

OPINION

Open Access



# The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community

Miten Jain , Hugh E. Olsen, Benedict Paten and Mark Akeson\*

## Abstract

Nanopore DNA strand sequencing has emerged as a competitive, portable technology. Reads exceeding 150 kilobases have been achieved, as have in-field detection and analysis of clinical pathogens. We summarize key technical features of the Oxford Nanopore MinION, the dominant platform currently available. We then discuss pioneering applications executed by the genomics community.

## Introduction

Nanopore sequencing was pioneered by David Deamer at the University of California Santa Cruz, and by George Church and Daniel Branton (both at Harvard University). Beginning in the early 1990s, academic laboratories reached a series of milestones towards developing a functional nanopore sequencing platform (reviewed in [1, 2]). These milestones included the translocation of individual nucleic acid strands in single file order [3], processive enzymatic control of DNA at single-nucleotide precision [4], and the achievement of single-nucleotide resolution [5, 6].

Several companies have proposed nanopore-based sequencing strategies. These involve either: the excision of monomers from the DNA strand and their funneling, one-by-one, through a nanopore (NanoTag sequencing (Genia), Bayley Sequencing (Oxford Nanopore)); or strand sequencing wherein intact DNA is ratcheted through the nanopore base-by-base (Oxford Nanopore MinION). To date, only MinION-based strand sequencing has been successfully employed by independent genomics laboratories. Where possible, this review focuses on peer-reviewed research performed using the MinION [1, 7–38].

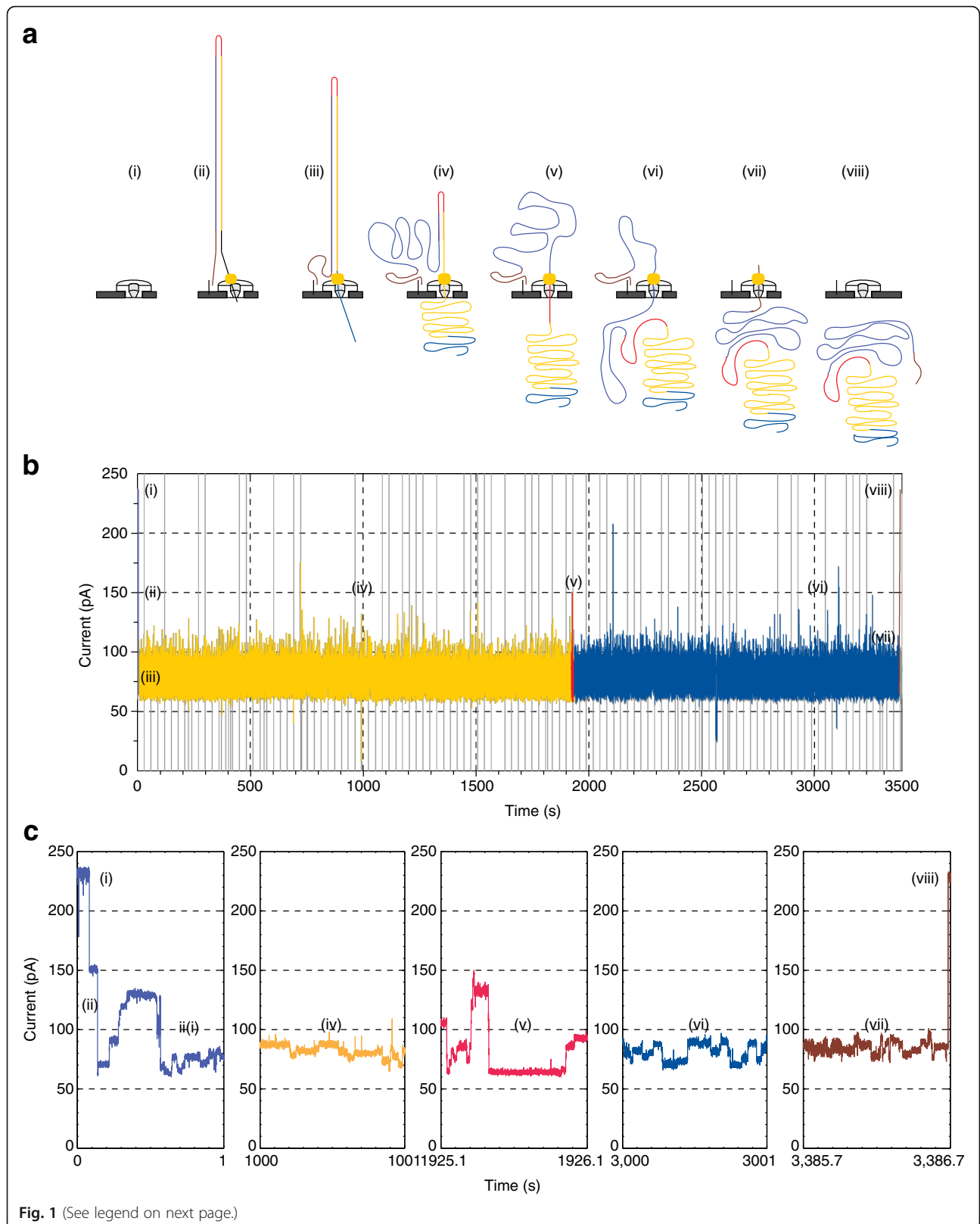
## DNA strand sequencing using the Oxford Nanopore MinION

Oxford Nanopore Technologies (ONT) licensed core nanopore sequencing patents in 2007, and began a strand sequencing effort in 2010 [2]. At the Advances in Genome Biology and Technology (AGBT) 2012 conference, Clive Brown (Chief Technical Officer of ONT) unveiled the MinION nanopore DNA sequencer, which was subsequently released to early-access users in April 2014 through the MinION Access Program (MAP).

The MinION is a 90-g portable device. At its core is a flow cell bearing up to 2048 individually addressable nanopores that can be controlled in groups of 512 by an application-specific integrated circuit (ASIC). Prior to sequencing, adapters are ligated to both ends of genomic DNA or cDNA fragments (Fig. 1). These adapters facilitate strand capture and loading of a processive enzyme at the 5'-end of one strand. The enzyme is required to ensure unidirectional single-nucleotide displacement along the strand at a millisecond time scale. The adapters also concentrate DNA substrates at the membrane surface proximal to the nanopore, boosting the DNA capture rate by several thousand-fold. In addition, the hairpin adapter permits contiguous sequencing of both strands of a duplex molecule by covalently attaching one strand to the other. Upon capture of a DNA molecule in the nanopore, the enzyme processes along one strand (the 'template read'). After the enzyme passes through the hairpin, this process repeats for the complementary strand (the 'complement read').

