REVIEW ARTICLE
https://doi.org/10.1038/s41587-021-01108-x

nature biotechnology

Nanopore sequencing technology, bioinformatics and applications

Yunhao Wang[1,3], Yue Zhao[1,2,3], Audrey Bollas[1,3], Yuru Wang[1] and Kin Fai Au[1,2] ✉

Rapid advances in nanopore technologies for sequencing single long DNA and RNA molecules have led to substantial improvements in accuracy, read length and throughput. These breakthroughs have required extensive development of experimental and bioinformatics methods to fully exploit nanopore long reads for investigations of genomes, transcriptomes, epigenomes and epitranscriptomes. Nanopore sequencing is being applied in genome assembly, full-length transcript detection and base modification detection and in more specialized areas, such as rapid clinical diagnoses and outbreak surveillance. Many opportunities remain for improving data quality and analytical approaches through the development of new nanopores, base-calling methods and experimental protocols tailored to particular applications.

Nanopore sequencing technology and its applications in basic and applied research have undergone substantial growth since Oxford Nanopore Technologies (ONT) provided the first nanopore sequencer, MinION, in 2014 (refs. [1,2]). The technology relies on a nanoscale protein pore, or 'nanopore', that serves as a biosensor and is embedded in an electrically resistant polymer membrane[1,3] (Fig. 1). In an electrolytic solution, a constant voltage is applied to produce an ionic current through the nanopore such that negatively charged single-stranded DNA or RNA molecules are driven through the nanopore from the negatively charged '*cis*' side to the positively charged '*trans*' side. Translocation speed is controlled by a motor protein that ratchets the nucleic acid molecule through the nanopore in a step-wise manner. Changes in the ionic current during translocation correspond to the nucleotide sequence present in the sensing region and are decoded using computational algorithms, allowing real-time sequencing of single molecules. In addition to controlling translocation speed, the motor protein has helicase activity, enabling double-stranded DNA or RNA–DNA duplexes to be unwound into single-stranded molecules that pass through the nanopore.

In this review, we first present an introduction to the technology development of nanopore sequencing and discuss improvements in the accuracy, read length and throughput of ONT data. Next, we describe the main bioinformatics methods applied to ONT data. We then review the major applications of nanopore sequencing in basic research, clinical studies and field research. We conclude by considering the limitations of the existing technologies and algorithms and directions for overcoming these limitations.

Technology development

Nanopore design. The concept of nanopore sequencing emerged in the 1980s and was realized through a series of technical advances in both the nanopore and the associated motor protein[1,4–8]. α-Hemolysin, a membrane channel protein from *Staphylococcus aureus* with an internal diameter of ~1.4 nm to ~2.4 nm (refs. [1,9]), was the first nanopore shown to detect recognizable ionic current blockades by both RNA and DNA homopolymers[10–12]. In a crucial step toward single-nucleotide-resolution nanopore sequencing, engineering of the wild-type α-hemolysin protein allowed the four DNA bases on oligonucleotide molecules to be distinguished, although complex sequences were not examined in these reports[13–15]. Similar results were achieved using another engineered nanopore, *Mycobacterium smegmatis* porin A (MspA)[16,17], that has a similar channel diameter (~1.2 nm)[18,19].

A key advance in improving the signal-to-noise ratio was the incorporation of processive enzymes to slow DNA translocation through the nanopore[20–22]. In particular, phi29 DNA polymerase was found to have superior performance in ratcheting DNA through the nanopore[23,24]. Indeed, this motor protein provided the last piece of the puzzle; in February 2012, two groups demonstrated processive recordings of ionic currents for single-stranded DNA molecules that could be resolved into signals from individual nucleotides by combining phi29 DNA polymerase and a nanopore (α-hemolysin[24] and MspA[25]). In contrast to the previous DNA translocation tests that were poorly controlled[13–17], the addition of the motor protein reduced the fluctuations in translocation kinetics, thus improving data quality. In the same month, ONT announced the first nanopore sequencing device, MinION[26]. ONT released the MinION to early users in 2014 and commercialized it in 2015 (ref. [2]) (Fig. 2a). There have been several other nanopore-based sequencing ventures, such as Genia Technologies's nanotag-based real-time sequencing by synthesis (Nano-SBS) technology, NobleGen Biosciences's optipore system and Quantum Biosystems's sequencing by electronic tunneling (SBET) technology[27,28]. However, this review focuses on ONT technology as it has been used in most peer-reviewed studies of nanopore sequencing, data, analyses and applications.

ONT has continually refined the nanopore and the motor protein, releasing eight versions of the system to date, including R6 (June 2014), R7 (July 2014), R7.3 (October 2014), R9 (May 2016), R9.4 (October 2016), R9.5 (May 2017), R10 (March 2019) and R10.3 (January 2020) (Fig. 2a). The original or engineered proteins used in the R6, R7, R7.3, R10 and R10.3 nanopores have not been disclosed by the company to date. R9 achieved a notable increase in sequencing yield per unit of time and in sequencing accuracy (~87% (ref. [29]) versus ~64% for R7 (ref. [30])) by using the nanopore Curlin sigma S-dependent growth subunit G (CsgG) from *Escherichia coli* (Fig. 2b and Supplementary Table 1). This nanopore has a translocation rate of ~250 bases per s compared to ~70 bases per s for R7 (ref. [31]).

