are capacitance or ionic current measurements (similar to Oxford Nanopore detection, except that the DNA strand will remain intact). To obtain a strong enough signal, the DNA needs to be trapped for interrogation only for a millisecond. Most of the work reported by the IBM group has been numerical, so far, through MD simulations.[98,99] Although 5 to 7 years of further development is expected to be required[27] before any commercial release, the idea of electronic detection combined with easy sample preparation offers the exciting potential for very cheap sequence readout.

Despite the challenges of achieving single-base resolution by means of current blockage measurements through a man-made nanopore, a number of groups have easily distinguished a translocation of ssDNA from dsDNA in a nanopore wide enough to accommodate the double strand.[87,100,101] Because coarse-grained resolution is easily achieved, researchers started looking into the creation of de novo sequencing techniques, by attaching hybridization probes to DNA fragments. Recently, Balagurusamy et al.[100] experimentally showed a translocation and successful electrical detection of two consecutive 12-mer long double strands through a nanopore drilled in a 20 nm thick silicon nitride membrane. Another solid-state nanopore study[101] reported the detection of dsDNA hybridized with peptide nucleic acid (PNA) probes threaded through a sub-5 nm nanopore in 30 nm thick membrane. These studies are promising for the realization of the sequencing by hybridization[102] (SBH) through nanopores, also known as the hybridization-assisted nanopore sequencing[103] (HANS) method. This technology has been licensed by NABsys, a DNA sequencing startup company founded by Brown physics professor Sean Ling in 2005. The company is working toward the development and commercialization of a computer chip to "electronically read" DNA. In practice, 6-mer hybridization probes will be bound to 100-kb genomic fragments, which are then driven electrophoretically through a solid-state nanopore, creating a current signal (Figure 7D). On the basis of this current tracing, the positioning of the probes and, thus, the sequence of small fragments is determined. The process is done in parallel with an entire library of probes, which will in principle allow the assembly of the whole genome length and readout. The company promises an eventual 4 orders of magnitude decrease in the cost of whole genome sequencing. However, an open question for the HANS technology is the sufficient resolution for accurate readout of the electronic signal.

■ LONG READ DNA EXTENSION METHODS

While short-read methods that rely on DNA fragments that are less than 400 bases long constitute the bulk of the current DNA sequencing technologies, several different, new approaches aim to sequence DNA up to several megabases in length. Recent reports have highlighted the limitations of short-read technologies for genome assembly of prokaryotes.[104] Mapping of more extended DNA regions can provide data on the number of repeats, deletions, insertions, and transpositions that are unobtainable with any of the currently available short-read methods.

Final Assembly by Optical Mapping. At the University of Wisconsin-Madison, Prof. David C. Schwartz and colleagues have created the only system (Optical Mapping) available to date with the capacity to strategically guide, validate, and complete the sequence assembly of whole, complex genomes. The Optical Mapping System constructs genome-wide, long-range, ordered restriction maps from large data sets comprising 5,000−2,000,000 individual genomic DNA molecules (∼500 kb), "bar-coded" by restriction digestion and directly imaged by fluorescence microscopy. This highly automated system is the first single-molecule platform proven capable of whole genome analysis.[105] The Optical Mapping System boasts computational tools that include alignment capabilities,[106] which position nascent sequence assemblies onto de novo optical maps[107] spanning entire genomes. Such alignments place orphan sequence assemblies, order and orient scaffolds and contigs, size gaps, and reveal assembly errors, in addition to accurate accounting of chromosome number and sizes. Early applications of Optical Mapping have centered on bacterial[108] and lower eukaryotic genomes;[109,110] however, more recently, Optical Mapping analysis has successfully guided the assembly and validation of complex genomes that included rice[111] and maize,[112−115] which are the most complex genomes ever sequenced. Because very large ∼500 kb genomic DNA molecules are analyzed, complex genomic regions near centromeres, or those rife with segmental duplications become measurable to reveal new structural variants not approachable by sequencing. This advantage allows discovery of many new structural variants as insertions or complex rearrangements within human genomes[116,117] that confound sequencing approaches and portend significant analytical approaches for dissecting breakpoints and rearrangements in cancer genomes (unpublished, Schwartz, et al.).