As the DNA passes through the pore, the sensor detects changes in ionic current caused by differences in the shifting nucleotide sequences occupying the pore. These ionic current changes are segmented as discrete events that have an associated duration, mean amplitude, and variance. This sequence of events is then interpreted computationally as a sequence of 3–6 nucleotide long kmers ('words') using graphical models.

\* Correspondence: makeson@soe.ucsc.edu  
UC Santa Cruz Genomics Institute and Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA



(See figure on previous page.)

**Fig. 1** Data for a 2D read of a full-length  $\lambda$  phage dsDNA from the MinION nanopore sequencer. **a** Steps in DNA translocation through the nanopore: (i) open channel; (ii) dsDNA with lead adaptor (*blue*), bound molecular motor (*orange*) and hairpin adaptor (*red*) is captured by the nanopore; capture is followed by translocation of the (iii) lead adaptor, (iv) template strand (*gold*), (v) hairpin adaptor, (vi) complement strand (*dark blue*) and (vii) trailing adaptor (*brown*); and (viii) status returns to open channel. **b** Raw current trace for the passage of the single 48-kb  $\lambda$  dsDNA construct through the nanopore. Regions of the trace corresponding to steps i–viii are labeled. **c** Expanded time and current scale for raw current traces corresponding to steps i–viii. Each adaptor generates a unique current signal used to aid base calling

The information from template and complement reads is combined to produce a high-quality ‘2D read’, using a pairwise alignment of the event sequences.

An alternate library preparation method does not use the hairpin to connect the strands of a duplex molecule. Rather, the nanopore reads only one strand, which yields template reads. This allows for higher throughput from a flow cell, but the accuracy for these ‘1D reads’ is slightly lower than that of a ‘2D read’.

### Benefits of MinION compared to other next generation sequencing platforms

#### Detection of base modifications

Next generation sequencing (NGS) technologies do not directly detect base modifications in native DNA. By contrast, single-molecule sequencing of native DNA and RNA with nanopore technology can detect modifications on individual nucleotides. Previously, Schreiber et al. [39] and Wescoe et al. [40] demonstrated that a single-channel nanopore system can discriminate among all five C-5 variants of cytosine (cytosine (C), 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC)) in synthetic DNA. The discrimination accuracies ranged from 92 to 98% for a cytosine of interest in a background of known sequences [40].

In 2016, two research groups independently demonstrated that MinIONs can detect cytosine methylation in genomic DNA [41, 42]. Rand et al. [41] developed a probabilistic method that combines a pair hidden Markov model (HMM) and a hierarchical Dirichlet process (HDP) mixture of normal distributions. They performed a three-way classification among C, 5-mC, and 5-hmC with a median accuracy of 80% in synthetic DNA [41]. Simpson et al. [42] performed a similar study in which they trained an HMM to perform a two-way classification among C and 5-mC, with 82% accuracy in human genomic DNA.

#### Real-time targeted sequencing

There are significant advantages to acquiring and analyzing DNA or RNA sequences in a few hours or less, especially for clinical applications. This is difficult using conventional NGS platforms, but relatively straightforward using the MinION because of its size, cost, simple library prep, and portability (see [14]). Beyond this, the

MinION platform permits real-time analysis because individual DNA strands are translocated through the nanopore, allowing decisions to be made during the sequencing run.

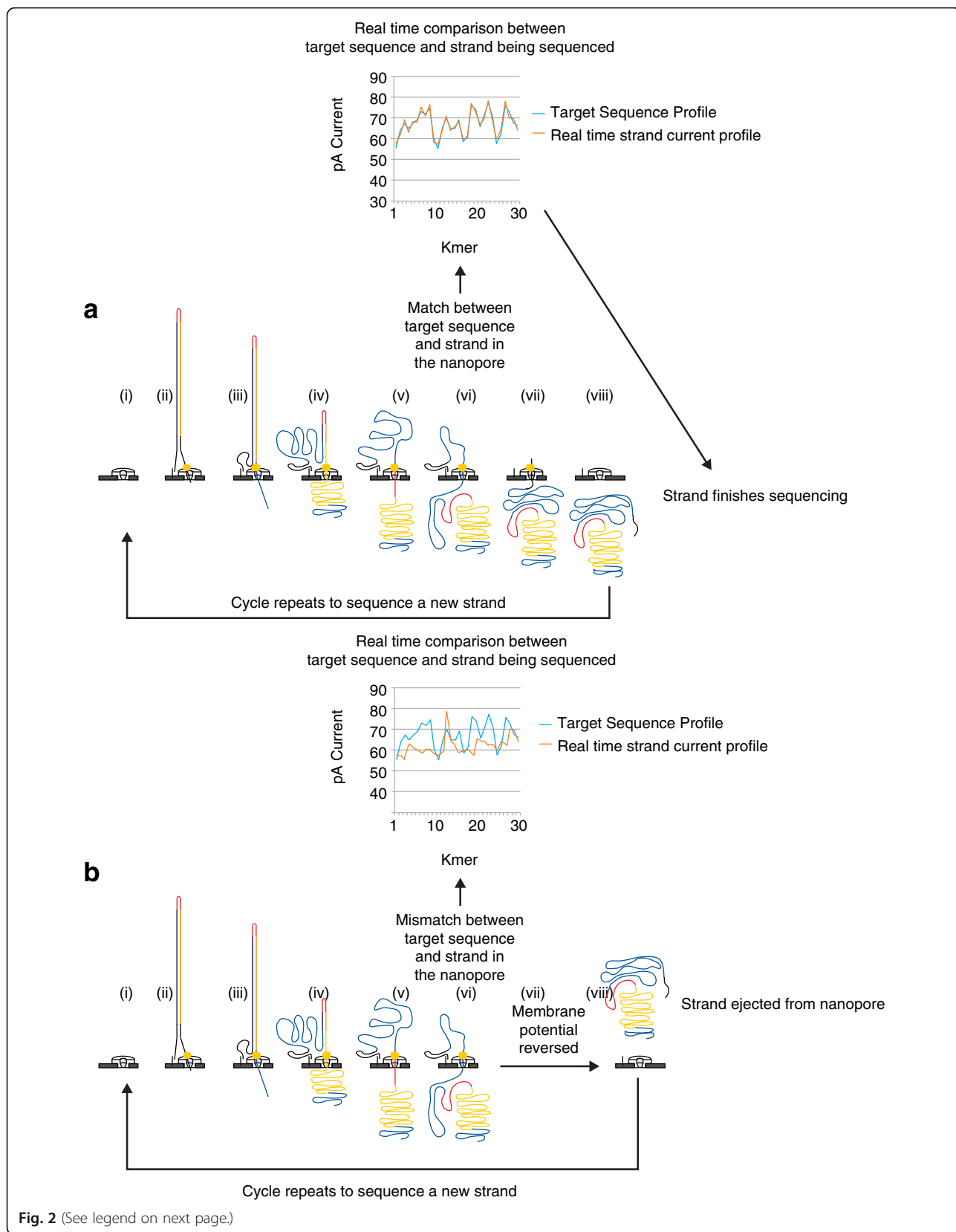
This real-time utility of MinION was first demonstrated by Loose et al. [43] in a manuscript that described targeted enrichment (‘Read Until’) of 5 and 10 kb regions from phage lambda double-stranded DNA (dsDNA). Briefly, a mixture of DNA fragments is applied to the MinION flow cell. While a DNA strand is captured and processed in the nanopore, the resulting event levels are aligned against the expected pattern for a target sequence. If the pattern matches, the sequencing continues (Fig. 2a). If the pattern does not match, the DNA strand is ejected from the nanopore so that a subsequent DNA strand can be captured and analyzed (Fig. 2b). In doing this, reads of the targeted strand are rapidly accumulated relative to the DNA strand population as a whole. ‘Read Until’ demonstrates how MinION sequencing could significantly reduce the time required from biological sampling to data inference, which is pertinent for in-field and point-of-care clinical applications.

