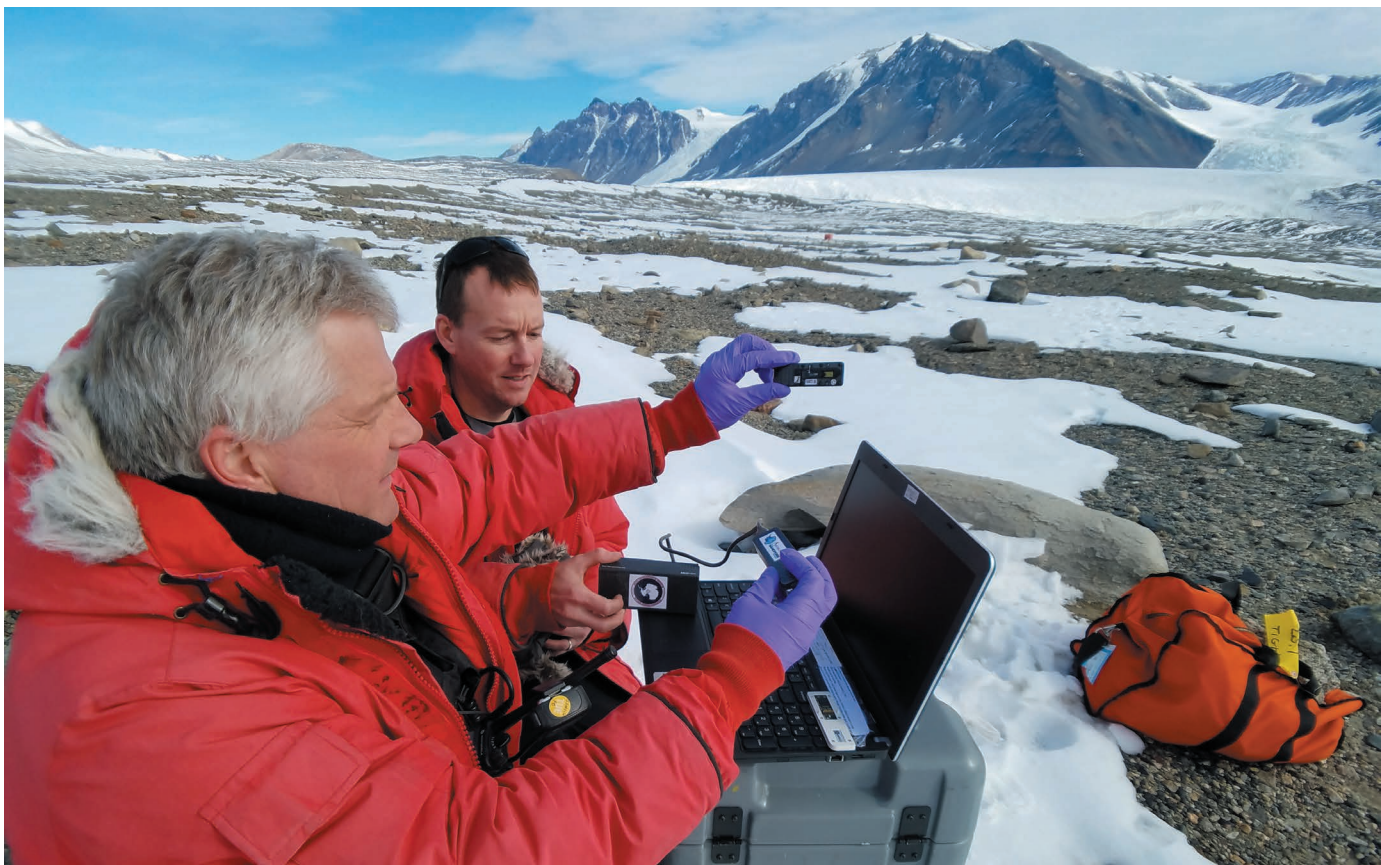# AN ACE IN THE HOLE FOR DNA SEQUENCING

Offering long reads and rapidly improving accuracy, nanopore sequencing has the potential to upend the DNA sequencing market.

