

# Nanopores: a sequencer in your backpack

Vivien Marx

It's been a long wait for nanopore sequencing technology. Over 1,000 labs are testing the first commercial device and publishing results. Researchers tell *Nature Methods* about their experiences putting these early instruments through their paces. And more technology is in the works.

Nanopore sequencing instruments are up and running, delivering data in labs eager to advance their projects with this single-molecule technology<sup>1–3</sup>. The data published as *Nature Methods* goes to press are from the only commercially launched device, the MinION, developed by Oxford Nanopore Technologies (ONT). It is a sequencer that fits in the palm of a hand.

Through the company's early access program, this device is now in more than 1,000 labs. Other nanopore devices are in development in labs and behind closed doors at Roche, through its recent acquisition of Genia, and, separately, at Quantapore.

Nanopore sequencing technology has been emerging in fits and starts<sup>4–6</sup>. The technology identifies a nucleic acid sequence by threading a molecule through a pore a few nanometers in diameter. That pore might be a protein such as  $\alpha$ -hemolysin that is embedded in a polymer membrane or a hole formed in a solid material such as silicon nitride (SiN)<sup>7</sup>. Voltage is applied across the membrane, creating an ionic current and an electrophoretic force that pulls the DNA through the opening. As the molecule zips through, it causes telltale fluctuations in the current that are specific to different DNA sequences. The technology can also be used to analyze RNA and proteins.

Nanopore sequencing conjures futuristic scenarios. Small sequencers could become ubiquitous sensors that test the environment for pathogens, for example. Plenty of challenges remain, including the need for higher accuracy, increased throughput and more data-analysis software, but the technology is rapidly evolving.

These sequencers can process DNA fragments that are 10,000–50,000 bases—and potentially 100,000 bases—in length.

As Clive Brown, ONT's chief technology officer, said in a recent company presentation, a nanopore read can be as long as the DNA fragment researchers prepare.

These long reads are unlike the output of most high-throughput sequencers, which produce many shorter snippets that must then be computationally glued together.

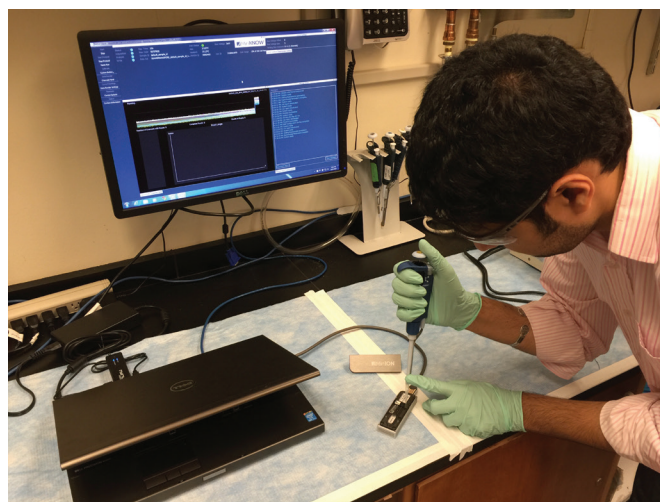
Longer reads help researchers build complete or near-complete assemblies more quickly.

The portable nanopore sequencer also opens up a slew of potential applications in the field, such as rapid surveillance of the flu, malaria, tuberculosis or Ebola, says Nicholas Loman of the University of Birmingham. "That's why we are excited about it," he says.

## Epidemic updates

In a few hours, Loman and colleagues used the MinION to identify the species of *Salmonella* responsible for an outbreak that had closed eight wards in a local hospital. They later confirmed results with sequencing performed on the Illumina MiSeq instrument.

In a different project, Joshua Quick, a researcher in Loman's lab, took three nanopore sequencers to Guinea to sequence the genomes of Ebola patients and to help



UCSC graduate student Miten Jain loads the MinION, a nanopore sequencer.

Hugh Olsen, UCSF

with efforts to monitor the outbreak using PCR-based diagnostics. Quick transmitted nanopore sequence data to Loman in Birmingham, who analyzed them and returned results both to Quick and to the World Health Organization to guide treatment. The project continues, says Loman. Despite a drop in case numbers, a few stubborn chains of viral transmission still need sorting out.

