Third-generation sequencing fireworks at Marco Island

David J Munroe & Timothy J R Harris

Advances in sequencing platforms promise to make this technology more accessible.

It was unseasonably cold in Florida during the Advances in Genome Biology and Technology (AGBT) meeting on Marco Island, on 24–27 February, but there was no cooling the enthusiasm and excitement of meeting participants over the new developments and innovations that continue to drive DNA sequencing technology. Even the lavish firework display could not upstage the sequencing pyrotechnics on offer from the newest generation of instruments showcased during the meeting.

Over the course of the past 5 years, the development of so-called 'next'- or 'second'-generation DNA sequencing, and the applications that this enabled, have firmly established DNA sequencing as the preeminent technology driving future developments in genomics. As reported at AGBT, the dominant second-generation sequencing platforms—HiSeq from Illumina (San Diego, CA) and SOLiD from Life Technologies (Foster City, CA)—have been optimized so that, by years end, they will not only have substantially reduced hands-on sample preparation time but also have their throughput increased to ≥100 Gb of mappable sequence per run. Improvements in the new Illumina platform (HiSeq 2000) include reagent optimization, the use of two flow cells and a dual surface imaging system, whereas the new SOLiD platform (SOLiD 4) makes use of a newly engineered DNA ligase, smaller bead size, reagent optimization, and improved software for bead detection and color calling. In addition to increased throughput, the SOLiD 4 boasts a >99.9% accuracy rate.

Last year, these platforms were joined by the commercial launch of another system, the arrayed nanoball system of Complete Genomics (Mountain View, CA), which is an iteration of the sequencing-by-ligation approach. Unlike the Illumina and Life Technologies sequencing businesses, which were positioned as instrument vendors, the Complete Genomics business model is to operate as a sequencing service rather than sell instrumentation and consumables. The Complete Genomics platform uses a proprietary combinatorial probe–anchor ligation strategy to sequence amplified DNA templates that are self-assembled into DNA nanoballs anchored onto patterned nanoarrays[1]. The ligation chemistry is complex, as is the data analysis inherent to all short-read platforms, two features that together translate into long turnaround times. Even so, a recent report detailing the sequencing of three human genomes demonstrates that this platform is highly accurate and is capable of generating an average of 45–87 fold coverage at a consumables cost of $4,400 per genome[1].

Although improvements to the second-generation continue to impress, perhaps the greatest 'buzz' at AGBT and elsewhere has been about the development of so-called third-generation DNA sequencing platforms. Designed to complement second-generation sequencing, third-generation platforms have several characteristics that distinguish them from their predecessors, including single-molecule templates, lower cost per base, easy sample preparation, significantly faster run times and simplified primary data analysis. Long-read lengths (hundreds of base pairs or more) enable *de novo* sequencing and simplify data analysis. In particular, a long-read length simplifies sequence assembly and facilitates a variety of data analysis functions such as detection of copy number variations (CNVs), translocations, splice variation, chimeric transcripts and haplotype phasing. The use of single-molecule templates translates into simplified template preparation and typically reduces the amount of sample needed for analysis. Third-generation sequencing platforms also have significantly faster run times compared with second-generation instruments (minutes as opposed to days). These short run times will facilitate application development and open the door to the routine use of sequencing as a diagnostic tool. Currently, several such platforms are in various stages of development. Four distinguish themselves from the rest: Pacific Biosciences (PacBio; Menlo Park, CA), Life Technologies (Carlsbad, CA), Oxford Nanopore (Oxford, UK) and Ion Torrent (Gilford, CT). The representatives of these companies were decked out in their brightly colored company regalia at the meeting, with each ensconced in their respective rooms like the pits of a Formula One race.

Of the emerging third-generation technologies, the PacBio and Life Technologies platforms are the most similar and closest to commercial release, with early-access partnerships scheduled for midyear and year-end, respectively. The similarities between these two platforms confer a shared set of strengths and weaknesses. Both the PacBio and Life Technologies instruments use DNA polymerase and terminal phosphate–labeled nucleotides[2] that allow long read lengths (1 kb and 1.5 kb, respectively) and short run times (15 min and 20 min, respectively). They both also use a charge-coupled diode (CCD) array detection system[3]. This means that the throughput of these platforms is restricted by the current state-of-the-art in CCD array technology. Simply put, these cameras have a finite amount of data-recording capacity. Until this capacity is increased, the per-run throughput of these platforms will be limited to a level no higher than that of the Illumina and SOLiD second-generation sequencers.