[1]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA. [2]Biomedical Informatics Shared Resources, The Ohio State University, Columbus, OH, USA. [3]These authors contributed equally: Yunhao Wang, Yue Zhao, Audrey Bollas. ✉e-mail: kinfai.au@osumc.edu

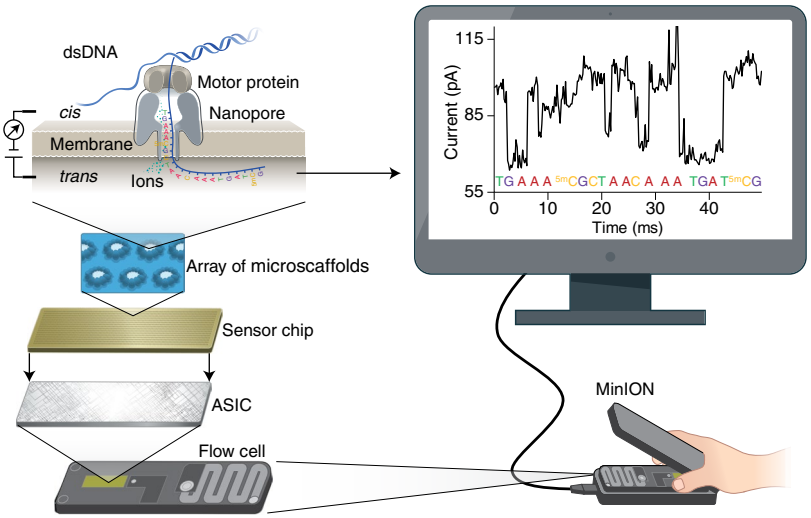

Fig. 1 | Principle of nanopore sequencing. A MinION flow cell contains 512 channels with 4 nanopores in each channel, for a total of 2,048 nanopores used to sequence DNA or RNA. The wells are inserted into an electrically resistant polymer membrane supported by an array of microscaffolds connected to a sensor chip. Each channel associates with a separate electrode in the sensor chip and is controlled and measured individually by the application-specific integration circuit (ASIC). Ionic current passes through the nanopore because a constant voltage is applied across the membrane, where the *trans* side is positively charged. Under the control of a motor protein, a double-stranded DNA (dsDNA) molecule (or an RNA–DNA hybrid duplex) is first unwound, then single-stranded DNA or RNA with negative charge is ratcheted through the nanopore, driven by the voltage. As nucleotides pass through the nanopore, a characteristic current change is measured and is used to determine the corresponding nucleotide type at ~450 bases per s (R9.4 nanopore).

Subsequently, a mutant CsgG and a new motor enzyme (whose origin was not disclosed) were integrated into R9.4 to achieve higher sequencing accuracy (~85–94% as reported in refs. [32–36]) and faster sequencing speeds (up to 450 bases per s). R9.5 was introduced to be compatible with the 1D^2 sequencing strategy, which measures a single DNA molecule twice (see below). However, the R9.4 and R9.5 have difficulty sequencing very long homopolymer runs because the current signal of CsgG is determined by approximately five consecutive nucleotides. The R10 and R10.3 nanopores have two sensing regions (also called reader heads) to aim for higher accuracy with homopolymers[37,38], although independent studies are needed to assess this claim.

Additional strategies to improve accuracy. Beyond optimizing the nanopore and motor protein, several strategies have been developed to improve accuracy. Data quality can be improved by sequencing each dsDNA multiple times to generate a consensus sequence, similar to the 'circular consensus sequencing' strategy used in the other single-molecule long-read sequencing method from Pacific Biosciences (PacBio)[39]. Early versions of ONT sequencing used a 2D library preparation method to sequence each dsDNA molecule twice; the two strands of a dsDNA molecule are ligated together by a hairpin adapter, and a motor protein guides one strand (the 'template') through the nanopore, followed by the hairpin adapter and the second strand (the 'complement')[40–42] (Fig. 3d, left). After removing the hairpin sequence, the template and complement reads, called the 1D reads, are used to generate a consensus sequence, called the 2D read, of higher accuracy. Using the R9.4 nanopore as an example, the average accuracy of 2D reads is 94% versus 86% for 1D reads[33] (Fig. 2b). In May 2017, ONT released the 1D^2 method together with the R9.5 nanopore; in this method, instead of being physically connected by a hairpin adapter, each strand is ligated separately to a special adapter (Fig. 3d, right). This special adapter provides a high probability (>60%) that the complement strand will immediately be captured by the same nanopore after the template strand, offering similar consensus sequence generation for dsDNA as the 2D library. The average accuracy of 1D^2 reads is up to 95% (R9.5 nanopore)[43] (Fig. 2b). Unlike the 2D library, the complement strand in the 1D^2 library is not guaranteed to follow the template, resulting in imperfect consensus sequence generation. However, ONT no longer offers or supports the 2D and 1D^2 libraries. Currently, for DNA sequencing, ONT only supports the 1D method in which each strand of a dsDNA is ligated with an adapter and sequenced independently (Fig. 3d, middle).

