

DNA SEQUENCING

Packing, trapping and sequencing

Ultralow concentrations of DNA can be optically sequenced with SMRT DNA sequencing.

In principle, optical DNA-sequencing protocols have the advantage of reading long strands of DNA in real time and at high speeds. In practice, however, reading long DNA strands is a challenge with current methods, which require high concentrations and suffer from short-chain loading bias. To overcome these limitations, a research team led by Meni Wanunu at Northeastern University in Boston has now developed an efficient voltage-controlled DNA-loading technology that enables single molecule, real time (SMRT) sequencing of long DNA strands at ultralow concentrations.

In SMRT sequencing, the replication of DNA by a single DNA polymerase is optically recorded using fluorescently labeled dNTP analogues. The DNA replication occurs in a zero-mode waveguide (ZMW), the base of a small cavity in an opaque film which allows molecules to freely flow in and out; but, unlike a standard waveguide, which guides light, the ZMW's diameter is too small for light to enter. Since the film and ZMWs are both opaque, the background fluorescence from the dNTP analogues in the surrounding medium is blocked from reaching the detectors. However, since dNTP analogues are integrated into the DNA at the bottom of the waveguide, the fluorescence emission can still exit the waveguide at that location and be recorded.

Until now, it has been a challenge to pack long strands of DNA into the zeptoliter-sized (a trillion times smaller than a nanoliter) waveguide cavity. For example, a 10 kilobase-pair (kb) DNA strand has an effective diameter, as it flows and bends in liquid, of over five times that of the ZMW. In previous work, the team introduced the nanopore ZMW (NZMW), in which the nanopore enabled the application of a voltage along the length of the waveguide cavity. The strong electric field pulls electrically charged DNA molecules into the NZMW, "allowing the DNA to cross over a huge entropic barrier," as described by the paper's first author, Joseph Larkin. This field "is actually compressing the DNA," he notes. The early NZMW technology,

GENE EXPRESSION

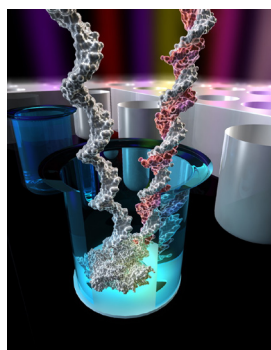
SINGLE-CELL RNA-SEQ—NOW WITH PROTEIN

Two new methods simultaneously measure epitope and transcriptome levels in single cells.

The molecular understanding of the cell has been greatly advanced by single-cell RNA-seq, a technique that generates a library of all transcripts (the 'transcriptome') in a single cell. The technique has revealed surprising heterogeneity in cell populations previously considered homogeneous, identified new and rare cell types, and extended our understanding of cellular development. The transcriptome is, however, only a proxy of the 'proteome', the collection of proteins in a cell that defines how the cell looks, acts, and reacts. Although the transcriptome provides valuable information, it does not necessarily reflect protein abundance in the cell. And while flow cytometry is an established strategy for profiling populations at single-cell resolution according to surface-protein levels, it cannot access the rich phenotypic information available in the full cellular transcriptome.

Now, independent efforts led by Marlon Stoeckius at the New York Genome Center (NYGC) and Vanessa Peterson at Merck have yielded approaches to measure levels of both gene and protein expression in single cells on a large scale.

Both CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing), developed by the NYGC group, and REAP-seq (RNA expression and protein sequencing assay), designed by the Merck group, use a similar approach. Proteins are detected by using antibodies conjugated to a tripartite DNA sequence that contains a primer for amplification and sequencing (PCR handle), a unique oligonucleotide that acts as an antibody barcode, and a poly(dA) sequence. The poly(dA) sequence allows for simultaneous extension of antibody-specific DNA sequences and cDNAs in the same poly(dT)-primed reaction. This generates a protein readout that is captured and sequenced along with the cell's transcriptome. The two approaches differ in how the DNA barcode is conjugated to the antibody. While antibodies used in CITE-seq are conjugated



Artistic rendering of DNA polymerase replicating DNA in a nanopore zero-mode waveguide. Image: Ella Maru Studio.

which demonstrated electrophoretic DNA packing, relied on a highly fluorescent silicon nitride substrate, and thus it was incompatible with SMRT sequencing.

“We had to replace the silicon nitride with another material,” recalls Wanunu. The team fabricated the waveguides on a 20-nm-thick glass layer deposited on a silicon nitride membrane. Then the silicon nitride was selectively etched, which left a glass substrate and reduced the photoluminescence background by 40-fold. With a nearly background-free substrate, researchers studied the loading and sequencing of DNA in the NZMW.

Not only did the team find a nearly uniform loading probability for DNA strands from 1–48.5-kb long, but it also found a loading efficiency five to six orders of magnitude greater than that possible by diffusion alone. By functionalizing the waveguide base with biotin, streptavidin on the DNA polymerase bound the biotin and held the DNA in the waveguide

for sequencing. With the biotin functionalization, DNA strands remained in the waveguide indefinitely; whereas without the surface chemistry, molecules escaped when the voltage was turned off.

The team concluded by sequencing a 3-pM sample of a 20-kb DNA strand with their redesigned NZMW. They demonstrated a 67% single-read accuracy with typical read lengths of 1.6 kb. For future work, “we are thinking about scalability,” explained Wanunu. “Currently each pore is made by hand,” Larkin said. By integrating the waveguides onto a porous substrate, the team hopes to fabricate large arrays of NZMWs for high-throughput SMRT DNA sequencing.

Zachary J Lapin

RESEARCH PAPERS

Larkin J. *et al.* Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nat. Nanotechnol.* <http://dx.doi.org/10.1038/nnano.2017.176> (2017).

to streptavidin that is noncovalently bound to biotinylated DNA barcodes, REAP-seq relies on covalent bonds between the antibody and aminated DNA barcode.

In a proof-of-principle study, Stoeckius and colleagues monitored ten surface proteins and the transcriptomes of 8,000 single cells from cord blood mononuclear cells. The CITE-seq analysis revealed cell profiles similar to those established by flow cytometry. In addition, the multimodal data from CITE-seq enhanced the phenotypic characterization of a specific type of immune cell, the natural killer cell, compared with single-cell RNA-seq alone.

REAP-seq was used to characterize the effect of a CD27 agonist on human naïve CD8⁺ T cells by employing 80 barcoded antibodies and monitoring the expression of more than 20,000 genes in a single workflow. The transcriptome data analysis identified several differentially expressed genes in treated versus untreated cells. But REAP-seq’s ability to quantify cell surface proteins also led researchers to determine that ICOS, an immune checkpoint protein, is increased on the surface of treated cells, regardless of the fact that this protein’s mRNA does not differ in abundance between treated and untreated cells. REAP-seq also identified a small and previously unknown cell population within the enriched naïve CD8⁺ lymphocytes.

While both CITE-seq and REAP-seq add to established methods for transcriptome analysis without affecting the quality of the data, the main limitation of both approaches is the quality of the antibodies used and the epitope location, which is currently restricted to the cell surface. Both research groups anticipate that the use of these tools will soon be extended to measure intracellular proteins.

Vesna Todorovic

RESEARCH PAPERS

Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

Peterson, V.M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* <http://dx.doi.org/10.1038/nbt.3973> (2017).