Sequencing the Ebola virus with the MinION takes around half an hour, says Loman. "It takes longer to make the libraries than it does to do the sequencing," he says. Together with Aaron Quinlan at the University of Virginia, he has developed software called poretools that converts MinION data into FASTA or FASTQ files for further analysis with traditional sequence-alignment or genome-assembly software. There is currently a flurry of software development for analyzing nanopore sequence data (**Box 1**).

## BOX 1 NANOPORE SOFTWARE TOOLS

A budding software-development community has sprouted around users' needs for the new technology. Here are some tools.

**GraphMap**: mapper for long, error-prone reads

**LINKS**: software to scaffold long reads

**marginAlign**, **marginCaller** and **marginStats**: tools to align nanopore reads to a reference genome

**MinoTour**: real-time analysis tools

**nanoCORR**: error-correction tool for nanopore sequence data

**Nano0k**: Software for pre- and post-alignment analysis of nanopore sequencing data, quality and error profiles

**nanopolish**: nanopore sequence analysis and genome assembly software

**nanopore**: variant-detection tool for nanopore sequence data

**nanocorrect**: error-correction tool for nanopore sequence data

**npReader**: real-time conversion and analysis of nanopore reads

**poRe**: tool for analyzing and visualizing nanopore sequence data

**poreSeq**: error-correction and variant-calling software for nanopore sequence data

**poretools**: nanopore sequence analysis and visualization software

**The Burrows-Wheeler Aligner** (BWA) and **LAST** have been tuned to work with nanopore reads.

Rapid sequence analysis helps map the relatedness of Ebola cases, says Loman. A virus might be transmitted from parent to child or across a village; rapid sequencing and analysis can reveal chains of transmission. During an outbreak a virus will also mutate as a result of selection pressure, which can cause diagnostic PCR primers to fail and lead to false negative results. Software such as poretools can help determine variants relative to a viral reference sequence generated early in an outbreak. The MinION has applications in the lab too, he says. For instance, it can help to quickly authenticate a cell line before the start of an experiment.

Nanopore sequencing directly analyzes a sample. "All other sequencing devices make copies, which masks modifications and doesn't work at all for proteins," says Mark Akeson of the University of California at Santa Cruz (UCSC), who codeveloped the technology and is an ONT consultant. In his work, accuracy with the MinION now reaches 92% for a single-molecule 2D read. The one systematic bias he has found relates to homopolymers longer than five bases, such as the sequence TTTTTT. "There are several ways to overcome this, one of which we are working on at UCSC," he says.

Benedict Paten, also at UCSC, anticipates that this technology will be able to read out the sequence of bases and their individual methylation statuses, too, which could potentially help to discern haplotypes.

### A pore in action

In the ONT approach to sequencing, an engineered protein nanopore is embedded in a proprietary polymer membrane. As a DNA molecule passes through the pore, it disrupts the current in a sequence-specific way.

To stop the DNA from zipping through too quickly, an enzyme is bound to one end, which ratchets the DNA through the pore. When a hairpin molecule is attached to the other end of the DNA, both strands can be sequenced for higher accuracy, an approach the company calls 2D sequencing. ONT's sequencer currently uses a proprietary pore protein called R7 and a ratcheting enzyme, or motor, named E5. The company continues to develop chemistries.

ONT's MinION, its first device, plugs into a computer's USB 3.0 port. The device has a flow cell with an array of 512 nanopores. Each pore produces one read at a time. When a sample is added to the device, it flows across the array, DNA enters a number of pores, the built-in electronics measure the current changes, and the instrument's control software processes those measurements.

The company is planning to beta-test its next sequencer, the PromethION, through an early access program. This benchtop system can fit 48 flow cells, each with 3,000 nanopores, which adds up to a total of 144,000 nanopores. Each flow cell can be run separately. This high-throughput device will

read a huge number of bases per day, says Akeson. "Will it compete with high-end Illumina devices? Possibly." The comparison depends on factors such as the number of reads a lab needs to achieve consensus accuracy, cost and any time advantage that stems from assembling 10-kb reads versus 400-base reads, he says.