The Schwartz laboratory has further advanced genome mapping approaches through the addition of sequence reads to long double-stranded molecules and the development of the Nanocoding System.[118] Nanocoding uses genomic analytes, and within a single reaction mix, nicking restriction enzymes selectively clip only one strand of the double stranded DNA, at cognate recognition sites. Newly created nicks are then tagged by polymerase-mediated nick translation using fluorochrome labeled nucleotides. Unique single-molecule barcodes emerge because the end products are full-length dsDNA molecules precisely decorated by fluorochromes at each enzyme recognition site. Decorated DNA molecules are loaded into a microfluidic chip to flow into channels on the order of 50 μm in size. These channels are bisected at a 45° angle by nanofluidic channels 1 μm wide and 100 nm deep. The combination of the microfluidic−nanofluidic channel angle and the nanofluidic channel width significantly reduce the entropic penalty required to fully stretch the DNA from a coiled form, while low ionic strength buffers greatly enhance molecular stretching within the nanoslits. Once stretched in the channel, fluorescence imaging [FRET (fluorescence resonance energy transfer), intercalated YOYO-1 dye (Donor), and Alexa Fluor 647 (acceptor)] and machine vision identify the locations of covalently incorporated fluorochromes for the construction of single-molecule bar-codes that are assembled into genome-wide physical maps.

A second company using nanofluidics, BioNanomatrix, was established with technology licensed from Princeton University. Their design also uses nanofluidic channels to stretch the DNA with a modified channel entrance design. The channels are on the order of 100 nm or less in width as well as in depth. To overcome the entropic barrier to entry, the channel transitions from micrometers deep to nanometers deep using a lithography pattern to introduce a pillar type pattern that gradually forces the DNA to uncoil and extend into the nanochannels for imaging.[56−62] These chips may also contain constrictions to force the DNA through a narrow gap. The BioNanomatrix chip has been used with formamide and controlled localized heating in the presence of YOYO-1 to partially denature the DNA and infer a sequence from a pattern.[119] A second technique identified landmarks on λ-DNA using a nicking enzyme to displace a recognition site on backfill with nucleotides and, subsequently, hybridize the displaced strand with a fluorophore-labeled probe.[120] A camera and imaging software were used for analysis. Of the 300 molecules imaged in 30 s, 85% of the two targeted sites were properly labeled.

Nonoptical, Stretched DNA Molecule Methods. The methods discussed here still stretch DNA over a surface, which is probed to read each individual base, but eschews the use of a video camera completely, for atomic imaging methods. Halcyon Molecular is a fourth-generation technology that relies on a rapid-scan tunneling electron microscope (TEM) method.[63] Individual DNA bases are labeled with distinct heavy atoms to differentiate between each base as described generally here.[121] ZS Genetics, where ZS stands for "Zero Science", has also pursued a TEM method but has yet to publish detailed methods or results. Scanning tunneling microscopy (STM) was reportedly used to identify guanine from nonguanine bases.[122] STM measures the density of electron flow through a scanning tip. Although 140 bases were read and compared to a reference gene sequence, a number of limitations, most notably speed, currently prevents commercial viability. A recent review[123] describes these techniques in greater detail.

