Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology

Alice McCarthy
DOI 10.1016/j.chembiol.2010.07.004

The Cornell University laboratories of Watt Webb and Harold Craighead are miles and climatologic worlds away from Menlo Park, California, but without them, the founders of Pacific Biosciences (PacBio) may not be enjoying their current wave of success—a success that is poised to reach new levels later this year.

In late 2010, the company will make commercially available its new Single Molecule Real Time (SMRT) DNA sequencing system. This instrument represents the first third-generation sequencing instrument to become publicly available. As this article goes to press, the company is shipping the first of its DNA sequencers to ten users for beta testing.

The technology platform allows for real-time detection of biological events at single molecule resolution.

-Joseph Puglisi

DNA Sequencing

DNA sequencing involves deciphering the continuum of nucleotide bases that make up a given DNA segment. Though early, laborious techniques to accomplish this task had been available since 1971, the interest in sequencing picked up in the late 1970s with the advent of the first rapid DNA sequencing technology, known as Sanger sequencing, after its inventor Frederick Sanger. Continuous DNA sequencing of approximately 1,000 bases can be achieved by this method. Things really got interesting in 1990, when the Human Genome Project was announced: the large-scale effort to fully sequence the entire human DNA sequence (genome). Approximately 13 years and an estimated cost of over 3 billion dollars were required to churn out the first instance of a genome 3 billion nucleotides long.

To make the goal of routine genome sequencing a reality, newer, faster technologies are necessary. Second generation technologies have truly revolutionized the field of DNA sequencing, genomics research, and the understanding of genome-related disease. These came in the form of technologies introduced by Roche (454 Life Sciences), Applied Biosystems (SOLiD), and Illumina (Genome Analyzer); however, there were tradeoffs even with these newer systems, mostly centered on the lack of reliable lengths of base reads. Second-generation read lengths range from 30 up to 450 bases, and the reliability of these reads is lower than with Sanger sequencing, introducing an inherent error rate.

Zero-Mode Waveguide Technology

PacBio's SMRT technology raises the DNA sequencing bar. At the heart of PacBio's new DNA sequencing instrument is zero-mode waveguide (ZMW) technology, initially described in a 2003 *Science* paper coauthored by Webb and Craighead (Levene et al., 2003).

"ZMWs are tiny nanoholes, 70 nm in diameter by 100 nm in depth, where a single molecule of the DNA polymerase enzyme can be placed to directly observe it sequencing a strand of DNA," explains Eric Schadt, PacBio's Chief Scientific Officer. "We can multiplex that process and observe this phenomenon as it happens over thousands or tens of thousands of these holes simultaneously. The DNA polymerase becomes the sequencing engine. That is a true revolution."

In the late 1990s, under the tutelage of Webb and Craighead, PacBio founders Jonas Korlach and Stephen Turner worked collaboratively on this initial technology (Levene et al., 2003). Today, they are Vice President, and Chief Technology Officer, respectively, at PacBio.

With PacBio's SMRT technology, the polymerase enzyme is affixed at the bottom of a ZMW well. Using a single DNA molecule as a template, the polymerase incorporates fluorescently labeled DNA bases as it reads the template. Each base has a different fluorescent dye, thereby emitting a signal out of the ZMW. A detector reads the fluorescent signal and names the base based on the color of the detected signal. Once the base is added, the fluorescent tag is cleaved by the polymerase.

At the time the technology was announced, Craighead commented that the "strategy is to eavesdrop on the naturally occurring mechanisms for reading DNA molecules while they do their work."

"One of the principal challenges to creating single nucleotide sequencing capability is finding a way to exclude unwanted background noise created by the biological building materials that must be present for this strategy to work," explains Turner. "The zero-mode waveguide technology solves this problem for the first time."

Third Generation Essentials

"Our viewpoint is that the third generation has to bring the best of the first and second generations forward: longer sequenced read length, flexibility, and quality of the first generation with the lower cost and higher throughput of second generation," says Turner. He explains that the PacBio system produces sequenced read lengths up to 10,000 bases, a full ten times higher compared with the longest reads produced with Sanger sequencing.

And it works more rapidly than second-generation systems. "It is roughly 10,000–20,000 times faster than the current

 ©2010 Elsevier Ltd All rights reserved

Univ. of Washington EX2031
0001 Guardant Health, Inc v. Univ. of Washington
IPR2022-00817

technologies,'' says Schadt. ''Instead of taking an hour/nucleotide to sequence, we are watching 1–3 nucleotides/second.''

The combination of quick time to result and longer read lengths lends itself more easily to the assembly of previously unknown genomes, a process researchers call de novo genome assembly.