In parallel, accuracy has been improved through new base-calling algorithms, including many developed through independent research[32,44] (see below). Taking the R7.3 nanopore as an example, the 1D read accuracy was improved from 65% by hidden Markov model (HMM)[45] to 70% by Nanocall[46] and to 78% by DeepNano[47].

Extending read length. Although the accuracy of ONT sequencing is relatively low, the read length provided by electrical detection has a very high upper bound because the method relies on the physical process of nucleic acid translocation[48]. Reads of up to 2.273 megabases (Mb) were demonstrated in 2018 (ref. [49]). Thus, ONT read lengths depend crucially on the sizes of molecules in the sequencing library. Various approaches for extracting and purifying high-molecular-weight (HMW) DNA have been reported or applied to ONT sequencing, including spin columns (for example, Monarch Genomic DNA Purification kit, New England Biolabs), gravity-flow columns (for example, NucleoBond HMW DNA kit, Takara Bio), magnetic beads (for example, MagAttract HMW DNA kit, QIAGEN), phenol–chloroform, dialysis and plug extraction[50] (Fig. 3a). HMW DNA can also be sheared to the desired size by sonication, needle extrusion or transposase cleavage (Fig. 3a). However, overrepresented small fragments outside the desired size distribution may decrease sequencing yield because of higher efficiencies of both adapter ligation and translocation through nanopores than long fragments. To remove overrepresented small DNA fragments, various size selection methods (for example, the gel-based BluePippin system of Sage Science, magnetic beads and the Short Read Eliminator kit of Circulomics) have been used to obtain the desired data distribution and/or improve sequencing yield (Fig. 3a).