#### Extending read lengths using the MinION

A virtue of nanopore DNA strand sequencing is read lengths that substantially exceed those of dominant NGS platforms. For example, 1D reads over 300 kb in length and 2D reads up to 60 kb in length have been achieved using *Escherichia coli* genomic DNA [44]. To demonstrate utility, Jain et al. [9] used 36-kb + MinION reads to resolve a putative 50-kb gap in the human Xq24 reference sequence. Previously, this gap in the reference sequence could not be completed because it contained a series of 4.8-kb-long tandem repeats of the cancer-testis gene CT47. This work established eight CT47 repeats in this region (Fig. 3).

#### Detection of structural variants

Mistakes arising in assemblies of 450-base-long NGS reads are also problematic when characterizing structural variants in human genomes. The problem is acute in cancer, where examples of copy number variants, gene duplications, deletions, insertions, inversions, and translocations are common. For reads that averaged 8 kb in length, Norris et al. [45] used the MinION to detect



(See figure on previous page.)

**Fig. 2** 'Read Until' strategy for selective sequencing of dsDNA molecules. The ionic current profile obtained during translocation of a DNA strand through the nanopore is compared in real time to the ionic current profile of a target sequence. **a** As sequencing of the template strand of DNA proceeds (during step iv), the measured current is compared to the reference current profile. If there is a match, sequencing of that strand continues to completion (steps v–vii). A new strand can now be captured. **b** Alternatively, if the measured current does not match the reference current profile, the membrane potential is reversed, sequencing of that strand stops, and the strand is ejected (at stage v). A new strand can now be captured. (Image based on the strategy of Loose et al. [43])

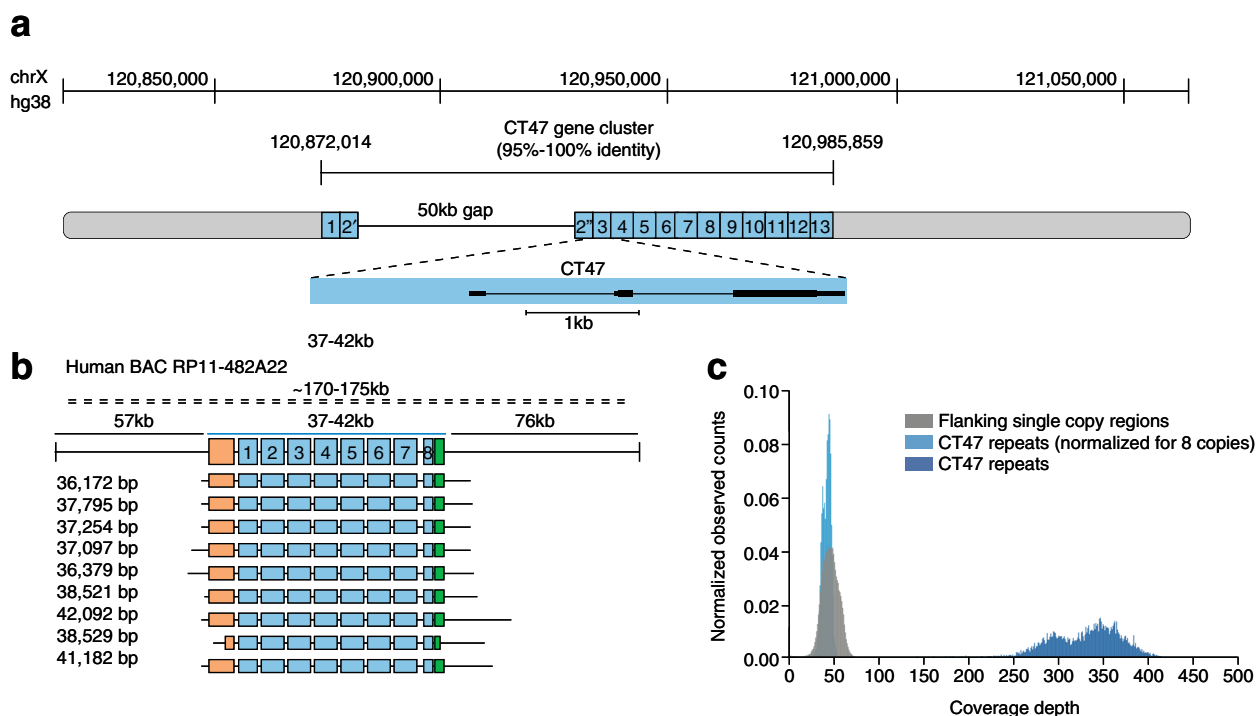
structural variants in a pancreatic cancer cell line. These authors concluded that the MinION allowed for reliable detection of structural variants with only a few hundred reads compared to the millions of reads typically required when using NGS platforms.

**RNA expression analysis**

RNA expression analysis is most often performed by NGS sequencing of cDNA copies. A drawback of this strategy is that the reads are relatively short, thus requiring assembly of cDNA reads into full-length transcripts. This is an issue for the accurate characterization of RNA splice isoforms because there

is often insufficient information to deconvolute the different transcripts properly. Full-length cDNA reads would avoid this problem and can be executed with either the PacBio or MinION platforms.