Most of the sequencing happening today in cancer genomes uses reference-based mapping to figure out what changes in the tumor cells have occurred to cause them to behave in a certain way. ''But we know lots of things happen in the tumor, like structural rearrangements in the genome,'' explains Schadt. ''The only way you can analyze that is a de novo assembly, which our long read lengths would allow you to do.''

The technology platform allows for real-time detection of biological events at single-molecule resolution. ''If you are able to take things down to their most fundamental unit of the molecule, that gives you the best possible resolution of what is going on,'' says Schadt. This is perhaps the main advantage of the SMRT technology.

Many of today's sequencing projects, known as genome-wide association studies (GWAS), involve the study of many genomes to find genetic variants linked to disease or any phenotype under study. The PacBio technology is not aiming to replace GWAS studies involving comparative genomics on lots of people or samples. ''With single-molecule sequencing, we are extending that research mission to be able to detect the rare mutations between people in terms of health down to the molecular level,'' says Turner.

Mapping Chemical Modifications

The other significant feature of PacBio's SMRT technology involves a time dimension, but not just in terms of how fast one can get results.

Because the technology allows the user to watch the DNA polymerase in action in real time, if the enzyme encounters any chemical modification on the DNA bases along the way, its kinetics change. ''That speaks to the epigenetic changes this process finds that are very important for regulation of the gene,'' says Schadt. The epigenome is a set of chemical modifications surrounding the bases of the DNA that have as much or more influence on the gene's activity as its sequence.

Common chemical modifications include methylation. As the sequencing occurs, the polymerase enzyme kinetics shift when it encounters a region of methylation or any other base modification. ''If there are chemical modifications on the bases of the DNA, like methylated C residues or evidence of 8-oxoG from DNA damage, the kinetics of the enzyme will change when it encounters those modified bases,'' says Schadt. This information is provided without extra protocols like bisulfite treatments, which are currently used to probe areas of methylation.

When the enzyme encounters chemically modified bases, it will slow down or speed up in a uniquely identifiable way. The kinetic patterns are specific to different modifications. ''It isn't just about highlighting a region of interest. It actually tells you what is there,'' says Turner.

PacBio researchers published findings on methylation patterns in *Nature Methods* (Flusberg et al., 2010). There they mapped the kinetic patterns of three methylation modifications: N6-methyladenine, 5-methylcytosine and 5-hydroxymethylcytosine. ''The polymerase enzyme's timing is affected in subtle and very distinct ways, so we can tell the difference, for example, between a methylcytosine and a hydroxymethylcytosine in the same place and the same context,'' adds Turner.

Observing Protein Synthesis

The single-molecule real time platform has shown applicability for nonsequencing efforts as well. PacBio collaborated with Stanford University researcher Joseph Puglisi to study the action of ribosomes, the body's protein production machinery. Puglisi's lab works on the basic mechanisms of how proteins are made and the role of ribosomes in particular. Ribosomes translate RNA messages to make proteins.

''The limitation we always faced was that in translation, you have to work at very high concentrations of the substrates in order to have translation occur efficiently, at micromolar concentrations,'' says Puglisi. With earlier single-molecule technologies, he had to work with artificially lower fluorescently labeled substrate concentrations to avoid background noise. His team had to work at 30–50 nM concentrations, a factor of 20–50-fold lower than the concentrations that occur in the cell. He says, ''It is just not the right context in order to watch protein synthesis.''

With the SMRT technology, Puglisi could observe translation in real time. ''With this method, we can simultaneously watch the ribosome performing its task but also the underlying messenger RNA sequence,'' he explains. By using fluorescently dyed amino acids – the building blocks of proteins – the instrument gives off differing pulse colors as translation occurs step by step (Uemura et al., 2010).

''One of the biggest advantages of this technology is being able to study biological processes in real time at relevant concentrations,'' says Puglisi. ''I believe that studying dynamic behavior of biological systems is one of the frontiers of us getting a full understanding of how biology occurs from a chemical perspective.''

Beyond Genomics

''Our technology is not just a DNA sequencing paradigm,'' says Turner. In the future, he believes it will well serve the integration of transcriptomics, epigenomics, genomics, metabolomics, clinical annotation, and phenotype information—a paradigm that is not fully developed yet. ''But that is where the power to transform medicine lies,'' he says. ''Our platform of looking at single molecules in real time and the fluorescent labeling of different biological systems is broadly applicable and ideally suited for that new environment.''

REFERENCES

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Nat. Methods *7*, 461–465.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Science *299*, 682–686.

Uemura, S., Aitken, C.E., Korlach, J., Flusberg, B.A., Turner, S.W., and Puglisi, J.D. (2010). Nature *464*, 1012–1017.

Alice McCarthy (alice@alicemccarthy.com) is a science writer based in Gloucester Massachusetts.

 ©2010 Elsevier Ltd All rights reserved