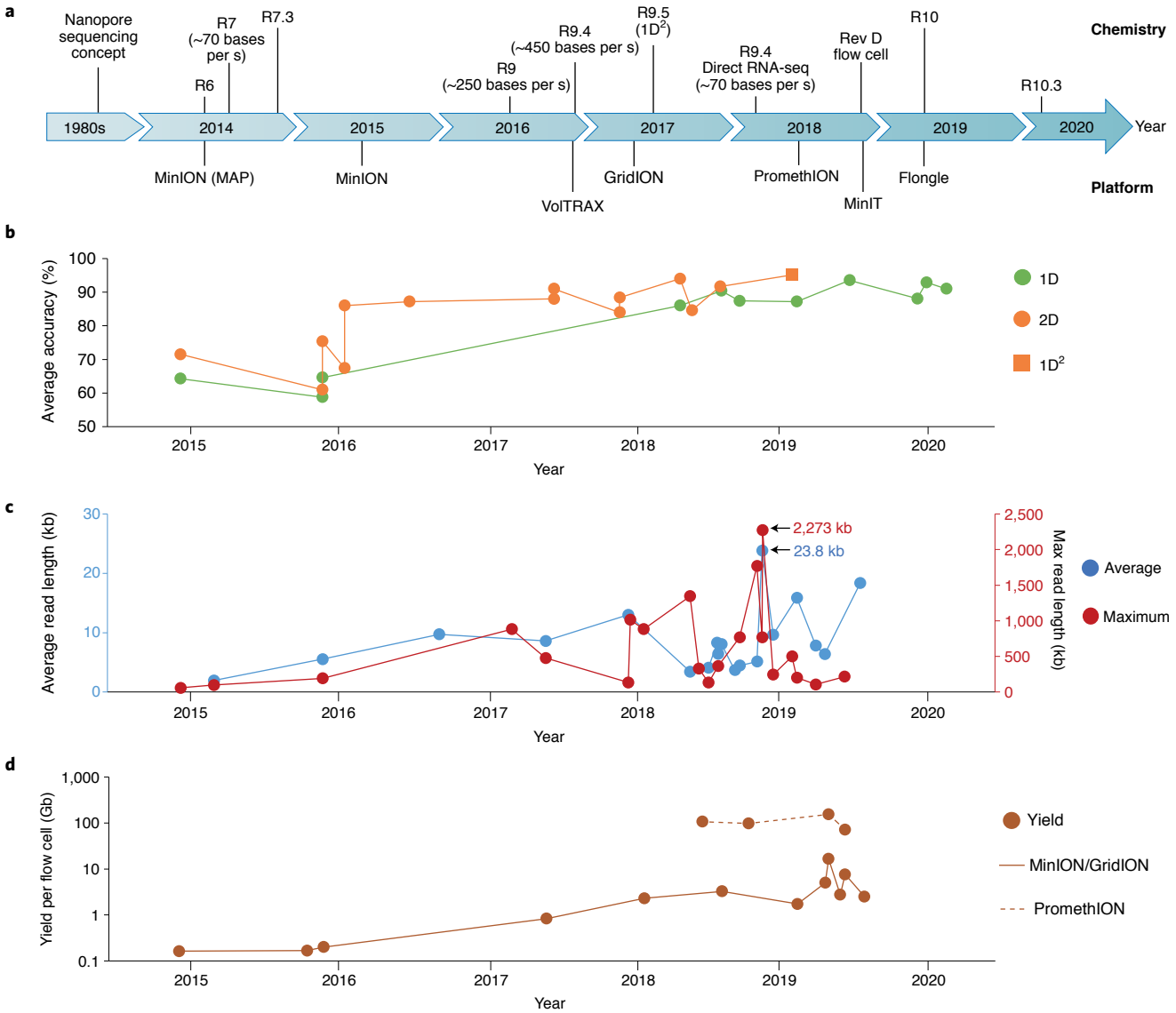

Fig. 2 | ONT sequencing data improvement over time. a, Timeline of the major chemistry and platform releases by ONT. **b**, Accuracy of 1D, 2D and 1D^2 reads. **c**, Average and maximum read lengths. Special efforts have been made in some studies to achieve ultralong read length. For example, by late 2019, the highest average sequencing length achieved has been 23.8 kilobases (kb) using a specific DNA extraction protocol[51]. The longest individual read is 2,273 kb, rescued by correcting an error in the software MinKNOW[49]. The DNA extraction and purification methods used in these independent studies are summarized in Supplementary Table 1. Read lengths are reported for 1D reads. **d**, Yield per flow cell (in $\log_{10}$ scale for *y* axis). Yields are reported for 1D reads. Data points shown in **b** (accuracy), **c** (read length) and **d** (yield) are from independent studies. Details for these data points are summarized in Supplementary Table 1.

With improvements in nanopore technology and library preparation protocols (Figs. 2a and 3a), the maximum read length has increased from <800 kb in early 2017 to 2.273 Mb in 2018 (ref. [49]) (Fig. 2c). The average read length has increased from a few thousand bases at the initial release of MinION in 2014 to ~23 kb (ref. [51]) in 2018 (Fig. 2c), primarily due to improvements in HMW DNA extraction methods and size selection strategies. However, there is a trade-off between read length and yield; for example, the sequencing yield of the HMW genomic DNA library is relatively low.

Sequencing RNA. ONT devices have been adapted to directly sequence native RNA molecules[52]. The method requires special library preparation in which the primer is ligated to the 3′ end of native RNA, followed by direct ligation of the adapter without conventional reverse transcription (Fig. 3c). Alternatively, a cDNA strand can be synthesized to obtain an RNA–cDNA hybrid duplex, followed by ligation of the adapter. The former strategy requires less sample manipulation and is quicker and thus is good for on-site applications, whereas the latter produces a more stable library for longer sequencing courses and therefore produces higher yields. In both cases, only the RNA strand passes through the nanopore, and therefore direct sequencing of RNA molecules does not generate a consensus sequence (for example, 2D or 1D^2). Compared to DNA sequencing, direct RNA sequencing is typically of lower average accuracy, around 83–86%, as reported by independent research[53,54].

Like conventional RNA sequencing, ONT can be used to perform cDNA sequencing by utilizing existing full-length cDNA synthesis methods (for example, the SMARTer PCR cDNA Synthesis kit of Takara Bio and the TeloPrime Full-Length cDNA Amplification kit