Scott Tighe (left) and David Goerlitz use the MinION device to sequence microbial DNA in Taylor Valley, Antarctica.

BY MICHAEL EISENSTEIN

Christopher Mason has a trick that he likes to break out at conferences. By harvesting DNA from swabs collected from a volunteer's phone, he and his colleagues can perform on-site ancestry analyses within an hour, and even recount details of a donor's day. "We were able to predict who had just eaten an orange, and who had eaten pork, from what was left on their phones," says Mason, a computational biologist at Weill Cornell Medicine in New York City.

Mason achieves this speedy analysis using a handheld sequencing device called MinION,

developed by the UK firm Oxford Nanopore Technologies (ONT). MinION reads sequence information by threading long DNA strands through a tiny aperture known as a nanopore and detecting minute changes in electrical current caused by DNA's four component nucleotides. Mason's demonstrations provide a light-hearted illustration of the device's capabilities, but early users have also racked up some high-profile scientific achievements. MinION played a prominent part in monitoring the 2015 Ebola virus outbreak, has voyaged to Antarctica and even gone into orbit.

But MinION — which is roughly the size of a deck of cards — accounts for a relatively small

fraction of the world's sequencing output, which is still dominated by Illumina of San Diego, California. Illumina has a nearly 10-year head start, but ONT and its users are also grappling with technical challenges, most notably higher error rates. Meanwhile, competing firms are hoping to surpass ONT with innovative spins on this conceptually simple but technically complex sequencing strategy.

A ROCKY START

Illumina's ubiquitous DNA sequencing technology generates vast numbers of exceedingly accurate 'short reads' by sequentially reading the bases that are incorporated into a ▶

▶ sample during a DNA replication reaction. These strings of sequence data, which span a few hundred nucleotides, can then be computationally assembled into overlapping 'contigs' containing millions of nucleotides.

For nanopore sequencing, entire DNA fragments are analysed directly by physically threading them through a nanoscale pore.

Costing US$500–900 each, MinION's flow cells contain hundreds of nanopores, so they can analyse many molecules in parallel. The system applies voltage across each pore, and as an enzyme steadily draws the DNA through, the nucleotides block the flow of ions and produce tiny changes in electrical current, which are interpreted by specialized software (see 'Nanopore sequencing'). The resulting 'long reads' can span thousands of nucleotides, simplifying assembly.

On its release in 2014, the genomics community was tantalized by MinION's potential, but early users encountered numerous challenges. "It was a lot of effort to sequence just a single bacterial genome because the output was low, and the single-read accuracy was also pretty low," says Nicholas Loman, a microbial genomicist at the University of Birmingham, UK, and one of MinION's earliest adopters. Whereas Illumina typically achieves an average accuracy of 99.9% for individual reads, the first-generation MinION incorrectly identified roughly three of every ten bases. There were other problems, too. "You'd get one flow cell that was amazing, but then you'd get one that only gave you three active pores for no clear reason, even though they were from the same batch," says Mason. ONT did not wish to publicly comment on this article.