CONCLUDING REMARKS

Technology and funding in the field of novel DNA sequencing technologies have been growing at a rate never before seen. As discussed in this Review, there has been a proliferation of vastly different approaches to DNA sequencing, across all generations of the newer technologies. Each technique has its own advantages and limitations; so, ultimately, the specific genotyping application must be evaluated in order to choose the appropriate sequencing platform. While second- and third-generation platforms boast considerable throughput, Sanger-based CE sequencing is still the gold standard for ultrahigh-accuracy sequencing and is the only technique that has so far provided both de novo sequencing and the de novo assembly of a human genome. In order to gain widespread recognition as the front-running next-generation sequencing technology, one of the second- or third-generation platforms must provide a side-by-side study with a first-generation CE-based platform and quantitatively compare the outcomes of the sequencing and assembly of the same de novo sample. This will provide concrete evidence of the true cost of de novo sequencing and will serve as the jumping off point from which current and future researchers can make decisions on how to tackle the next wave of human genome sequencing projects or the de novo sequencing of similarly sized complex genomes. Currently, on the basis of the current limitations of sequencing technologies, it appears that several of these technologies must be used in tandem to achieve the benefits of high-throughput, accuracy, long contiguous read lengths, and long-range mapping that would be needed to catalog such a complete complex genome, de novo.

AUTHOR INFORMATION

Corresponding Author

*Address: Stanford University, Department of Bioengineering, W300B James H. Clark Center, 318 Campus Drive, Stanford, CA 94305-5444. E-mail: aebarron@stanford.edu. Phone: (650) 796-4002. Fax: (650) 498-8004.

BIOGRAPHIES

Thomas P. Niedringhaus studied Chemical Engineering at Vanderbilt University where he earned a B.E. summa cum laude in 2005. Currently, he is a Ph.D. candidate in the Department of Chemical Engineering at Stanford University under the supervision of Professor Annelise E. Barron and Professor Eric Shaqfeh. His research focuses on understanding both physical and chemical properties of polymer solutions with applications in microchannel electrophoresis of DNA.

Denitsa Milanova received her MS in Mechanical Engineering at Stanford University in 2010 and her B.S. (summa cum laude) from University of Central Florida in 2008. She is a Ph.D. candidate at Stanford University and works under the supervision of Professor Annelise E. Barron and Professor Juan G. Santiago. Denitsa currently studies electrokinetic transport of DNA and small molecules in gel-free solutions in nanochannel devices.

Matthew B. Kerby holds a Ph.D. in Biomedical Engineering and is licensed in California as a Professional Chemical Engineer (PE). He has over 10 years of industrial research and development experience in the fields of microfluidics, instrumentation, and biomolecular analysis and a further 7 years in academic research. In 1993, he earned a B.S. in Chemical Engineering from the University of California, Davis. He obtained a M.S. in Biomedical Engineering from CSU Sacramento after creating a mathematical model of diabetic severity, while simultaneously developing software systems for an electronics company. In 1998, he created ultrasound contrast agent at Point Biomedical and, in 2000, integrated microfluidic tools at Caliper Technologies for pharmaceutical screening. He founded his own consulting company in 2003 engineering cell transfection and lyophilization systems. In 2004, he joined the Chemical and Biochemical Engineering laboratory of Dr. Anubhav Tripathi, an expert in fluid dynamics, at Brown University as a Ph.D. student. His thesis work integrated quantitative analysis of kinetics, microfluidics, molecular biology, and viral diagnostics. He received a Ph.D. from Brown University in Biomedical Engineering in 2007 before moving to Stanford. He currently holds an appointment as a postdoctoral scholar at Stanford University in the lab of Dr. Annelise Barron in the Bioengineering department. He works on fourth generation DNA sequencing and integrated fluidic systems for stem cells.

Michael Snyder is the Stanford Ascherman Professor and Chair of Genetics and the Director of the Center of Genomics and Personalized Medicine. Dr. Snyder received his Ph.D. training at the California Institute of Technology and carried out postdoctoral training at Stanford University. He is a leader in

the field of functional genomics and proteomics. His laboratory study was the first to perform a large-scale functional genomics project in any organism and, currently, carries out a variety of projects in the areas of genomics and proteomics both in yeast and humans. These include the large-scale analysis of proteins using protein microarrays and the global mapping of the binding sites of chromosomal proteins. His laboratory built the first proteome chip for any organism and the first high resolution tiling array for the entire human genome.