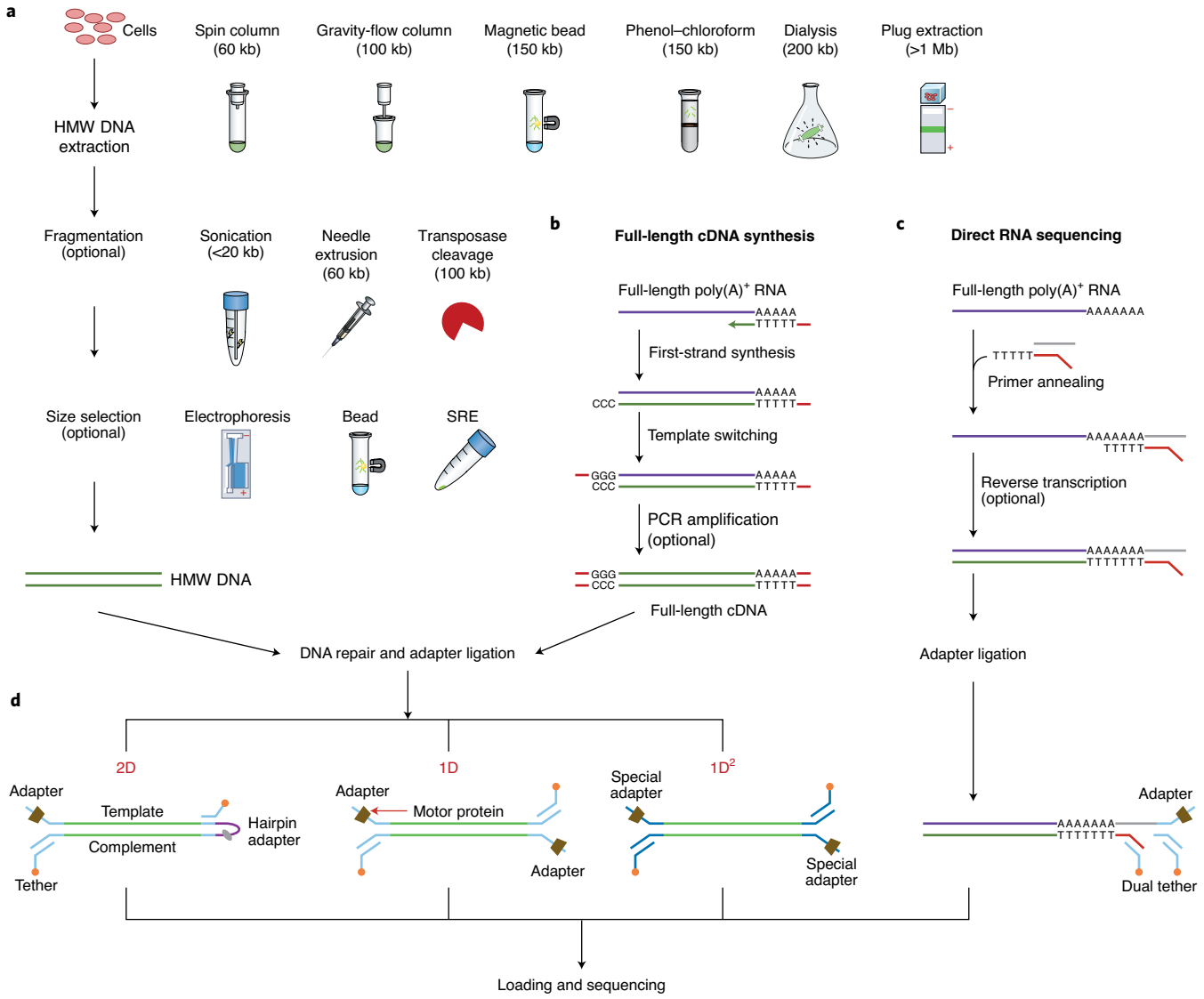

Fig. 3 | Library preparation workflow for ONT sequencing. a, Special experimental techniques for ultralong genomic DNA sequencing, including HMW DNA extraction, fragmentation and size selection. **b**, Full-length cDNA synthesis for direct cDNA sequencing (without a PCR amplification step) and PCR-cDNA sequencing (with a PCR amplification step). **c**, Direct RNA-sequencing library preparation with or without a reverse transcription step, where only the RNA strand is ligated with an adapter and thus only the RNA strand is sequenced. **d**, Different library preparation strategies for DNA/cDNA sequencing, including 2D (where the template strand is sequenced, followed by a hairpin adapter and the complement strand), 1D (where each strand is ligated with an adapter and sequenced independently) and $1D^2$ (where each strand is ligated with a special adapter such that there is a high probability that one strand will immediately be captured by the same nanopore following sequencing of the other strand of dsDNA); SRE, short read eliminator kit (Circulomics).

of Lexogen) followed by PCR amplification[42,55] (Fig. 3b). ONT also offers a direct cDNA sequencing protocol without PCR amplification, in contrast to many existing cDNA sequencing methods. This approach avoids PCR amplification bias, but it requires a relatively large amount of input material and longer library preparation time, making it unsuitable for many clinical applications. A recent benchmarking study demonstrated that ONT sequencing of RNA, cDNA or PCR-cDNA for the identification and quantification of gene isoforms provides similar results[56].

Increasing throughput. In addition to sequencing length and accuracy, throughput is another important consideration for ONT sequencing applications. To meet the needs of different project scales, ONT released several platforms (Box 1). The expected data output of a flow cell mainly depends on (1) the number of active nanopores, (2) DNA/RNA translocation speed through the nanopore and (3) running time.

Early MinION users reported typical yields of hundreds of megabases per flow cell, while current throughput has increased to ~10–15 gigabases (Gb) (Fig. 2d, solid line) for DNA sequencing through faster chemistry (increasing from ~30 bases per s by R6 nanopore to ~450 bases per s by R9.4 nanopore) and longer run times with the introduction of the Rev D ASIC chip. Subsequent devices, such as PromethION, run more flow cells with more nanopores per flow cell. An independent study reported a yield of 153 Gb from a single PromethION flow cell with an average sequencing speed of ~430 bases per s (ref. [57]) (Fig. 2d, dashed line). By contrast, direct RNA sequencing currently produces about 1,000,000 reads (1–3 Gb) per MinION flow cell due in part to its relatively low sequencing speed (~70 bases per s).