These limitations pigeonholed MinION mostly to applications where speed and simplicity were paramount, such as pathogen detection. "If you were asking whether there's anthrax in a letter, you could do that on a MinION very rapidly, even without perfect read accuracy," says Adam Phillippy, head of genome informatics at the US National Human Genome Research Institute in Bethesda, Maryland.
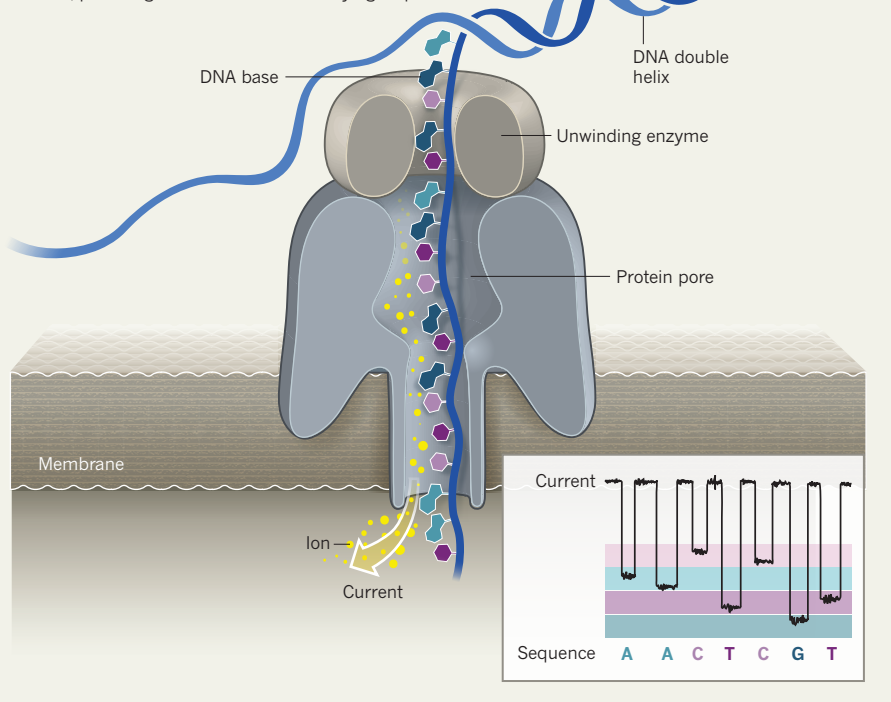
LEAPS AND BOUNDS
Today, MinION has matured, says Keith Robison, who is principal scientist at the drug-discovery company Warp Drive Bio in Cambridge, Massachusetts. More importantly, ONT has overcome the scepticism and dismissive critiques that followed the company's prelaunch announcements. "They've proven all of those people wrong, and have repeatedly delivered," says Robison.

A pore protein derived from the bacterium *Escherichia coli*, together with improvements in the flow-cell chemistry, have slashed the error rate for individual reads to 2–5% for many experiments. A big boost in data output allows researchers to better identify errors by sequencing many more molecules in parallel, and read

NANOPORE SEQUENCING
At the heart of the MinION device, an enzyme unwinds DNA, feeding one strand through a protein pore. The unique shape of each DNA base causes a characteristic disruption in electrical current, providing a readout of the underlying sequence.

DNA double helix

DNA base

Unwinding enzyme

Protein pore

Membrane

Ion

Current

Current

Sequence   A   A   C   T   C   G   T

length has jumped from about 7,000 nucleotides initially to upwards of 100,000 nucleotides today. Loman's team has pushed single-read lengths to nearly one million bases. "In our early runs, we were getting a few hundred megabases of sequence per run," says Jared Simpson, a bioinformatician at the Ontario Institute for Cancer Research in Toronto, Canada. "With the new pore, we quite quickly saw that go up to gigabases of sequence, and now up to maybe 5 or 10 gigabases per run."

The platform has also benefited from aggressive software development. Contig assembly is easier with long reads than short reads because there is more overlap, but nanopore reads are also more error-laden, and manipulating longer sequences can be computationally intensive. To address this, Phillippy's group devised an algorithm called Canu, in which conventional short-read assembly processes are tailored to compensate for the quirks of long-read data. Another software tool, ONT's Scrappie, addresses sequence errors that can arise when homopolymers — sequences containing multiple adjacent instances of a nucleotide, such as AAAAA — cause the system to hiccup.

A MATCH FOR MICROBES
MinION has proved particularly popular among infectious-disease researchers. Loman, for instance, has teamed up with colleagues in the world's virological 'hot zones' to monitor the spread of Ebola in West Africa

and Zika in Brazil[1,2]. "They were basically able to get a sequencing laboratory up and running in 48 hours, packed in luggage you could take below-decks on an airplane," says Mark Akeson, a biophysicist at the University of California, Santa Cruz, who conducted some of the foundational research in nanopore sequencing and is a member of ONT's advisory board. Loman says that this portability was a massive boon, but he notes that the massive data output could be overwhelming. "We just about managed in Brazil, but killed my Mac by overheating!"