Annelise E. Barron studied chemical engineering at the University of Washington in Seattle, earning a B.S. cum laude in 1990, and the University of California, Berkeley, earning her Ph.D. in 1995. After postdoctoral research stints at Soane (later ACLARA) Biosciences and collaboratively at UCSF and the Chiron Corp. in Emeryville, CA, she joined the chemical and biological engineering faculty at Northwestern University in Evanston, IL, earning tenure in 2003 and full professorship in 2006. In August 2007, she moved to Stanford University, where she is now the W.M. Keck associate professor of bioengineering. Dr. Barron's research merges the disciplines of polymer science and hydrogel engineering with applications of novel materials and bioconjugates in biotechnology and medicine. The physical properties of novel polymers and gels are correlated with performance for a given application, to inform optimization and elucidate mechanism of action. Dr. Barron started work in the field of capillary electrophoresis research in 1990 and continues to develop innovative materials and strategies to enhance bioanalytical separations on microfluidic devices. She has served as Deputy Editor for *Electrophoresis* and organized Microscale Bioseparations 2011 (San Diego).

REFERENCES

(1) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463–5467.
(2) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (2), 560–564.
(3) Sanger, F.; et al. *Nature* **1977**, *265* (5596), 687–695.
(4) Ansorge, W. J. *New Biotechnol.* **2009**, *25* (4), 195–203.
(5) Franca, L. T.; Carrilho, E.; Kist, T. B. *Q. Rev. Biophys.* **2002**, *35* (2), 169–200.
(6) Slatko, B. E.; Albright, L. M.; Tabor, S.; Ju, J. DNA sequencing by the dideoxy method. In *Current Protocols in Molecular Biology*; Need Publisher: Need Place of publication, 2001.
(7) Lander, E. S.; et al. *Nature* **2001**, *409* (6822), 860–921.
(8) Venter, J. C.; et al. *Science* **2001**, *291* (5507), 1304.
(9) Conti, R.; et al. *Med. Decis. Making* **2010**, *30* (3) 328−340.
(10) Ginsburg, G. S.; Donahue, M. P.; Newby, L. K. *J. Am. Coll. Cardiol.* **2005**, *46* (9), 1615–1627.
(11) Ginsburg, G. S.; Willard, H. F. *Transl. Res.* **2009**, *154* (6), 277–287.
(12) Philippidis, A. After a Decade, JGI Retires the Last of Its Sanger Sequencers. *GenomeWeb Daily News*; GenomeWeb LLC: New York, 2010.
(13) Hert, D. G.; Fredlake, C. P.; Barron, A. E. *Electrophoresis* **2008**, *29* (23), 4618–4626.
(14) Shendure, J.; Ji, H. *Nat. Biotechnol.* **2008**, *26* (10), 1135–1145.
(15) Metzker, M. L. *Nat Rev Genet* **2010**, *11* (1), 31–46.
(16) Fuller, C. W.; et al. *Nat. Biotechnol.* **2009**, *27* (11), 1013–1023.
(17) Levy, S.; et al. *PloS Biology* **2007**, *5* (10), 2113–2144.
(18) Wheeler, D. A.; et al. *Nature* **2008**, *452*, 872–877.
(19) Bentley, D. R.; et al. *Nature* **2008**, *456*, 53–59.
(20) Ley, T. J.; et al. *Nature* **2008**, *456*, 66–72.
(21) Wang, J.; et al. *Nature* **2008**, *456*, 60–65.
(22) Pushkarev, D.; Neff, N. F.; Quake, S. R. *Nat. Biotechnol.* **2009**, *27* (9), 847–850.
(23) Herper, M. A First: Diagnosis By DNA. In *Forbes*; Forbes: New York, 2010
(24) Reporter, P. In *The Daily Scan*; GenomeWeb LLC: New York, 2010.
(25) Karow, J. Life Tech Streamlines Sample Prep for Ion Torrent; Moves Away from Traditional Emulsion PCR. In *Sequence*; GenomeWeb LLC: New York, 2011.