To illustrate, Bolisetty et al. [8] used the MinION to determine RNA splice variants and to detect isoforms for four genes in *Drosophila*. Among these is *Dscam1*, the most complex alternatively spliced gene known in nature, with 18,612 possible isoforms ranging in length from 1806 bp to 1860 bp [8]. They detected over 7000 isoforms for *Dscam1* with >90% alignment identity. Identifying these isoforms would be impossible with 450-base-long NGS reads.



**Fig. 3** Estimate CT47-repeat copy-number on human chromosome Xq24. **a** BAC end sequence alignments (RP11-482A22: AQ630638 and AZ517599) span a 247-kb region, including 13 annotated CT47 genes [69] (each within a 4.8-kb tandem repeat), and a 50-kb scaffold gap in the GRCh38/hg38 reference assembly. **b** Nine MinION reads from high molecular weight BAC DNA span the length of the CT47-repeat region, providing evidence for eight tandem copies of the repeat. The insert (dashed line), whose size is estimated from pulse-field gel electrophoresis, with flanking regions (black lines) and repeat region (blue line) are shown. Single-copy regions before and after the repeats are shown in orange (6.6 kb) and green (2.6 kb), respectively, along with repeat copies (blue) and read alignment in flanking regions (gray). The size of each read is shown to its left. **c** Shearing BAC DNA to increase sequence coverage provided copy-number estimates by read depth. All bases not included in the CT47 repeat unit are labeled as flanking regions (gray distribution; mean of 46.2-base coverage). Base coverage across the CT47 repeats was summarized over one copy of the repeat to provide an estimate of the combined number (dark blue distribution; mean of 329.3-base coverage) and was similar to single-copy estimates when normalized for eight copies (light blue distribution; mean of 41.15-base coverage). (Figure reproduced from Jain et al. [9])

### Bioinformatics and platform advances

The first manuscript to discuss MinION performance was based on limited data and ill-suited analysis, and thus yielded misleading conclusions about the platform's performance [24]. Over the subsequent 9-month period, ONT optimized MinION sequencing chemistry and base-calling software. Combined with new MinION-specific bioinformatics tools (Table 1), these refinements improved the identity of sequenced reads, that is, the proportion of bases in a sequencing 'read' that align to a matching base in a reference sequence, from a reported 66% in June 2014 [9] to 92% in March 2015 [44]. Links to these tools are provided in Table 1 and highlighted in the sections that follow.

### De novo base-calling

The base-calling for MinION data is performed using HMM-based methods by *Metrichor*, a cloud-based computing service provided by ONT. *Metrichor* presently requires an active internet connection [46, 47] and is a closed source. However, its base-calling source code is now available to registered MinION users under a developer license. To create a fully open-source alternative, earlier in 2016, two groups independently developed base-callers for MinION data. *Nanocall* [46] is an HMM-based base-caller that performs efficient 1D base-calling locally

without requiring an internet connection at accuracies comparable to *Metrichor*-based 1D base-calling. *DeepNano* [47], a recurrent neural network framework, performs base-calling and yields better accuracies than HMM-based methods. Being able to perform local, offline base-calling is useful when performing in-field sequencing with limited internet connectivity [30].

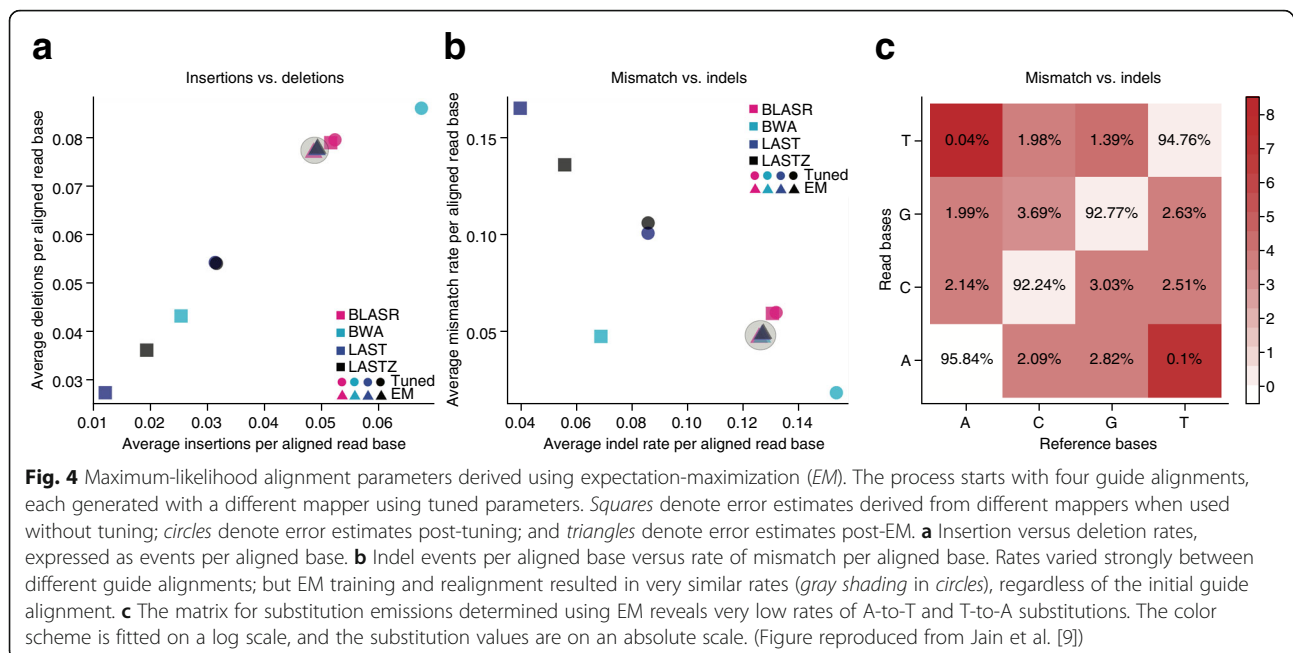
### Sequence alignment

When the MAP began, the first attempts at aligning MinION reads to reference sequences used conventional alignment programs. Most of these are designed for short-read technologies, such as the 250-nucleotide highly accurate reads produced by the Illumina platform. Not surprisingly, when applied to lower accuracy 10-kb MinION reads, these aligners disagreed in their measurement of read identity and sources of error, despite parameter optimization (Fig. 4). *MarginAlign* was developed to improve alignments of MinION reads to a reference genome by better estimating the sources of error in MinION reads [9]. This expectation-maximization-based approach considerably improves mapping accuracy, as assayed by improvements in variant calling, and yielded a maximum likelihood estimate of the insertion, deletion, and substitution errors of the reads (Fig. 4). This