Some groups are exploring clinical microbiology applications. Lachlan Coin, a bioinformatician at the University of Queensland in St Lucia, Australia, has developed real-time data-analysis algorithms to detect drug-resistant bacteria in blood samples. In early pilot tests using cultured bacteria and older flow cells, Coin's team could identify all of the antibiotic-resistance genes in a sample within 10 hours[3]. Current technology could halve that time, he says, but working with real-world samples in which human DNA overwhelms bacterial DNA is complicating the process. "I think that in a year or so, we'll be able to identify the antibiotic-resistance genes in patient samples within six hours," Coin says.

Other researchers are pursuing metagenomics, in which the goal is to comprehensively profile all of the organisms in a sample. In principle, every nanopore in the

flow cell could be put to work detecting a different genome at the same time. "You can get a complete genetic portrait of the species that are there — bacteria, viruses, human DNA," says Mason. He has used nanopore sequencing to conduct a metagenomic census of the famously filthy New York City subway system, and has ambitious plans for even more inhospitable environments — including Mars. Working with scientists at NASA, Mason has shown that MinION can perform robustly in zero gravity aboard the International Space Station. He and his colleagues hope one day to ship this technology to the red planet, where the technology could aid the ongoing search for extraterrestrial life.

Back on Earth, geneticist Scott Tighe at the University of Vermont in Burlington ran the MinION in Antarctica's McMurdo Dry Valleys, where his team spent more than two hours sequencing microbial samples. "The reason the run stopped was that it was too cold outside: the battery ended up dying," explains Mason, who has collaborated with Tighe on several projects.

GO BIG OR GO HOME

Nanopore veterans such as Phillippy consider microbial genome assembly to be "a solved problem". Now they are taking aim at bigger game: mammalian genomes encompassing billions, rather than millions, of nucleotides. This year, a multi-institutional team including Phillippy, Loman and Simpson reported the assembly of an entire human genome using only MinION data, achieving high contiguity and accuracy[4]. The average contig sizes ran into the megabases, Simpson says, with accuracy values as high as 99.44%. Complementary use of Illumina's short-read technology refined the team's accuracy to 99.96%, although this still trails the 99.99% 'gold standard' accuracy that assembly projects typically strive for.

In other aspects of human genome analysis, however, nanopores really could excel. Human-genome assemblies are still incomplete, for instance, because highly repetitive regions are resistant to short-read analysis. A team led by genomics researcher Karen Miga at the University of California, Santa Cruz, showed that nanopores could help researchers to fill in those blanks[5]. Miga's team used 150-kilobase-pair reads to reconstruct a human centromere — the ultra-repetitive genomic stretches at the pinched waists of eukaryotic chromosomes, which previously represented a daunting void. A truly complete genome sequence could be just a few years away, predicts Akeson, who collaborated with Miga.

Nanopore analysis is also well suited for mapping epigenetic marks — minor chemical modifications to individual nucleotides that can influence gene expression. Most sequencing platforms use sample-preparation methods that erase those marks, but nanopore platforms can directly analyse modified DNA. Simpson and Winston Timp at Johns Hopkins University in Baltimore, Maryland, showed that they could train software to distinguish the electrical signatures of methylated cytosine nucleotides from normal cytosines with roughly 90% accuracy[6]. Akeson has achieved similar success[7]. "We've been able to detect any modification we've attempted to see," says Akeson. "It can distinguish differences as little as two hydrogen atoms."

MORE TO COME

Although still alone in the nanopore market, ONT faces an established long-read competitor. Pacific Biosciences (PacBio), based in Menlo Park, California, has built a reputation on generating extremely accurate data from DNA fragments measuring tens of thousands of bases. PacBio's platform is bulkier and more expensive than MinION — its smallest system, the Sequel, is roughly the size of a refrigerator, and costs $350,000. It also can't quite reach ONT's read-length extremes.