(26) Ronaghi, M.; et al. *Anal. Biochem.* **1996**, *242* (1), 84–89.
(27) Karow, J. For 454 Technology, Roche Bets on Specific Apps, Taking Share from Sanger. In *Sequence*; GenomeWeb LLC: New York, 2010.
(28) Wetterstrand, K. A. *DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program*. Available at: www.genome.gov/sequencingcosts. Accessed [April 18, 2011].
(29) Lander, E. S.; et al. *Nature* **2001**, *409* (6822), 860–921.
(30) Levene, M. J.; et al. *Science* **2003**, *299* (5607), 682–686.
(31) Rigneault, H.; et al. *Phys. Rev. Lett.* **2005**, *95* (11), 117401 (1−4).
(32) Popov, E.; et al. *J. Opt. Soc. Am. A: Opt., Image, Sci., Vision* **2006**, *23* (9), 2342–2348.
(33) Samiee, K. T.; et al. *Biophys. J.* **2005**, *88* (3), 2145–2153.
(34) Samiee, K. T.; et al. *Biophys. J.* **2006**, *90* (9), 3288–3299.
(35) Foquet, M.; et al. *J. Appl. Phys.* **2008**, *103* (3), 034301 (1−9).
(36) Lundquist, P. M.; et al. *Opt. Lett.* **2008**, *33* (9), 1026–1028.
(37) Korlach, J.; et al. *Nucleosides, Nucleotides Nucleic Acids* **2008**, *27* (9), 1072–1083.
(38) Korlach, J.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (4), 1176–1181.
(39) Eid, J.; et al. *Science* **2009**, *323* (5910), 133–138.
(40) Travers, K. J.; et al. *Nucleic Acids Res.* **2010**, *38*, (15) e159 (1−8).
(41) Chin, C. S.; et al. *New Eng. J. Med.* **2011**, *364* (1), 33−42.
(42) Flusberg, B. A.; et al. *Nat. Methods* 7(6), 461−465.
(43) Uemura, S.; et al. *Nature 464* (7291), 1012−1017.
(44) Strezoska, Z.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (22), 10089–10093.
(45) Drmanac, R.; et al. *Electrophoresis* **1992**, *13* (8), 566–573.
(46) Drmanac, R.; et al. *Science* **1993**, *260* (5114), 1649–1653.
(47) Drmanac, S.; et al. *Nat. Biotechnol.* **1998**, *16* (1), 54–58.
(48) Schirinzi, A.; et al. *Genet. Test.* **2006**, *10* (1), 8–17.
(49) Shendure, J.; et al. *Science* **2005**, *309* (5741), 1728–1732.
(50) Drmanac, R.; et al. *Science* **2010**, *327* (5961), 78−81.
(51) Reid, C. *Future Oncol.* **2011**, *7*(2), 219−221.
(52) McKernan, K. J.; et al. *Genome Res.* **2009**, *19* (9), 1527–1541.
(53) Margulies, M.; et al. *Nature* **2005**, *437* (7057), 376–380.
(54) Li, J. B.; et al. *Genome Res.* **2009**, *19* (9), 1606–1615.
(55) The International HapMap Consortium. *Nature* **2007**, *449* (7164), 851−861.
(56) Roach, J. C.; et al. *Science* **2010**, *328* (5978), 636−639.
(57) Lee, W.; et al. *Nature* **2010**, *465* (7297), 473−477.
(58) Oh, J. D.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (26), 9999–10004.
(59) Nyren, P.; Pettersson, B.; Uhlen, M. *Anal. Biochem.* **1993**, *208* (1), 171–175.
(60) Nyren, P. *Anal. Biochem.* **1987**, *167* (2), 235–238.
(61) Hyman, E. D. *Anal. Biochem.* **1988**, *174* (2), 423–436.
(62) Ronaghi, M.; Uhlen, M.; Nyren, P. *Science* **1998**, *281* (5375), 363.
(63) Gharizadeh, B.; et al. *Anal. Biochem.* **2002**, *301* (1), 82–90.
(64) Esfandyarpour, H.; et al. *J. Vac. Sci. Technol., B* **2008**, *26* (2), 661–665.
(65) Esfandyarpour, H.; et al. *Biomicrofluidics* **2008**, *2* (2) 024102 (1−11).
(66) Esfandyarpour, H.; Ronaghi, M. Obtaining sequence information from single stranded DNA template comprises providing primer region of DNA template, placing multiple copies of template DNA, adding DNA polymerization mixture, and measuring temperature change; patent application no. US2008166727-A1 US959317, 2007.