**Table 1** Software tools developed specifically for MinION sequence data; there are existing tools that can also be made to work with nanopore data (not shown)

Name	Applications	Link
Poretools [22]	Sequence data extraction and statistics	<a href="https://github.com/arq5x/poretools">https://github.com/arq5x/poretools</a>
poRe [37]	Sequence extraction and basic statistics	<a href="https://sourceforge.net/projects/rpore/">https://sourceforge.net/projects/rpore/</a>
BWA MEM [49]	Sequence alignment	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
LAST [48]	Sequence alignment	<a href="http://last.cbrc.jp/">http://last.cbrc.jp/</a>
NanoOK [20]	Sequence alignment, statistics, and visualization	<a href="https://documentation.tgac.ac.uk/display/NANOOK/">https://documentation.tgac.ac.uk/display/NANOOK/</a>
marginAlign [9]	Sequence alignment, SNV calling, and statistics	<a href="https://github.com/benedictpaten/marginAlign">https://github.com/benedictpaten/marginAlign</a>
Nanopolish [50]	Signal alignment and SNV calling	<a href="https://github.com/jts/nanopolish">https://github.com/jts/nanopolish</a>
GraphMap [12]	Sequence alignment and SNV calling	<a href="https://github.com/isovic/graphmap">https://github.com/isovic/graphmap</a>
minimap	Fast approximate mapping	<a href="https://github.com/lh3/minimap">https://github.com/lh3/minimap</a>
miniasm	De novo assembly	<a href="https://github.com/lh3/miniasm">https://github.com/lh3/miniasm</a>
CANU [70]	De novo assembly	<a href="https://github.com/marbl/canu">https://github.com/marbl/canu</a>
Nanocorrect [48]	De novo assembly	<a href="https://github.com/jts/nanocorrect">https://github.com/jts/nanocorrect</a>
PoreSeq [53]	De novo assembly and SNV calling	<a href="https://github.com/tszalay/poreseq">https://github.com/tszalay/poreseq</a>
NaS [23]	De novo assembly	<a href="https://github.com/institut-de-genomique/NaS">https://github.com/institut-de-genomique/NaS</a>
Nanocorr [13]	De novo assembly	<a href="https://github.com/jgurtowski/nanocorr">https://github.com/jgurtowski/nanocorr</a>
Mash [71]	Species identification and fast approximate alignments	<a href="https://github.com/marbl/mash">https://github.com/marbl/mash</a>
minoTour [72]	Real-time data analysis	<a href="https://github.com/minoTour/minoTour">https://github.com/minoTour/minoTour</a>
Read Until [43]	Selective sequencing	<a href="https://github.com/mattloose/RUscripts">https://github.com/mattloose/RUscripts</a>
Nanocall [46]	Local base-calling	<a href="https://github.com/mateidavid/nanocall">https://github.com/mateidavid/nanocall</a>
DeepNano [47]	Recurrent neural network (RNN)-based base-calling	<a href="https://bitbucket.org/vboza/deepnano">https://bitbucket.org/vboza/deepnano</a>

SNV single nucleotide variant



was later used by a MAP consortium to achieve a 92% read accuracy for the *E. coli* k12 MG1655 genome [44].

MarginAlign refines alignments generated by a mapping program, such as LAST [48] or BWA mem [49], and is therefore reliant on the accuracy of the initial alignment. GraphMap [12] is a read mapper that employs heuristics that are optimized for longer reads and higher error rates. In their study, Sović et al. [12] demonstrated that GraphMap had high sensitivity (comparable to that of BLAST) and that GraphMap's estimates of error rates were in close agreement with those of marginAlign.

### De novo assembly

The current error profile of MinION reads makes them largely unsuitable for use with de novo assembly methods that are designed for short reads, such as de Bruijn graph-based methods. This is principally for two reasons. First, these methods rely on a sufficient fraction of all possible k-mers sequenced being reconstructed accurately; the overall indel and substitution error rates produced by MinION are unlikely to meet this demand. Second, de Bruijn graphs, in their structure, do not exploit the longer-read information generated by the MinION. Instead, nanopore sequencing is helping to mark a return to overlap-consensus assembly methods [50], a renaissance that largely started with the earlier advent of SMRT sequencing [51]. Overlap-consensus methods were principally developed for lower-error-rate Sanger-based sequencing, and so novel strategies are required to error correct the reads before they are assembled. The first group to demonstrate this approach

achieved a single contig assembly of the *E. coli* K-12 MG1655 genome at 99.5% base level accuracy using only MinION data [50]. Their pipeline, 'nanocorrect', corrected errors by first aligning reads using the graph-based, greedy partial order aligner method [52], and then by pruning errors that were apparent given the alignment graph. The error-corrected reads were then assembled using the Celera Assembler. This draft assembly was then further improved using Loman and co-worker's polishing algorithm, 'nanopolish' [50].

### Single-nucleotide variant calling

Reference allele bias, the tendency to over-report the presence of the reference allele and under-report non-reference alleles, becomes more acute when the error rate of the reads is higher, because non-reference variants are more likely to be lost in noisy alignments. To overcome this problem for MinION reads, several academic laboratories have developed MinION-specific variant calling tools.

The marginCaller module in marginAlign [9] uses maximum-likelihood parameter estimates and marginalization over multiple possible read alignments to call single nucleotide variants (SNVs). At a substitution rate of 1% (in silico), marginCaller detected SNVs with 97% precision and 97% recall at 60× coverage. Similarly, by optimizing read level alignments, Sović et al. [12] used their GraphMap approach, for accurate mapping at high identity, to detect heterozygous variants from difficult-to-analyze regions of the human genome with over 96% precision. They also used in silico tests to demonstrate that GraphMap could detect structural

variants (insertions and deletions of different lengths) with high precision and recall.

Nanopolish [50] uses event-level alignments to a reference for variant calling. This algorithm iteratively modifies the starting reference sequence to create a consensus of the reads by evaluating the likelihood of observing a series of ionic current signals given the reference nucleotide sequence. At each iteration, candidate modifications to the consensus sequence are made and the sequence with the highest likelihood is chosen. At termination of iteration, the alignment of the final consensus to the final reference sequence defines the variants (differences) between the reads and the reference. This approach was used to demonstrate the feasibility of real-time surveillance as part of a study in West Africa in which Quick et al. [30] identified ebola virus sub-lineages using the MinION with ~80% mean accuracy.

