nature biotechnology

The long view on sequencing

Ultra-long read, single-molecule nanopore sequencers are beginning to find an expanding set of research applications in both human genomics and pathogenomics.

Not so long ago, nanopore sequencing was more dream than reality. Just six years have passed since it was first demonstrated that a nanopore could discriminate individual bases of a single-stranded DNA sequence. In the interim, commercial nanopore devices no bigger than a USB memory stick have shown their power in tracking full-length genome sequences of Ebola and Zika viruses during disease outbreaks in West Africa and Brazil, respectively. In this issue, an international collaboration reports the first application of nanopore sequencing to successfully assemble *de novo* a complete, human genome (p. 338). Another report demonstrates its use in resolving long tracts of near-identical tandem repeats in a Y chromosome centromere (p. 321). Nanopore technology still lags far behind other sequencing platforms in terms of consensus base-calling accuracy. But with nanopore sequencers now in the hands of several thousand laboratories across the world, and platform optimization proceeding apace, an expanding set of compelling applications is coming into view.

The concept of using biological pores embedded in a membrane to directly sequence a single strand of DNA via changes in electric current was first articulated in the 1980s. It took until 2012, however, for this idea to be reduced to practice. One key insight was to hook the DNA sample to a motor protein (e.g., phage phi29 polymerase or helicase), which sufficiently slows DNA transport through the pore to allow discrimination of single nucleotides. In 2014, Oxford Nanopore Technologies (ONT), the UK startup pioneering nanopore sequencing, announced an early access program to make available the MinION device for exploration by the research community. The MinION uses a single consumable flow cell that can run up to 512 nanopore channels at a time. After a commercial launch in 2015, a new generation of flow cells was created using a different nanopore—the *Escherichia coli* Curlin sigma S-dependent growth subunit G (CsgG). Current CsgG-based flow cells are capable of reading not only one DNA strand (1D), but also both strands of DNA sequentially ($1D^2$).

Today, ONT's MinION can process up to 450 bases of DNA per second per channel. Read-lengths in excess of 1.3 Mb (ultra-long reads) are being reported in different laboratories; indeed, read-length currently seems to be limited primarily by mechanical shearing of DNA during manipulation and by how much DNA can be physically loaded onto flow cells.

These ultra-long reads are important because they can traverse previously intractable sequences and be used to assist in genome assembly. A compelling demonstration is provided on p. 338 by five groups from the United Kingdom, Canada and the United States, led by Matt Loose and Nick Loman; they use a MinION to sequence the human cell line GM12878. Using nanopore reads alone, they assemble a genome with an NG50 contig size of ~6.4 Mb that covers >85% of the reference human genome with 99.88% accuracy; polishing of this assembly with short reads (Illumina) improved accuracy even further, to 99.96%. These authors also achieve several other important firsts specifically enabled by ultra-long reads: successful sequencing of the human major histocompatibility (MHC) locus on a single contig, phased over its full length; the closure of 12 large (>50 kb) gaps in the reference human genome (GRCh38); and the most contiguous human genome assembly reported to date.

Another paper in this issue (p. 321) reports the use of a MinION to sequence and assemble the Y-chromosome centrosome—the smallest centromere in the human genome. Karen Miga and her colleagues use bacterial artificial chromosome (BAC) clones of the centromere to generate a consensus assembly and thereby resolve the organization of the 52 tandem repeats within the structure—a feat that has eluded other third-generation sequencing technologies, such as Pacific Bioscience's Sequel and Illumina's Tru-Seq.

What is becoming increasingly clear is that nanopores can decipher sequences no other platforms can reach. Beyond MHC loci, nanopores may also help crack cytochrome P450 genes, T-cell receptor or antibody VDJ segments, segmental duplications, structural variants, and telomeres, to name a few. In transcriptomics, direct RNA nanopore sequencing promises profiling of low-abundance transcripts without PCR bias and more facile identification of RNA isoforms (*Nat. Methods* **15**, 201–206, 2018). Beyond sequence, nanopores can also directly detect modified bases, such as 5-methylcytosine and 5-hydroxymethylcytosine (*Nat. Methods* **14**, 407–410; 411–413, 2017), which may prove particularly attractive in epitranscriptomics, where existing technology often leads to false positives (*Nat. Chem. Biol.* **14**, 215–225, 2018).

Make no mistake, however, nanopore technology remains at an early stage in its development. Commercial production still suffers problems—users encounter flow cells with blocked or non-functional pores. The inferior accuracy of base calling means it will be some time before nanopores rival, let alone surpass, other established long-read technologies. But ultimately, nanopores look likely to be part of the solution for a low cost, end-to-end assembly of a phased human genome, in which every chromosome is complete, including telomeres and centromeres.

Improving base-calling algorithms, extracting more of the information present in the raw signal from a nanopore read, and adapting software to ultra-long reads will be instrumental in increasing raw sequence accuracy. Pacific Biosciences has encouraged open-source software development for its platform and ONT is following suit, with open-source tools that include its base-caller training pipeline software Sloika and base-calling algorithm Scrappie.

Above all, though, it is the nanopore sequencer's flexibility, portability and ease of use that make it unique. Combined with mobile phone technology, nanopore devices offer researchers the chance to sequence in the field—from pathogens to metagenomes and beyond—all at lower price-points than competing platforms.

In the long view, nanopore sequencers could bring about the democratization of sequencing. One day, perhaps, every lab will own one.

Corrected after print 15 May 2018.

Erratum: The long view on sequencing

Anonymous

Nat. Biotechnol. **36, 287 (2018); published online 5 April 2018; corrected after print 15 May 2018**

In the version of this article initially published, paragraphs 2 and 3, it was stated that Oxford Nanopore Technologies (ONT) "announced an early access program to make available the MinION—a device containing the *Mycobacterium smegmatis* porin A (MspA) nanopore. The following year, ONT launched its first commercial product containing 'flow cells' featuring >500 MspA pores in a lipid monolayer. " As ONT has never disclosed what pores the early-access MinION used, the references to MspA have been removed. In addition, since the company has never used lipids and its membrane material is a trade secret, the second sentence has also been corrected. In paragraph 4, sentence 1 "Today, ONT's MinION can process up to 450 bases of DNA per second. Read-lengths in excess of 800,000 bases (ultra-long reads) are being reported in different laboratories…" had several inaccuracies. The MinION processes 450 bases of DNA per second "per channel." Rather than read-lengths in excess of "800,000 bases," reports are of read-lengths in excess of "1.3 Mb." In the paragraph preceding the penultimate paragraph, the sentence "ONT would do well to take a leaf from Pacific Biosciences, which has encouraged open-source software development for its platform" overlooked two contributions by ONT, and has been revised. The errors have been corrected in the HTML and PDF versions of the article.