But it is the differences, rather than the similarities, between the PacBio and Life Technologies platforms that are most pertinent. The reactions in the PacBio RS sequencer are performed in 80,000 zero-mode waveguide (ZMW) 'wells', each holding 20 zeptoliters (10^{-21} liters)[4–7](**Fig. 1a**). In addition to *de novo* sequencing capabilities, the first release of the PacBio instrument will also offer redundant re-sequencing and strobe-sequencing applications. Redundant sequencing generates multiple independent reads of each template molecule, resulting in accuracy rates exceeding 99.9%. The second application, strobe sequencing, is a simplified alternative to second-generation sequencing's mate-pair application. Strobe sequencing was developed as a solution to the problem that continuous illumination required by the excitation laser inflicts photo

David J. Munroe and Timothy J. R. Harris are in the Advanced Technology Program, SAIC-Frederick, Inc., National Cancer Institute–Frederick, Frederick, Maryland, USA.
e-mail: dmunroe@ncifcrf.gov

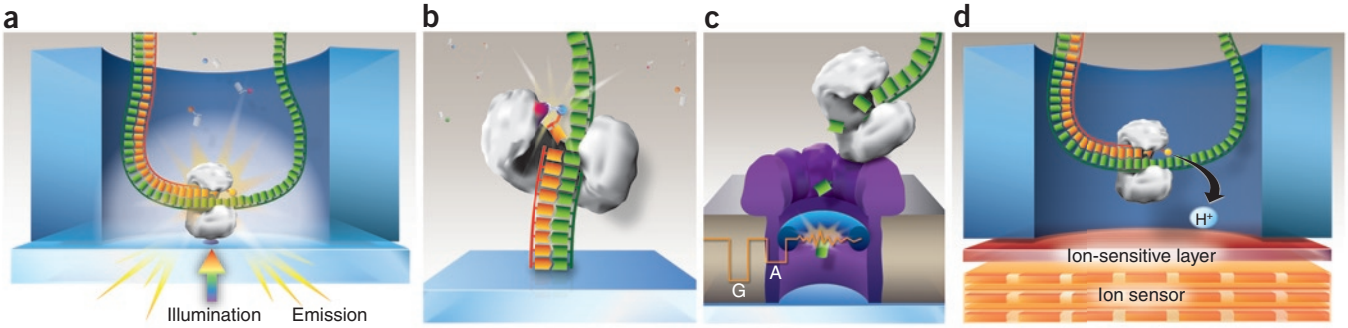

Figure 1 Third-generation sequencing platforms. (**a**) Pacific Biosciences SMRT (single-molecule real-time) DNA sequencing method. The platform uses a DNA polymerase anchored to the bottom surface of a ZMW (pictured in cross section). Differentially labeled nucleotides enter the ZMW via diffusion and occupy the 'detection volume' (white translucent halo area) or microseconds. During an incorporation event, the labeled nucleotide is 'held' within the detection volume by the polymerase for tens of milliseconds. As each nucleotide is incorporated, the label, located on the terminal phosphate, is cleaved off and diffuses out of the ZMW. (**b**) Life Technologies FRET sequencing platform uses base fluorescent labeling technology, a DNA polymerase modified with a quantum dot and DNA template molecules immobilized onto a solid surface. During an incorporation event, energy is transferred from the quantum dot to an acceptor fluorescent moiety on each labeled base. Light emission can only emanate from labeled nucleotides as they are being incorporated. (**c**) The Oxford nanopore sequencing platform uses an exonuclease coupled to a modified α-hemolysin nanopore (purple, pictured in cross section) positioned within a lipid bilayer. As sequentially cleaved bases are directed through the nanopore, they are transiently bound by a cyclodextrin moiety (blue), disturbing current through the nanopore in a manner characteristic for each base. (**d**) The Ion Torrent sequencing platform uses a semiconductor-based high-density array of microwell reaction chambers positioned above an ion-sensitive layer and an ion sensor. Single nucleotides are added sequentially, and incorporation is recorded by measuring hydrogen ions released as a by-product of nucleotide chain elongation.

damage on the polymerase in the ZMW guide wells, thus limiting read lengths. Strobe sequencing addresses this issue by periodically 'turning off' the excitation laser. While the laser is 'off,' no sequence data can be collected, but the polymerase can continue to traverse the template molecule without incurring damage; and the distant sequence is then read when the laser is turned back on. The net effect is that multiple sequence reads (totaling an average of 1 kb) can be collected across longer stretches of each contiguous template molecule.