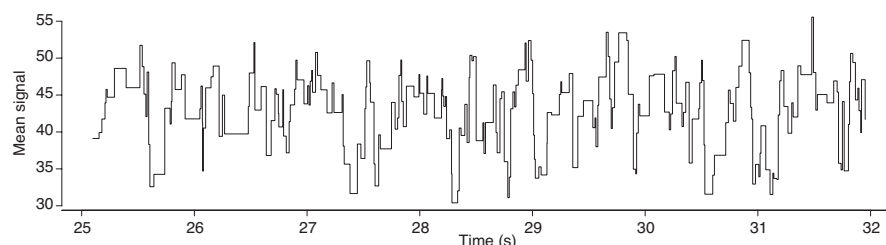
One technology need not necessarily replace another. "You use several technologies together to get to the best answer," says Akeson, which might mean, for example, combining Illumina short-read sequencers, molecular barcoding with 10X Genomics and an ONT device. He believes that ONT's advantages over time will be that it offers small, inexpensive devices for work in the field and that the instruments can also directly read RNA and analyze proteins.

To accelerate sequencing throughput, ONT is also developing a 'fast mode' for its flow cells, with the tradeoff of a slight drop in accuracy. The company has also now separated the flow cell and sensor array from the reusable electronics; that application-specific integrated circuit will be used in the PromethION and in the next generation of the MinION. There are also plans to speed up the chemistry such that the instrument can detect and sequence 1,000 bases per second per pore.

### Nanopore output

The instrument's rawest of raw data are numerical values that reflect the disruption of current by a molecule passing through a nanopore, says The Roslin Institute's Mick Watson, who has been working with the MinION. The instrument's built-in software determines which signals are noise and which are 'events' that can be interpreted as base sequence. "The software needs to decide when an event has occurred," says Watson. Data about events are delivered to users in the form of a squiggle plot.

ONT has been studying the squiggle patterns, says Watson, and developed a statistical model based on a string of five to six bases, to reliably match events to sequence. Watson likes that ONT's technology measures the actual DNA strand, and not the incorporation of a particular label as many other technologies do. This measures "the full noise in DNA space," he says, detecting A's, C's, G's and T's, also in their many varied forms. "That comes out as noise and inaccuracy when in actual fact, it's measuring something that we just don't know about yet," says Watson.



The MinION's raw data are a 'squiggle plot', numerical values that reflect the way current is disrupted when a molecule passes through the nanopore.

Nanopore sequencing could help epigenetics researchers in new ways. They have been particularly interested in detecting methylated DNA bases and want to measure many types of such changes that current technology cannot measure directly, says Watson. For example, bisulfite sequencing cannot distinguish between methyl cytosine and 5-hydroxymethylcytosine. With the MinION, Watson says, these changes have different signals.

### Wish list

Nanopore sequencers still need higher accuracy base calling. As a single-molecule technology, it's inevitably noisier than approaches with clones, says Loman. He looks forward to ONT's automated sample-preparation cartridge called VolTRAX, which is in development and which docks onto the sequencer. "There's no point in having a portable sequencer that you can have in your backpack if you are still going to need a bunch of lab equipment that needs power and things like that," he says.

ONT had some manufacturing issues as it rolled out the MinION to early access testers, which led to uneven results in many labs, including Loman's. The issues were hard to troubleshoot because researchers couldn't be sure of the source of errors. These challenges are becoming less pronounced, says Loman, but he is concerned that some researchers may have been too quickly discouraged after trying earlier versions of the device.

To date, MinION sequence data have not been of satisfactory quality for Stephen Richards at Baylor College of Medicine, whose research focus is insect genomics. His collaborator Charles Brockhouse at Creighton University has had the MinION for a while, and together they have been sequencing the blackfly genome. They have also struggled with extracting DNA of adequate quality.

Richards now has his own MinION. To decide how he will use it, he will be running

control DNAs from a test suite of species: three bacterial species, chosen for their differing GC contents, and, down the line, fruit fly and human DNA.

Richards previously tried multiple sequencing platforms and settled on PacBio sequencers because they let him obtain an entire bacterial genome without gaps. Average PacBio read lengths are around 15 kb, with many reads longer than 20 kb, says Richards. Short-read sequencers often lead to fragmentary genome assemblies, he says, but long-read technologies help to build higher quality assemblies, which also lead to better annotations.

The ONT device can potentially deliver 100-kb reads, says Richards, which would be "absolutely transforming for reference genome assembly, and thus could be transforming for the many researchers working on species without genome references." He is glad that ONT's devices can now work with nanograms of material. Human samples from tumors, for example, can be small. And whereas a single praying mantis might yield enough DNA for sequence analysis, blackflies need to be sequenced in bulk in order to obtain enough starting material; this runs into the problem of genomic sequence variation between individuals.