Box 1 | ONT devices

- MinION is a flow cell containing 512 channels, with four nanopores per channel. Only one nanopore in each channel is measured at a time, allowing concurrent sequencing of up to 512 molecules.
- GridION, for medium-scale projects, has five parallel MinION flow cells.
- PromethION, a high-throughput device for large-scale projects, has 24 or 48 parallel flow cells (up to 3,000 channels per flow cell).
- Flongle, for smaller projects, is a flow cell adapter for MinION or GridION with 126 channels.
- VolTRAX is a programmable device for sample and library preparation.
- MinIT is a data analysis device that eliminates the need for a computer to run MinION.
- SmidgION is a smartphone-compatible device under development.

Data analysis

Bioinformatics analysis of ONT data has undergone continued improvement (Fig. 4). In addition to in-house data collection and specific data formats, many ONT-specific analyses focus on better utilizing the ionic current signal for purposes such as base calling, base modification detection and postassembly polishing. Other tools use long read length while accounting for high error rate. Many of these, such as tools for error correction, assembly and alignment, were developed for PacBio data but are also applicable to ONT data (Table 1).

Because ONT devices do not require high-end computing resources or advanced skills for basic data processing, many laboratories can run data collection themselves. MinKNOW is the operating software used to control ONT devices by setting sequencing parameters and tracking samples (Fig. 4, top left). MinKNOW also manages data acquisition and real-time analysis and performs local base calling and outputs the binary files in fast5 format to store both metadata and read information (for example, current measurement and read sequence if base calling is performed). The fast5 format organizes the multidimensional data in a nested manner, allowing the piece-wise access/extraction of information of interest without navigating through the whole dataset. Previous versions of MinKNOW output one fast5 file for each single read (named single-fast5), but later versions output one fast5 file for multiple reads (named multi-fast5) to meet the increasing throughput. Both fast5 and fastq files are output if the base-calling mode is applied during the sequencing experiment. In addition to official ONT tools (for example, ont_fast5_api software for format conversion between single-fast5 and multi-fast5 and data compression/decompression), several third-party software packages[40,58–62] have been developed for quality control, format conversion (for example, NanoR[63] for generating fastq files from fast5 files containing sequence information), data exploration and visualization of the raw ONT data (for example, Poretools[64], NanoPack[65] and PyPore[66]) and for after base-calling data analyses (for example, AlignQC[42] and BulkVis[49]) (Fig. 4, top right).

Base calling. Base calling, which decodes the current signal to the nucleotide sequence, is critical for data accuracy and detection of base modifications (Fig. 4, top center). Overall, method development for base calling went through four stages[32,44,58,67,68]: (1) base calling from the segmented current data by HMM at the early stage and by recurrent neural network in late 2016, (2) base calling from raw current data in 2017, (3) using a flip–flop model for identifying individual nucleotides in 2018 and (4) training customized base-calling models in 2019. ONT developed new base callers as 'technology demonstrator' software (for example, Nanonet, Scrappie and Flappie), which were subsequently implemented into the officially available software packages (for example, Albacore and Guppy). Albacore development is now discontinued in favor of Guppy, which can also run on graphics processing units in addition to central processing units to accelerate base calling.

ONT devices take thousands of current measurements per second. Processive translocation of a DNA or RNA molecule leads to a characteristic current shift that is determined by multiple consecutive nucleotides (that is, *k*-mer) defined by the length of the nanopore sensing region[1]. The raw current measurement can be segmented based on current shift to capture individual signals from each *k*-mer. Each current segment contains multiple measurements, and the corresponding mean, variance and duration of the current measurements together make up the 'event' data. The dependence of event data on neighboring nucleotides is Markov chain-like, making HMM-based methods a natural match to decode current shifts to nucleotide sequence, such as early base callers (for example, cloud-based Metrichor by ONT and Nanocall[46]). The subsequent Nanonet by ONT (implemented into Albacore) and DeepNano[47] implemented a recurrent neural network algorithm to improve base-calling accuracy by training a deep neural network to infer *k*-mers from the event data. In particular, Nanonet used a bidirectional method to include information from both upstream and downstream states on base calling.