In contrast to PacBio, the Life Technologies platform covalently binds the end of the DNA template molecule to a glass array surface (**Fig. 1b**). The DNA polymerase used in the Life Technologies system is modified with a quantum dot fluorescent donor molecule that enables a fluorescence resonance energy transfer (FRET)–based labeling strategy offering two distinct advantages. First, light emission can only emanate from labeled nucleotides as they are being incorporated, leading to a significantly lower background. Second, because a FRET-based system does not require continuous high-energy laser excitation, significantly less photodamage is inflicted on the polymerase, which should ultimately lead to much longer read lengths. With the initial release of this platform, Life Technologies will also offer a redundant sequencing application that will push accuracy rates to >99.9%. Currently in development is an ultra-long-read-length application (>100 kb), in which single template molecules are stretched in nanotubes and sequenced by several polymerase molecules simultaneously. As this platform comes closer to commercial release, we will see to what extent these differences translate into advantages.

Slightly further from commercial release is the Oxford Nanopore Technologies instrument. Rather than using a sequencing-by-synthesis method, this technology employs an exonuclease-based 'sequencing by deconstruction' approach. At the heart of this technology is an exonuclease coupled to a modified α-hemolysin nanopore (**Fig. 1c**). The modified nanopores are positioned within a lipid bilayer over a microwell that contains a pair of electrodes on either side of the lipid bilayer. When an electrical potential is applied, the high intrinsic resistance of the bilayer directs a cation-modulated current through the nanopore. As a DNA sample is introduced, the exonuclease functions to 'capture' the DNA molecule and direct the sequentially cleaved bases through the nanopore. As each cleaved base traverses the nanopore, the current is disturbed in a manner characteristic for each base, creating an 'electrical trace' unique to each nucleotide[8].

Distinct advantages of this system include a low instrument fabrication and operation cost due to the lack of labeled nucleotides and optical detection systems (that is, laser and CCD camera). In addition, the Oxford Nanopore platform is compatible with direct RNA sequencing and the detection of modified bases[8] by virtue of each individual base's characteristic ability to disturb electrical current, which should enable epigenetics applications. A clear disadvantage, however, is that because the template molecule is digested during sequencing, redundant sequencing (and the associated high accuracy) is not possible. However, this drawback could be eliminated by simply replacing the exonuclease coupled to the nanopore with a DNA polymerase. Several other noteworthy groups, including GE Healthcare (Little Chalfont, UK), are also developing nanopore-based sequencing platforms, the details of which have not yet been made public.

Arguably, though, the most heat at AGBT was generated by the Ion Torrent Systems platform. This technology uses a semiconductor-based high-density array of microwells that function as reaction chambers (**Fig. 1d**). As DNA polymerase traverses each single-molecule template, nucleotide incorporation events are recorded using a unique and imaginative readout system that measures hydrogen ions released as a natural by-product of chain elongation—a kind of sequencing pH meter. Like other nanopore-based technologies, the Ion Torrent platform has the advantage of low instrument fabrication and operation costs owing to the lack of labeled nucleotides and optical detection systems. Ion Torrent currently claims 100–200 base reads in 1–2 h on an instrument the size of a typical microwave oven with a projected sales price of ~$50,000. Although highly anticipated, no release date has yet been scheduled.

Today, we stand at the edge of an era when new sequencing technologies and the greatly reduced cost of generating sequencing data, will open up a host of possibilities in basic research, translational medicine and diagnostics that were unimaginable a decade ago. Up until this point, the complexity and cost of large-scale capillary and second-generation DNA sequencing largely limited its practice to large, specialized centers. Third-generation sequencing technology promises to remove these barriers. The simple sample preparation, short run times and relative ease of operation inherent to single-molecule sequencing make it significantly more accessible and will translate into many more genomes or parts of genomes being sequenced. This will require continuing very substantial investments in data storage and analysis to keep pace with the sequencing machines.

Perhaps the greatest impact, though, will be felt in clinical medicine and personalized healthcare, given that the characteristics of third-generation sequencing make these platforms particularly well suited to molecular diagnostics. Areas in which we might expect to see this new sequencing technology playing a more immediate role include haplotyping, mutation detection, companion diagnostics and real-time monitoring of pathogen evolution. Costs aside, it is clear that third-generation DNA sequencing is likely to produce fireworks lasting considerably longer than the ones at AGBT.

ACKNOWLEDGEMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Drmanac, R. *et al. Science* **327**, 78–81 (2010).
2. Korlach, J. *et al. Nucleosides Nucleotides Nucleic Acids* **27**, 1072–1083 (2008).
3. Lundquist, P.M. *et al. Optics Lett.* **33**, 1026–1028 (2008).
4. Levene, M.J. *et al. Science* **299**, 682–686 (2003).
5. Foquet, M. *et al. J. Appl. Phys.* **103**, 034301 (2008).
6. Korlach, J. *et al. Proc. Natl. Acad. Sci. USA* **105**, 1176–1181 (2008).
7. Eid, J. *et al. Science* **323**, 133–138 (2009).
8. Clarke, J. *et al. Nat. Nanotechnol.* **4**, 265–270 (2009).