At first the MinION was rolled out to non-core facility labs, and sequencing centers were side-stepped. But ONT says that the device is now in labs of all sizes, and company labs, too. There have been a series of kit, chemistry and software updates.

In Paten's view, changes such as the shift from version R6 to R7.3 chemistry have improved data quality, read length and throughput, which are the three desirables in sequencing that matter most in combination. "Within the bacterial genomics community, I think it is starting to have a broad impact," he says of the technology. Next, the technology must prove itself to the wider research community.

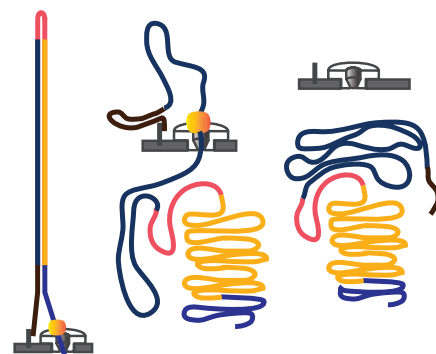
### New tech, new tools

For analysis of nanopore sequencing data, ONT spun out a company called Metrichor that runs proprietary software on the cloud. MinION users can analyze their data through Metrichor or with their own software. More generally, Metrichor targets citizen scientists and so-called self-quantifiers, who might eventually use ONT's MinION to monitor their own health.

Watson has developed poRe, software that lets scientists collect statistics about sequencing runs and visualize data such as read-length histograms. It is designed to help scientists less familiar with using tools from a Linux command line, he says.

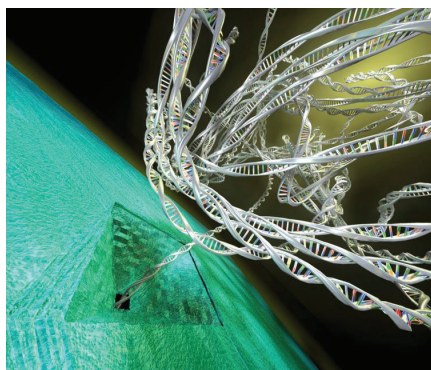
The emerging software must keep stride with the technology's rapid maturation. "We have to be very agile and be very quick about these things," says Watson. He informs users about software updates through his blog and social media. Some older software has been tweaked to work with nanopore data, such as the Burrows-Wheeler Aligner, or the sequence-alignment tool LAST, which "works really, really well on nanopore data," Watson says.

Heng Li, a researcher at the Broad Institute of Harvard and MIT, says he has not yet worked much on nanopore data analysis. He hopes for a push-button ONT assembler that can finish a bacterial genome on a single workstation or even a desktop computer. Li also imagines software able to perform real-time assembly. The assembler could refine an intermediate assembly right as the instrument delivers sequence reads. "When the assembler thinks the assembly is good enough, it lets users know so that



In nanopore sequencing, double-stranded DNA (here one strand is dark blue and the other is yellow) binds to an oligomer that is intercalated in the membrane supporting the nanopore. A 'molecular motor' (orange) recognizes a leader sequence (lighter blue) and ratchets the DNA through the pore. A hairpin (red) links the two strands.





C. Dekker lab, TU Delft/Tremari

Solid-state nanopores (here, an artistic rendering) can be made of various materials such as silicon nitride or graphene. One advantage is that the pores can be manufactured with fixed specifications and integrated with electronics and microfluidics.

they can stop sequencing and load the next sample,” he says. “This would be a killer feature if it came true.”

Watson thinks scientists might not need “base space” and could avoid base-calling altogether before alignment. “You could have a theoretical squiggle, which is your reference squiggle, and then basically align your observed squiggle to your reference squiggle and start calling differences, not in base space but in squiggle space,” he says. “Why not just work with that raw signal, which has all that rich information in there?”

The platform offers many possibilities, including a new field of bioinformatics that Loman calls real-time sequence analysis. High-throughput sequencing leads to concerns about moving and storing data, especially at sequencing centers with a daily terabase output. “The alternative approach is you don’t do that; you do all of your analysis in real time, and actually what you end up with is the answer,” he says. Sequencing could become an assay to directly answer a particular biological question.

The MinION does not produce a lot of data, but the upcoming model, the PromethION, does. “I think the real-time approach is really quite exciting,” says Loman, and it can potentially solve many data-management challenges.

