GENOMICS

Cheap third-generation sequencing

By covalently attaching cyclodextrin to a hemolysin nanopore, researchers show single-molecule, label-free sequencing at very high accuracy.

The quest for the Holy Grail of the $1,000 genome has led to big improvements in sequencing technologies. The original Sanger sequencing brought us the first human genome sequence for a price tag of billions of dollars; then second-generation sequencers—Illumina's Genome Analyzer, Roche's 454 sequencer and Applied Biosystems' SOLiD System—resequenced a human genome for under $100,000. Now a third generation of sequencers, able to determine the base composition of single DNA molecules, has joined the race.

Currently three technology platforms comprise this third generation. Two achieve single-molecule sequencing by incorporating and detecting fluorescently labeled nucleotides: Helicos' Genetic Analysis System is already commercially available, and Pacific Biosciences is ready to launch
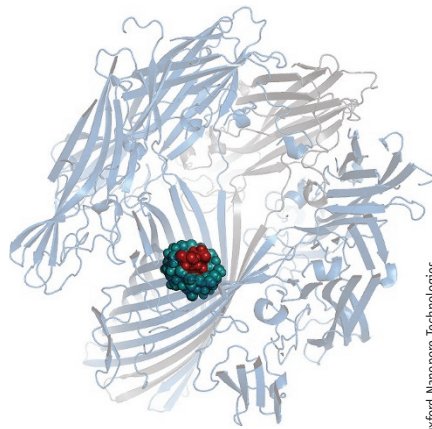
α-hemolysin nanopore (ribbon diagram) with covalently attached cyclodextrin (teal) transiently binds a base (red) traversing the pore.

Oxford Nanopore Technologies

commercial distribution of their Single Molecule Real Time (SMRT) technology next year. The third, Oxford Nanopore's nanopore sequencing, is not ready for commercial prime time, but is likely to be the cheapest of the three.

"Nanopore sequencing has the advantage that it does not require any labeling of the DNA, no expensive fluorescent reagents or really expensive CCD [charge-coupled device] cameras to record from optical chips," says Hagan Bayley from Oxford Nanopore Technologies.

In their most recent work, Bayley and his team focused on improving a nanopore that can read individual nucleoside monophosphates. Building on their previous work, they designed the nanopore from α-hemolysin with the molecular adapter cyclodextrin covalently attached to the inside of the pore. An exonuclease digests single-stranded DNA, and as single bases fall into the pore, they transiently interact with the cyclodextrin and block an electrical current that runs through the pore. The current amplitude—characteristic for the individual bases A, G, C and T as well as methylcytosine—is then easily converted to DNA sequence. Each base has a characteristic mean dwell time in the millisecond range, its dissociation rate constant is

GENOMICS

DEEP SEQUENCING OF RIBOSOME FOOTPRINTS

Combining ribosome profiling and deep sequencing, researchers present a method to monitor protein translation genome-wide.

The power of deep sequencing continues to transform many areas of biological research. In some cases, such as whole-genome sequencing, the speed and economy of new technology allows experiments to be done at a scale that was previously unfeasible. In other cases, RNA-Seq, for instance, deep sequencing allows a more quantitative analysis and a wider dynamic range than do other methods. In still other cases, such as the translational profiling approach recently published by Jonathan Weissman and colleagues at the University of California, San Francisco, the application of deep sequencing provides not only an efficient and quantitative tool, but yields qualitatively new types of information as well.

The translating ribosome protects an approximately 30-nucleotide segment of its mRNA template from digestion by nuclease. By using short-read Illumina technology to sequence DNA libraries corresponding to all protected RNA fragments in a cell, Weissman and colleagues obtain a comprehensive picture of protein translation in the cell.

This approach will have many potential applications. First, it is likely to be immensely useful in helping to define the proteome. As

Weissman puts it: "For complex genomes like the human genome, you really can't annotate what the expressed polypeptides are. This approach basically gives you an objective and comprehensive way of doing that." In their present work, Weissman and colleagues apply the method to budding yeast, but in principle it may be used in any organism. What is more, in combination with epitope-tagged ribosomes, it may prove useful in defining tissue-specific translation. "For something like molecular neuroanatomy," says Weissman, "I think this will usher in a new era."

Second, ribosome profiling is a more accurate way of looking at protein production than measuring mRNA abundance. The researchers used ribosome profiling to map the density of ribosome footprints on thousands of different mRNAs and then derived reproducible rates of protein translation from these measurements. They report that translation rates are better than mRNA abundance measurements for predicting protein abundance. "The nice thing for us about quantitative proteomics," says Weissman, "is that it's an objective way of seeing how well we're doing." In fact, by correcting for the higher density of ribosomes that they observe at the 5′ ends of messages, the researchers found that they could improve the correlation of translation rate with protein abundance even more (for an R^2 of 0.6).

voltage-dependent and a potential of +180 mV ensures that the base is swept out of the pore on the other side.