PoreSeq [53] is a similar algorithm to Nanopolish, published around the same time, that also iteratively maximizes the likelihood of observing the sequence given a model. Their model, which like Nanopolish uses MinION event-level data, accounts for the uncertainty that can arise during the traversal of DNA through the nanopore. PoreSeq can achieve high precision and recall SNV-calling at low coverages of sequence data. Using a 1% substitution rate in the M13 genome, Szalay and Golovchenko [53] demonstrated that PoreSeq could detect variants with a precision and recall of 99% using 16× coverage. This is around the same accuracy as marginAlign on the same data, but at a substantially lower coverage, demonstrating the power of the event-level, iterative approach.

### Consensus sequencing for high accuracy

The read accuracy of 92% currently achieved by MinION is useful for some applications, but at low coverage it is insufficient for applications such as haplotype phasing and SNV detection in human samples, where the number of variants to be detected is smaller than the published variant-detection error rates of algorithms using MinION data. One method previously used to improve the quality of single-molecule sequence employed rolling circle amplification [51]. In a parallel method for the MinION, Li et al. [54] used rolling circle amplification to generate multiple copies of the 16S ribosomal RNA (rRNA) gene in one contiguous strand. MinION nanopore sequencing of each contiguous strand gave a consensus accuracy of over 97%. This allowed sensitive profiling in a mixture of ten 16S rRNA genes.

### Current applications of the MinION

#### Analysis of infectious agents at point-of-care

Next-generation sequencing can detect viruses, bacteria, and parasites present in clinical samples and in a hospital

environment [11, 14, 27, 34]. These pathogen sequences enable the identification and surveillance of host adaptation, diagnostic targets, response to vaccines, and pathogen evolution [30]. MinIONs are a new tool in this area that provide substantial advantages in read length, portability, and time to pathogen identification, which is documented to be as little as 6 h from sample collection [14]. Pathogen identification can be performed in as little as 4 min once the sample is loaded on the MinION [14]. The breadth of clinical applications demonstrated to date include studies of chikungunya virus [14], hepatitis virus C [14], *Salmonella enterica* [28], and *Salmonella typhimurium* [7], as well as work on antibiotic resistance genes in five Gram-negative isolates and on the *mecA* gene in a methicillin-resistant *Staphylococcus aureus* (MRSA) isolate [17].

Arguably, the most inspired clinical use of the MinION to date involved teams of African and European scientists who analyzed ebola samples on-site in West Africa [30, 55]. The recent viral epidemic was responsible for over 28,599 ebola cases and more than 11,299 deaths [56]. In the larger of the two studies, Quick and colleagues [30] transported a MinION field sequencing kit (weighing <50 kg, and fitting within standard suitcases) by commercial airline to West Africa. Once there, they sequenced blood samples from 142 ebola patients in a field laboratory. Ebola virus sequence data were generated within 24 h after sample delivery, with confirmation of ebola sequences taking as little as 15 min of MinION run time. To our knowledge, these studies by Quick et al. [30] and by Hoenen et al. [55] are the first applications of any sequencing device for real-time on-site monitoring of an epidemic.

### Teaching and citizen science

The low cost of entry and portability of the MinION sequencer also make it a useful tool for teaching. It has been used to provide hands-on experience to undergraduate students as part of a recently taught course at Columbia University [57] and to teach graduate students at the University of California Santa Cruz. Every student was able to perform their own MinION sequencing. Similarly, the short and simple process of preparing a sequencing library allowed researchers at Mount Desert Island Biological Laboratory in Maine to train high school students during a summer course and have them run their own MinION experiments. Their Citizen Science initiative intends to address questions pertaining to health and environment that would otherwise be implausible [58].

### Aneuploidy detection

One of the immediate applications of the MinION is aneuploidy detection in prenatal samples. The typical



turnaround time for aneuploidy detection in such samples is 1–3 weeks when using NGS platforms [59]. Wei and Williams [38] used the MinION to detect aneuploidy in prenatal and miscarriage samples in under 4 h. They concluded that the MinION can be used for aneuploidy detection in a clinical setting.

### MinIONs in space

At present, it is hard to detect and identify bacteria and viruses on manned space flights. Most of these analyses, along with understanding the effects of space travel on genomes, occur when the samples are brought back to Earth. As a first step to resolve this shortcoming, NASA plans to test MinION-based real-time sequencing and pathogen identification on the International Space Station (ISS) [60, 61]. In a proof-of-concept experiment, Castro-Wallace et al. [62] demonstrated successful sequencing and de novo assembly of a lambda phage genome, an *E. coli* genome, and a mouse mitochondrial genome. They noted that there was no significant difference in the quality of sequence data generated on the ISS and in control experiments that were performed in parallel on Earth [62].

### Outlook

#### PromethION

The MinION allows individual laboratories to perform sequencing and subsequent biological analyses, but there is a part of the research community that is interested in high-throughput sequencing and genomics. Realizing this need, ONT has developed a bench-top instrument, PromethION, that is projected to provide high-throughput and is modular in design. Briefly, it will contain 48 flow cells that could be run individually or in parallel. The PromethION flow cells contain 3000 channels each, and are projected to produce up to 6 Tb of sequencing data each day. This equates to about 200 human genomes per day at 30× coverage.

#### Read accuracy

Single read accuracy is 92% for the current MinION device [44], which is often sufficient for applications such as the identification of pathogens or mRNA (cDNA) splice variants. However, some medical applications, such as the detection of individual nucleotide substitutions or base adducts in a single mitochondrial genome, would require read accuracies exceeding 99.99%. Given prior experience, it is reasonable that ONT will continue to improve their chemistry and base-calling software. Nevertheless, it is probable that Q40 nanopore sequencing will entail a single strand re-read strategy [2].

As is true for all sequencing platforms, MinION's base-call accuracy is improved using consensus-based methods. For example, for an *E. coli* strain where single

reads averaged ~80% accuracy, consensus accuracy improved to 99.5% at 30× coverage [50]. The remaining 0.5% error appears to be non-random. This improvement is in part due to the inability of the present MinION platform to resolve homopolymers longer than the nanopore reading head (six nucleotides), and to the absence of training in the detection of base modifications. It is plausible that resolving these two issues will push nanopore consensus accuracy to  $\geq 99.99\%$ .

#### Read length

With the advent of single-molecule sequencing technologies (PacBio and MinION), the average read lengths increased from 250 nucleotides to 10 kb. More recently, reads of more than 150 kb have routinely been achieved with the MinION (Akeson, unpublished findings), and this is expected to improve in the next few months. Achieving long reads will allow progress in understanding highly complex and repetitive regions in genomes that are otherwise hard to resolve.