But PacBio chief scientific officer Jonas Korlach notes that the system reliably delivers average reads of 10–18 kilobases, topping out at about 100 kilobases, and PacBio is trusted in the genomics world for its high-quality data. "You can get up to 99.99% accuracy pretty easily," says Phillippy, who adds that this technology is still his go-to when building a large genome assembly for the first time. It is also faster for large-scale projects: the recently reported nanopore-only human genome assembly[4] took ten times longer than it would have on a PacBio machine, says Phillippy.

Some users have found that nanopore sample preparation kits can be unsettlingly unpredictable, with some DNA samples requiring extensive optimization. "Some people are doing spectacularly well and getting amazing results, while others just struggle," says Robison. At a presentation in December 2016, ONT chief technology officer Clive Brown announced that: "A lot of effort is being put in … to giving people debugged protocols for specific sample types that will help them optimize the yield they get."

Similarly, one of the company's biggest assets — routine reinvention and refinement to improve performance — can leave loyal users scrambling. For example, ONT's newest pore created a technical headache for customers who were familiar with the older technology. "That sort of thing has happened to us several times now," says Loman. "That's the flip side of being on the cutting edge — things don't last quite as long as you'd hope."

These issues represent opportunities for competitors. Farthest along is Roche, headquartered in Basel, Switzerland, which acquired the California-based nanopore start-up Genia Technologies in 2014. Roche remains secretive about its system, but a 2016 publication from Genia describes a "nanopore sequencing by synthesis" strategy in which a DNA-synthesizing enzyme is coupled to a protein nanopore[8]. The enzyme reads the target DNA, building a complementary sequence from chemically tagged nucleotides. As each base is incorporated into the growing strand, its tag is released, passing through the nanopore to produce a distinctive electrical signal.

Neil Gunn, head of sequencing solutions at Roche, says that although that research predates Roche's acquisition, the core principles of the technology are largely unchanged. "That's very much in line with the design of the product," he says. "Since then, we've proceeded to focus on improving accuracy, reading rate and the sequencing length that we can acquire." Gunn notes that Roche's platform will be squarely targeted at the *in vitro* diagnostics space, with the goal of surpassing competitors' accuracy and reproducibility. Robison sees Roche's platform as a potentially strong contender, with early publications indicating accuracy ranging from 78% to 99% for any given nucleotide. "Their device could be very interesting, but the devil is always in the details," he says. "We'd need to see real data at a large scale."

But ONT is not resting on its laurels: its two latest bench-top systems can deliver much greater data volumes than previous models. GridION, released in March, essentially runs multiple MinION devices in parallel. By contrast, PromethION uses an entirely different type of flow cell, and is designed for human-genome-scale projects. "They've obviously targeted it at as something that could be competitive with the Illumina platform in terms of output," says Loman. Instruments have already shipped to 'early access' users, but no data have yet been publicly released.

However these developments pan out, nanopore sequencing is undeniably ascendant. And its promise of low-cost, reliable sequencing for the masses has researchers excited. "As a computer scientist, I'm always hungry for data," says Phillippy. "The thought of all of these microbiology labs and all of these college classrooms being able to generate sequence data is tantalizing." ∎

Michael Eisenstein *is a freelance science writer based in Philadelphia, Pennsylvania.*

1. Quick, J. *et al. Nature* **530,** 228–232 (2016).
2. Faria, N. R. *et al. Nature* **546,** 406–410 (2017).
3. Cao, M. D. *et al. GigaScience* http://dx.doi.org/10.1186/s13742-016-0137-2 (2016).
4. Jain, M. *et al.* Preprint at bioRxiv http://dx.doi.org/10.1101/128835 (2017).
5. Jain, M. *et al.* Preprint at bioRxiv http://dx.doi.org/10.1101/170373 (2017).
6. Simpson, J. T. *et al. Nature Meth.* **14,** 407–410 (2017).
7. Rand, A. C. *et al. Nature Meth.* **14,** 411–413 (2017).
8. Fuller, C. W. *et al. Proc. Natl Acad. Sci. USA* **113,** 5233–5238 (2016).

"It can distinguish differences as little as two hydrogen atoms."