The ability to directly read the fifth base, methylcytosine, without prior bisulfite conversion, is unique to this nanopore technology and, judging by the feedback in the blogosphere, is already creating great excitement among researchers seeking to sequence the epigenome.

Nanopore sequencing promises to meet the needs for most sequencing users: the 99.8% read accuracy is high, and errors will be easier to correct computationally because the amplitude-based recordings may cause an ambivalent call between two but not between all four bases. Homopolymer stretches are recorded without difficulty, as the pore records each base irrespective of what comes before or after. Reads will be long. "It does have the potential to read thousands of bases," says Bayley, "and there is no degradation of sequence quality. Even if there is a glitch in the middle, it will pick up again."

But there are two important technical issues that still need to be addressed before Oxford Nanopore's sequencers are ready for prime time. One is the optimal attachment of the exonuclease to the pore. "The real challenge," says Bayley, "is to get it to drop every single base into the pore." The other is parallelization, which would be achieved by creating a chip with tens of thousands of pores to ensure a fast overall sequencing process.

With cheap, long and accurate reads, nanopore sequencing may well bring the Holy Grail within grasping distance.

Nicole Rusk

RESEARCH PAPERS
Clarke, J. *et al*. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol*. advance online publication 22 February, 2009.

A third area that will be amenable to analysis by ribosome profiling is translational control. In their present work, the researchers use profiling to study the translational response to amino acid starvation in yeast. The method will no doubt be used to examine regulation of protein synthesis in disease and other stress states in higher organisms as well.

Lastly, the method has high spatial precision, allowing determination of the reading frame being translated. It could therefore be used to study programmed frameshifts and stop-codon readthrough. Or, as done by Weissman and colleagues in their current work in yeast, it could be used to map unorthodox translation initiation sites in the 5′ untranslated regions of mRNAs.

As Weissman sums it up, "it's now possible to directly make comprehensive high-quality measurements of the rates of protein translation. This can be used to define what proteins are being made and how much of those proteins are being made, as well as being a very nice analytical tool for looking at the process of translation itself."

Natalie de Souza

RESEARCH PAPERS
Ingolia, N.T. *et al*. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, published online 12 February, 2009.

Analyzing the prenylome

Protein prenylation refers to the attachment of farnesyl or geranylgeranyl groups to target proteins via prenyltransferases. To explore the total mammalian prenylome, Nguyen *et al*. developed a biotin-labeled geranylpyrophosphate substrate, as well as a series of prenyltransferases capable of conjugating the substrate to target proteins. This approach allowed all prenylated proteins to be detected by western blotting and mass spectrometry.
Nguyen, U.T.T. *et al. Nat. Chem. Biol.* advance online publication 15 February 2009.

Identifying lincRNAs

Mammalian genomes encode a number of large intervening noncoding RNAs (lincRNAs), but their functional importance is not well-understood. Guttman *et al*. used chromatin immunoprecipitation followed by massively parallel sequencing to identify chromatin domains indicative of active transcriptional units residing outside of protein-coding gene loci. They found that a majority of the lincRNAs were highly evolutionarily conserved, indicating that they are likely functional. They also describe an approach for predicting the function of these lincRNAs.
Guttman, M. *et al. Nature* **458**, 223–227 (2009).

Whole cell electron microscopy

de Jonge *et al*. describe a method for imaging whole cells in their native liquid state, using scanning transmission electron microscopy. Using gold nanoparticles as tags, they imaged single epidermal growth factor molecules bound to their receptors with a spatial resolution of 4 nanometers, higher than current super-resolution imaging methods. The rapid imaging speed may allow snapshots of dynamic information to be captured.
de Jonge, N. *et al. Proc. Natl. Acad. Sci. USA* **106**, 2159–2164 (2009).

A gene expression atlas of the mouse brain

Ng *et al*. present the Anatomic Gene Expression Atlas (AGEA) of the adult C57Bl/6J mouse brain. This publicly available resource was created from more than 4,000 gene expression profiles from *in situ* hybridization data from the Allen Brain Atlas. The AGEA allows users to explore three-dimensional gene expression–based correlation and clustering maps of the brain, and a search tool allows users to retrieve lists of enriched genes at points of interest.
Ng, L. *et al. Nat. Neurosci.* **12**, 356–362 (2009).

Predicting membrane protein structures

Membrane protein structures are notoriously difficult to solve experimentally, but they also pose a challenge for *in silico* structural prediction. Barth *et al*. describe a computational method for determining the structures of large membrane proteins by enforcing constraints on helix-helix packing interactions as predicted from the protein sequence or from experiments. They obtained near-native models for 9 of 12 membrane proteins.
Barth, P. *et al. Proc. Natl. Acad. Sci. USA* **106**, 1409–1414 (2009).