#### Direct RNA sequencing

Sequencing of direct RNA with nanopore technology is an active area of development at ONT and in academic research groups. Single-molecule detection of tRNA has been previously demonstrated in single-channel and solid-state nanopores [63, 64]. Nanopore sensing can also detect nucleotide modifications in both DNA [39–42] and tRNA [65]. Direct RNA sequencing will reveal insights in RNA biology that presently can get lost due to issues with reverse transcription and PCR amplification.

#### Single-molecule protein sensing

At present, mass spectrometry is the preferred technique for performing a comprehensive proteomics analysis [66], but there are limitations to the sensitivity, accuracy, and resolution of any one analytical technique [66]. In 2013, Nivala et al. [67] demonstrated enzyme-mediated translocation of proteins through a single-channel nanopore. Their study showed that sequence-specific features of the proteins could be detected. They then engineered five protein constructs bearing different mutations and rearrangements, and demonstrated that these constructs could be discriminated with accuracies ranging from 86 to 99%. Protein sequencing will allow studies of complex interactions among cells in different tissues [68].

### Conclusions

Nanopore DNA strand sequencing is now an established technology. In the short interval since the ONT MinION was first released, performance has improved rapidly, and the technology now routinely achieves read lengths of 50 kb and more and single-strand read accuracies of better than 92%. Improvement in read lengths,

base-call accuracies, base modification detection, and throughput is likely to continue. Owing to its portability, the MinION nanopore sequencer has proven utility at the point-of-care in challenging field environments. Further miniaturization of the platform (SmidgION) and associated library preparation tools (Zumbador, Vol-TRAX) promise an age of ubiquitous sequencing. Parallel applications, including direct RNA sequencing, are on the horizon.

#### Abbreviations

5-hmC: 5-hydroxymethylcytosine; 5-mC: 5-methylcytosine; C: Cytosine; dsDNA: Double-stranded DNA; HMM: Hidden Markov model; ISS: International Space Station; MAP: MinION Access Program; NGS: Next generation sequencing; ONT: Oxford Nanopore Technologies; rRNA: Ribosomal RNA; SNV: Single nucleotide variant;

#### Acknowledgements

The authors thank the National Human Genome Research Institute of the US National Institutes of Health for funding their research under award numbers HG006321 (MA), HG007827 (MA), and U54HG007990 (BP). The authors also thank Ariah Mackie for proofreading the manuscript.

#### Author contributions

All authors contributed to the writing, editing, and completion of the manuscript. All authors read and approved the final manuscript.

#### Competing interests

MA is a member of the Technology Advisory Board of Oxford Nanopore Technologies, Oxford, UK, for which he receives compensation, including stock options of unknown value that may exceed \$10,000. He is an inventor on the following nanopore-sequencing-related US patents owned at least in part by the University of California: 6,015,714; 6,267,872; 6,465,193; 6,746,594; 6,936,433; 7,060,507; 7,189,503; 7,238,485; 7,625,706; 7,947,454; 8,500,982; 8,673,556; and 8,679,747. The other authors declare that they have no competing interests.

Published online: 25 November 2016

#### References

- Branton D, Daniel B, Deamer DW, Andre M, Hagan B, Benner SA, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol.* 2008;26:1146–53.
- Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol.* 2016;34:518–24.
- Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A.* 1996;93:13770–3.
- Cherf GM, Lieberman KR, Hytham R, Lam CE, Kevin K, Mark A. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat Biotechnol.* 2012;30:344–8.
- Ayub M, Bayley H. Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.* 2012;12:5637–43.
- Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Nathaniel G, et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol.* 2012;30:349–53.
- Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol.* 2015;33:296–300.
- Bolisetty MT, Rajadinakaran G, Graveley BR. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* 2015;16:204.
- Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. *Nat Methods.* 2015;12:351–6.
- Quick J, Quinlan A, Loman N. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *GigaScience.* 2014;3:22.
- Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, et al. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience.* 2015;4:12.
- Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun.* 2016;7:11307.
- Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* 2015;25:1750–6.
- Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 2015;7:99.
- Hargreaves AD, Mulley JF. Assessing the utility of the Oxford Nanopore MinION for snake venom gland cDNA sequencing. *Peer J.* 2015;3:e1441.
- Cao MD, Ganesamoorthy D, Elliott A, Zhang H, Cooper MA, Coin LJM. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION™ sequencing. *GigaScience.* 2016;5:32.
- Judge K, Harris SR, Reuter S, Parkhill J, Peacock SJ. Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *J Antimicrob Chemother.* 2015;70:2775–8.
- Karlsson E, Lärkeryd A, Sjödin A, Forsman M, Stenberg P. Scaffolding of a bacterial genome using MinION nanopore sequencing. *Sci Rep.* 2015;5:11996.
- Kchouk M, Mehdi K, Mourad E. Error correction and DeNovo genome Assembly for the Minlon sequencing reads mixing Illumina short reads. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015. doi:10.1109/bibm.2015.7359962.
- Leggett RM, Heavens D, Caccamo M, Clark MD, Davey RP. NanoOK: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics.* 2016;32:142–4.
- Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol.* 2015;13:787–94.
- Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics.* 2014;30:3399–401.
- Madoui M-A, Engelen S, Cruaud C, Belsler C, Bertrand L, Alberti A, et al. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics.* 2015;16:327.
- Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour.* 2014;14:1097–102.
- Miles G, Hoisington-Lopez J, Duncavage E. Nanopore sequencing of a DNA library prepared from formalin-fixed paraffin-embedded tissue. *Lab Invest.* 2015;95 Suppl 1:520–1.
- Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. Metagenomics for pathogen detection in public health. *Genome Med.* 2013;5:81.
- Pallen MJ. Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections. *Parasitology.* 2014;141:1856–62.
- Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 2015;16:114.
- Quick J, Loman NJ. Bacterial whole-genome read data from the Oxford Nanopore Technologies MinION™ nanopore sequencer. *GigaScience Database.* 2014. doi:10.5524/100102.
- Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for ebola surveillance. *Nature.* 2016;530:228–32.
- Quick J, Quinlan AR, Loman NJ. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience.* 2014;3:1–6.
- Ramgren AC, Newhall HS, James KE. DNA barcoding and metabarcoding with the Oxford Nanopore MinION. *Genome.* 2015;58:268.
- Risse J, Thomson M, Patrick S, Blakely G, Koutsouvolos G, Blaxter M, et al. A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. *Gigascience.* 2015;4:60.
- Wang J, Moore NE, Deng Y-M, Eccles DA, Hall RJ. MinION nanopore sequencing of an influenza genome. *Front Microbiol.* 2015;6:766.
- Wang JR, Jones CD. Fast alignment filtering of nanopore sequencing reads using locality-sensitive hashing. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015. doi:10.1109/bibm.2015.7359668.
- Ward AC, Kim W. MinION™: new, long read, portable nucleic acid sequencing device. *J Bacteriol Virol.* 2015;45:285.
- Watson M, Thomson M, Risse J, Talbot R, Santoyo-Lopez J, Gharbi K, et al. poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics.* 2015;31:114–5.