(67) Thayer, A. M. *Chem. Eng. News* **2010**, *88* (34), 18−18.
(68) Rothberg, J.e.a. *Methods and apparatus for measuring analytes using large scale FET arrays*; USPTO, Ed., Ion Torrent Systems Incorporated: USA, 2009, patent application number: 20090127589.
(69) Rothberg, J.e.a. *Methods and apparatus for measuring analytes using large scale FET arrays*; USPTO, Ed., Ion Torrent Systems Incorporated: USA, 2009, patent application number: 20090026082.
(70) Karow, J. *Ion Torrent Unveils New $50k Electronic Sequencer*; 2010; Available from http://www.genomeweb.com/sequencing/ion-torrent-unveils-new-50k-electronic-sequencer.
(71) Branton, D.; et al. *Nat. Biotechnol.* **2008**, *26* (10), 1146–1153.
(72) Bayley, H. *Curr. Opin. Chem. Biol.* **2006**, *10* (6), 628–637.
(73) Kasianowicz, J. J.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (24), 13770–13773.
(74) Deamer, D. W.; Branton, D. *Acc. Chem. Res.* **2002**, *35* (10), 817–825.
(75) Tabard-Cossa, V.; et al. *ACS Nano* **2009**, *3* (10), 3009–3014.
(76) Tropini, C.; Marziali, A. *Biophys. J.* **2007**, *92* (5), 1632–1637.
(77) Wiggin, M.; et al. *Biophys. J.* **2008**, *95* (11), 5317–5323.
(78) Zwolak, M.; Di Ventra, M. *Nano Lett.* **2005**, *5* (3), 421–424.
(79) Gracheva, M. E.; Aksimentiev, A.; Leburton, J. P. *Nanotechnology* **2006**, *17* (13), 3160–3165.
(80) Soni, G. V.; Meller, A. *Clin. Chem.* **2007**, *53* (11), 1996–2001.
(81) Sauer-Budge, A. F.; et al. *Phys. Rev. Lett.* **2003**, *90* (23), 238101/1–238101/4.
(82) McNally, B.; et al. *Nano Lett.* **2010**, *10* (6), 2237–2244.
(83) Kasianowicz, J. J.; et al. *Annu. Rev. Anal. Chem.* **2008**, *1*, 737–766.
(84) Zwolak, M.; Di Ventra, M. *Rev. Mod. Phys.* **2008**, *80* (1), 141–165.
(85) Meller, A.; Branton, D. *Electrophoresis* **2002**, *23* (16), 2583–2591.
(86) Li, J. L.; et al. *Nat. Mater.* **2003**, *2* (9), 611–615.
(87) Fologea, D.; et al. *Nano Lett.* **2005**, *5* (10), 1905–1909.
(88) Towards the 15-minute genome. In *The Economist*; The Economist Newspaper Limited: London, 2011; from the print ed. Technology Quarterly; http://www.economist.com/node/18304268#footnote1.
(89) Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotechnol.* **2001**, *19* (7), 636–639.
(90) Astier, Y.; Braha, O.; Bayley, H. *J. Am. Chem. Soc.* **2006**, *128* (5), 1705–1710.
(91) Clarke, J.; et al. *Nat. Nanotechnol.* **2009**, *4* (4), 265–270.
(92) Wu, H. C.; et al. *J. Am. Chem. Soc.* **2007**, *129* (51), 16142–16148.