However, information may be lost when converting raw current measurement into event data, potentially diminishing base-calling accuracy. Raw current data were first used for classifying ONT reads into specific species[69]. Later, ONT's open-source base caller Scrappie (implemented into both Albacore and Guppy) and the third-party software Chiron[70] adopted neural networks to directly translate the raw current data into DNA sequence. Subsequently, ONT released the base caller Flappie, which uses a flip–flop model with a connectionist temporal classification decoding architecture and identifies individual bases instead of *k*-mers from raw current data. Furthermore, the software Causalcall uses a modified temporal convolutional network combined with a connectionist temporal classification decoder to model long-range sequence features[35]. In contrast to generalized base-calling models, ONT introduced Taiyaki (implemented into Guppy) to train customized (for example, application/species-specific) base-calling models by using language processing techniques to handle the high complexity and long-range dependencies of raw current data. Additionally, Taiyaki can train models for identifying modified bases (for example, 5-methylcytosine (5mC) or N^6-methyladenine (6mA)) by adding a fifth output dimension. The R10 and R10.3 nanopores with two sensing regions may result in different signal features compared to previous raw current data, which will likely drive another wave of method development to improve data accuracy and base modification detection. To date, Guppy is the most widely used base caller because of its superiority in accuracy and speed[32] (Table 1).

Detecting DNA and RNA modifications. ONT enables the direct detection of some DNA and RNA modifications by distinguishing their current shifts from those of unmodified bases[52,71–74] (Fig. 4, middle center), although the resolution varies from the bulk level to the single-molecule level. A handful of DNA and RNA modification detection tools have been developed over the years (Table 1). Nanoraw (integrated into the Tombo software package) was

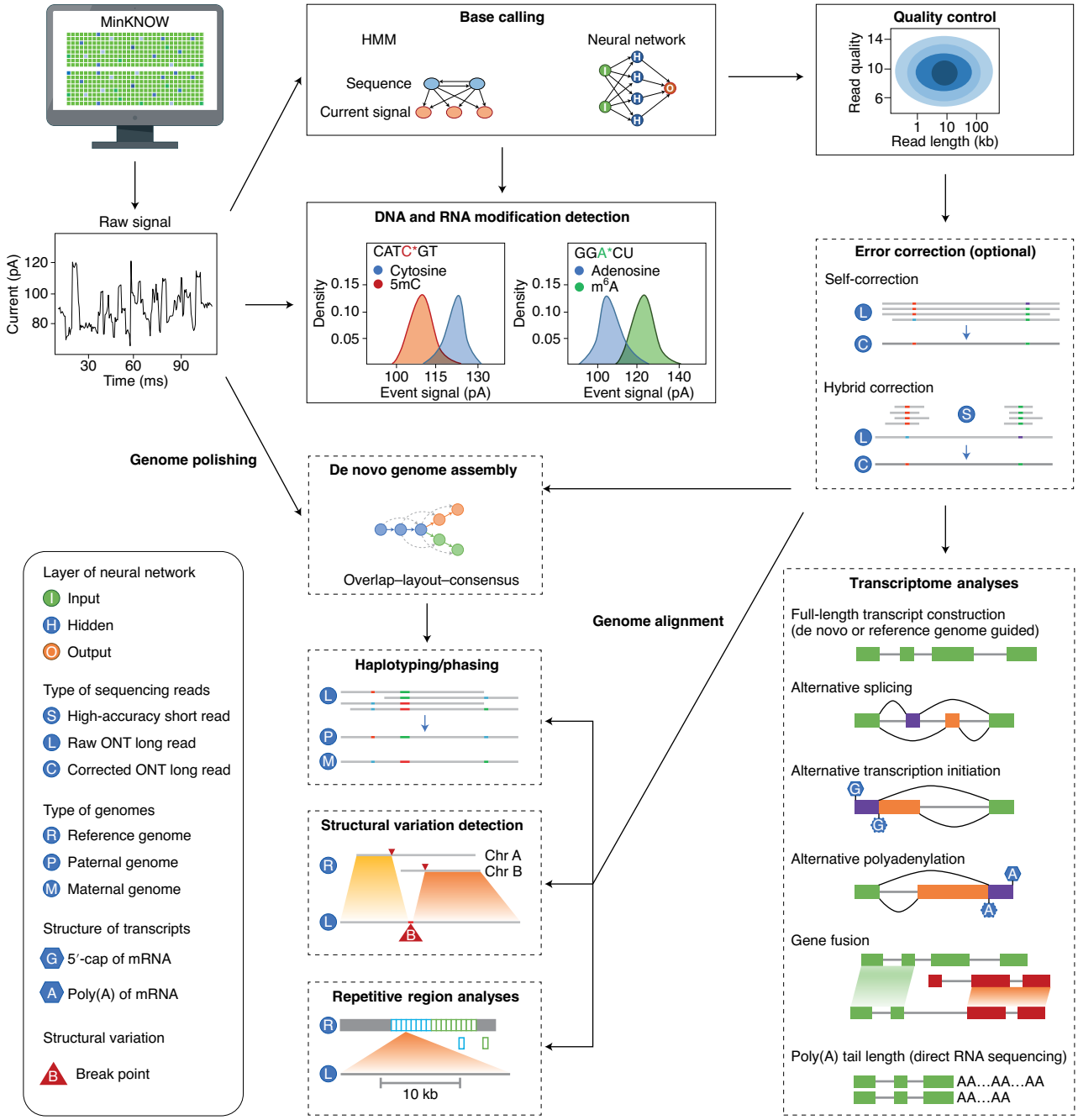

Fig. 4 | Analyses of ONT sequencing data. Typical bioinformatics analyses of ONT sequencing data, including the raw current data-specific approaches (for example, quality control, base calling and DNA/RNA modification detection), and error-prone long read-specific approaches (in dashed boxes; for example, error correction, de novo genome assembly, haplotyping/phasing, structural variation (SV) detection, repetitive region analyses and transcriptome analyses).