38. Wei S, Williams Z. Rapid short-read sequencing and aneuploidy detection using MinION nanopore technology. *Genetics*. 2016;202:37–44.
39. Schreiber J, Wescoe ZL, Abu-Shumays R, Vivian JT, Baatar B, Karplus K, et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc Natl Acad Sci U S A*. 2013;110:18910–5.
40. Wescoe ZL, Schreiber J, Akeson M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc*. 2014;136:16582–7.
41. Rand AC, Jain M, Eizenga J, Musselman-Brown A, Olsen HE, Akeson M, et al. Cytosine variant calling with high-throughput nanopore sequencing. *bioRxiv*. 2016. doi:10.1101/047134.
42. Simpson JT, Workman R, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA methylation using the Oxford Nanopore Technologies MinION sequencer. *bioRxiv*. 2016. doi:10.1101/047142.
43. Loose M, Malla S, Stout M. Real time selective sequencing using nanopore technology. *Nat Methods*. 2016;13:751–4.
44. Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. MinION analysis and reference consortium: phase 1 data release and analysis. *F1000Res*. 2015;4:1075.
45. Norris AL, Workman RE, Fan Y, Eshleman JR, Timp W. Nanopore sequencing detects structural variants in cancer. *Cancer Biol Ther*. 2016;17:246–53.
46. David M, Dursi LJ, Yao D, Boutros PC, Simpson JT. Nanocall: an open source basecaller for Oxford nanopore sequencing data. *Bioinformatics*. 2016. doi:10.1093/bioinformatics/btw569.
47. Boža V, Břejová B, Vinař T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. *arXiv.org*. 2016. arXiv:1603.09195 [q-bio.GN].
48. Frith MC, Hamada M, Horton P. Parameters for accurate genome alignment. *BMC Bioinformatics*. 2010;11:80.
49. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.org*. 2013. arXiv:1303.3997 [q-bio.GN].
50. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods*. 2015;12:733–5.
51. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323:133–8.
52. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics*. 2002;18:452–64.
53. Szalay T, Golovchenko JA. De novo sequencing and variant calling with nanopores using PoreSeq. *Nat Biotechnol*. 2015;33:1087–91.
54. Li C, Chng KR, Boey JHE, Ng HQA, Wilm A, Nagarajan N. INC-Seq: accurate single molecule reads using nanopore sequencing. *GigaScience*. 2016;5:34.
55. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, et al. Nanopore sequencing as a rapidly deployable ebola outbreak tool. *Emerg Infect Dis*. 2016;22:331–4.
56. World Health Organisation. Ebola Situation Report – 11 November 2015. World Health Organisation. 2015. <http://apps.who.int/ebola/current-situation/ebola-situation-report-11-november-2015>. Accessed 21 Jun 2016.
57. Zaaier S. Columbia University Ubiquitous Genomics 2015 Class, Erlich Y. *Elife*. 2016;5:e14258.
58. Krol A. Citizen sequencers: taking Oxford Nanopore's MinION to the classroom and beyond. *Bio-IT World*. 9 Dec 2015. [www.bio-itworld.com/2015/12/9/citizen-sequencers-taking-oxford-nanopores-minion-classroom-beyond.html](http://www.bio-itworld.com/2015/12/9/citizen-sequencers-taking-oxford-nanopores-minion-classroom-beyond.html). Accessed 29 Jun 2016.
59. Chen S, Li S, Xie W, Li X, Zhang C, Jiang H, et al. Performance comparison between rapid sequencing platforms for ultra-low coverage sequencing strategy. *PLoS One*. 2014;9:e92192.
60. Regalado A. Now they're sequencing DNA in outer space. *MIT Technology Review*. 10 Jun 2016. [www.technologyreview.com/s/601669/now-theyre-sequencing-dna-in-outer-space/](http://www.technologyreview.com/s/601669/now-theyre-sequencing-dna-in-outer-space/). Accessed 29 Jun 2016.
61. Dunn A. Sequencing DNA in the palm of your hand. 29 Sep 2015. [www.nasa.gov/mission\\_pages/station/research/news/biomolecule\\_sequencer](http://www.nasa.gov/mission_pages/station/research/news/biomolecule_sequencer). Accessed 29 Jun 2016.
62. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, et al. Nanopore DNA sequencing and genome assembly on the International Space Station. *bioRxiv*. 2016. doi:10.1101/077651.
63. Smith AM, Abu-Shumays R, Akeson M, Bernick DL. Capture, unfolding, and detection of individual tRNA molecules using a nanopore device. *Front Bioeng Biotechnol*. 2015;3:91.
64. Henley RY, Ashcroft BA, Farrell I, Cooperman BS, Lindsay SM, Wanunu M. Electrophoretic deformation of individual transfer RNA molecules reveals their identity. *Nano Lett*. 2016;16:138–44.
65. Akeson M. Progress at UC Santa Cruz: long DNA fragments, tRNA and modified bases | Vimeo. 26 May 2016. <https://vimeo.com/168851338>. Accessed 19 Oct 2016.
66. Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstetrician Gynaecologist*. 2011;13:189–95.
67. Nivala J, Marks DB, Akeson M. Unfoldase-mediated protein translocation through an  $\alpha$ -hemolysin nanopore. *Nat Biotechnol*. 2013;31:247–50.
68. Hood LE, Omenn GS, Moritz RL, Aebersold R, Yamamoto KR, Amos M, et al. New and improved proteomics technologies for understanding complex biological systems: addressing a grand challenge in the life sciences. *Proteomics*. 2012;12:2773–83.
69. Chen Y-T, Iseli C, Venditti CA, Old LJ, Simpson AJG, Jongeneel CV. Identification of a new cancer/testis gene family, CT47, among expressed multicopy genes on the human X chromosome. *Genes Chromosomes Cancer*. 2006;45:392–400.
70. Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol*. 2015;33:623–30.
71. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol*. 2016;17:132.
72. minoTour. Real time data analysis tools for the minION sequencing platform. *GitHub*. 2016. <https://github.com/minoTour/minoTour>. Accessed 26 Jun 2016.