(93) Lieberman, K. R.; et al. *J. Am. Chem. Soc.* **2010**, *132* (50), 17961−17972.
(94) Stoddart, D.; et al. *Nano Lett.* **2010**, *10* (9), 3633−3637.
(95) Kang, X. F.; et al. *Angew. Chem., Int. Ed.* **2005**, *44* (10), 1495–1499.
(96) Merchant, C. A.; et al. *Nano Lett.* **2010**, *10* (8), 2915–2921.
(97) Garaj, S.; et al. *Nature* **2010**, *467* (7312), 190−193.
(98) Luan, B. Q.; et al. *Phys. Rev. Lett.* **2010**, *104* (23), 238103 (1−4).
(99) Polonsky, S.; Rossnagel, S.; Stolovitzky, G. *Appl. Phys. Lett.* **2007**, *91* (15), 153103 (1−3).
(100) Balagurusamy, V. S. K.; Weinger, P.; Ling, X.S. *Nanotechnology* **2010**, *21* (33), 335102 (1−9).
(101) Singer, A.; et al. *Nano Lett.* **2010**, *10* (2), 738−742.
(102) Drmanac, R.; et al. *Genomics* **1989**, *4* (2), 114–128.
(103) Ling, X. S. B.; Pertsinidis, B. A. *Hybridization-assisted nanopore sequencing of nucleic acids*; US patent application no. 2007 0190542, 2007.
(104) Cahill, M. J.; et al. *PLoS One* **2010**, *5* (7), e11518.
(105) Lin, J. Y.; et al. *Science* **1999**, *285* (5433), 1558–1562.
(106) Valouev, A.; et al. *J. Comput. Biol.* **2006**, *13* (2), 442–462.
(107) Valouev, A.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (43), 15770–15775.
(108) Zhou, S.; et al. *J. Bacteriol.* **2004**, *186* (22), 7773–7782.
(109) Armbrust, E. V.; et al. *Science* **2004**, *306* (5693), 79–86.
(110) Haas, B. J.; et al. *Nature* **2009**, *461* (7262), 393–398.
(111) Zhou, S.; et al. *BMC Genomics* **2007**, *8*, 278.
(112) Schnable, P. S.; et al. *Science* **2009**, *326* (5956), 1112–1115.
(113) Wei, F.; et al. *PLoS Genet.* **2009**, *5* (11), e1000715.
(114) Wei, F.; et al. *PLoS Genet.* **2009**, *5* (11), e1000728.
(115) Zhou, S.; et al. *PLoS Genet.* **2009**, *5* (11), e1000711.
(116) Kidd, J. M.; et al. *Nature* **2008**, *453* (7191), 56–64.
(117) Teague, B.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (24), 10848–10853.
(118) Jo, K.; Schramm, T. M.; Schwartz, D. C. *Methods Mol. Biol.* **2009**, *544*, 29–42.
(119) Reisner, W.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (30), 13294–13299.
(120) Das, S. K.; et al. *Nucleic Acids Res.* **2010**, *38* (18), e177.
(121) Nagayama *Method of determining base sequence of DNA or RNA using heavy element labeling and imaging by transmission electron microscopy*; USPTO, Ed.; USA, 2008.
(122) Tanaka, H.; Kawai, T. *Nat. Nanotechnol.* **2009**, *4* (8), 518–522.
(123) Schadt, E. E.; Turner, S.; Kasarskis, A. *Hum. Mol. Genet.* **2010**, *19* (R2), R227–R240.