### Solid-state nanopores

Some companies are exploring solid-state nanopores. At Quantapore, which recently closed a \$35 million financing round, molecular biologists, single-molecule physicists, engineers and software developers are building one such system but not divulging many details.

Unlike ONT, Quantapore is in stealth mode; it may or may not be alpha-testing instruments in academic labs. What occurs in its device in order to sequence nucleic acids is a “purely physical reaction,” says Quantapore founder Martin Huber.

Unlike the electrical approach applied by ONT, the Quantapore technology involves labeling both the nucleic acid and the pore such that the pore can deliver an optical, enzyme-free readout. The system does not need a detector for each nanopore, says Huber, so the nanopores can be parallelized.

Electrical nanopore sequencing may ultimately face throughput limitations, says Huber. For example, there can be challenges with parallelizing electrical sensors in a nanopore array, or problems with the molecular motors used to slow down the translocation. These motor proteins “can do funny things such as stop, skip bases or even go backwards,” all of which can increase the error rate, he says. He believes his company’s approach with optical nanopore sequencing circumvents such issues.

Cees Dekker at Delft University of Technology has been working on DNA translocation through nanopores since 2000. “I love the conceptual simplicity of it,” he says. “One threads a molecule head to tail through a tiny pore and thus probes its local properties.”

Together with Hagan Bayley from the University of Oxford, Dekker has developed hybrid nanopore configurations; they are currently testing instruments with both alpha-hemolysin and solid-state nanopores. As Dekker explains, the combination delivers the best of both worlds. “Certainly, protein pores are in the lead regarding DNA sequencing,” he says. In his view, solid-state nanopores have strengths because of how they can be integrated with electronics and microfluidics, owing to their overall material stability and the option to choose and build a desired pore width. It is also possible to run many of them in parallel, and they can be integrated into microchip-based devices using existing manufacturing approaches.

Solid-state pores are often made of SiN and are around 5–30 nanometers thick. Dekker has also been working with graphene pores, which can have a membrane that is just one atom thin. Graphene is a conducting material, which opens up new measurement modalities, he says.

And with plasmonic nanopores, Raman spectroscopy can be used to detect what passes through them. Separately, Dekker and

his team are also working on various ways to modulate the speed with which DNA passes through the pore.

The nanopore field has been driven by its prospects in the world of DNA sequencing, which is a major application, says Dekker. But his emphasis has been on using nanopores as biophysical tools, for example, to probe the presence of proteins or knots along DNA, to manipulate DNA in the pore or to measure physical forces within a molecule.

### Dreamers welcome

Nanopore technology invites dreaming. Yaniv Erlich, a computational biologist at Columbia University and the New York Genome Center, believes that miniaturized nanopore sequencers such as the MinION usher in an era of ubiquitous sequencing, with sequencing being an integrated part of sensors<sup>8</sup>. “DNA-aware home appliances” could monitor drinking water for pathogens, for example, and a “smart toilet system” could collect information about the microbiomes of household members.

Watson says that one day chicken coops could be equipped with embedded sequencing to monitor animals’ health and dietary needs. Nanopore sequencers could detect antimicrobial resistance in hospitals or pathogens in sewage pipes and set off alerts. It might all sound futuristic, he says, but a decade ago, a USB-driven sequencer around the size of a smartphone, such as ONT’s MinION, seemed an unreasonable idea.

The MinION’s size and portability enable a new class of applications that bring sequencing to venues such as hospitals, factories, many stages of food production and agricultural arenas, says Paten. “I used to think that was long-off science fiction, but not anymore.”

1. Quick, J., Quinlan, A. & Loman, N. *Gigascience* **3**, 22 (2014).
2. Goodwin, S. *et al. bioRxiv* doi:10.1101/013490 (2015).
3. Jain, M. *et al. Nat. Methods* **12**, 351–356 (2015).
4. Kasianowicz, J.J. *et al. Proc. Natl. Acad. Sci. USA* **93**, 13770–13773 (1996).
5. Branton, D. *et al. Nat. Biotechnol.* **26**, 1146–1153 (2008).
6. Venkatesan, B.M. & Bashir, R. *Nat. Nanotechnol.* **6**, 615–624 (2011).
7. Muthukumar, M., Calin, P. & Dekker, C. *Phys. Today* **68**, 40 (2015).
8. Erlich, Y. *bioRxiv* doi:10.1101/019018 (2015).

Vivien Marx is technology editor for *Nature* and *Nature Methods* (v.marx@us.nature.com).