the first tool to identify the DNA modifications 5mC, 6mA and N^4-methylcytosine (4mC) from ONT data[74]. Several other DNA modification detection tools followed, including Nanopolish (5mC)[75], signalAlign (5mC, 5-hydroxymethylcytosine (5hmC) and 6mA)[71], mCaller (5mC and 6mA)[76], DeepMod (5mC and 6mA)[76], DeepSignal (5mC and 6mA)[77] and NanoMod (5mC and 6mA)[78]. Nanpolish, Megalodon and DeepSignal were recently benchmarked and confirmed to have high accuracy for 5mC detection with single-nucleotide resolution at the single-molecule level[79,80]. Compared to PacBio, ONT performs better in detecting 5mC but has lower accuracy in detecting 6mA[68,75,81].

The possibility of directly detecting N^6-methyladenosine (m^6A) modifications in RNA molecules was demonstrated using PacBio in 2012 (ref. [82]), although few follow-up applications were published. Recently, ONT direct RNA sequencing has yielded robust data of reasonable quality, and several pilot studies have detected bulk-level RNA modifications by examining either error distribution profiles (for example, EpiNano (m^6A)[73] and ELIGOS (m^6A and 5-methoxyuridine (5moU))[83]) or current signals (for example, Tombo extension (m^6A and m^5C)[74] and MINES (m^6A)[84]). However, detection of RNA modifications with single-nucleotide resolution at the single-molecule level has yet to be demonstrated.

Table 1 | Computational tools and experimental assays for ONT data analysis and applications

Computational tools for ONT data analysis

Data analysis	Computational tool	
Base calling	ONT's (https://github.com/nanoporetech/)	Guppy, Metrichor, Nanonet, Albacore, Scrappie, Flappie, Taiyaki, Bonito
	Third party	Nanocall[46], Chiron[70], Causalcall[35], DeepNano[47], DeepNano-blitz[246], basecRAWller[247], WaveNano[248], MinCall[249], URnano[250], CATCaller[251], SACall[252], Halcyon[253], Fast-Bonito[254]
Quality control	Prealignment	NanoPack[65], LongQC[255], PycoQC[256], MinIONQC[257], RabbitQC[258], NanoR[63], poRe[60], pyPore[66]
	Postalignment	AlignQC[42], SQANTI3[259], NanoOK[61], pyPore
Processing and visualization	Processing	Tombo (https://github.com/nanoporetech/tombo), Poretools[64], HPG pore[62]
	Visualization	SquiggleKit[260], BulkVis[49], Methylartist[261], NanoMethViz[262], Methplotlib[263]
DNA modification detection	5mC (CpG)	Nanopolish[75], Megalodon (https://github.com/nanoporetech/megalodon), DeepSignal[77], Tombo, Guppy, DeepMod[76], SignalAlign[71], NanoMod[78]
	5mC (GpC)	Nanopolish, NP-SMLR[72]
	6mA	Guppy, Tombo, mCaller[174], NanoMod, DeepMod, DeepSignal
	5hmC, 4mC	SignalAlign (5hmC), Tombo (4mC)
RNA modification detection	m^6A	EpiNano[73], xPore[264], MINES[84], Nanocompore[265], Nanom6A[266], Yanocomp[267], ELIGOS[83], DRUMMER[268], Tombo
	Pseudouridine (Ψ), m^5C	nanoRMS (Ψ)[104], Tombo (m^5C)
Error correction	Hybrid (graph based)	FMLRC[93], LoRDEC[91], Jabba[269], ECTools[270], HG-ColoR[271], NaS[272], Ratatosk[273]
	Hybrid (alignment based)	pacBioToCA[274], LSC[90], Nanocorr[45], proovread[275], Hercules[276], PBcR[274]
	Hybrid (dual alignment/graph based)	HALC[92], ColorMap[277]
	Self	Canu[88], daccord[278], LoRMA[89], MECAT[279], pbdagcon[280], FLAS[281], MARVEL[282], NanoReviser[283]
	Splice-aware	TALC (hybrid)[284], iSONcorrect (self)[285]
Genome alignment	Minimap2 (ref. [96]), GraphMap[95], NGMLR[98], LAST[97], BWA-MEM[286], BLAST[94], LRA[287], Winnowmap2 (ref. [288]), MashMap2 (ref. [289]), NanoBLASTer[290], mapAlign[291], GraphAligner[292], smsMap[293], lordFAST[294], S-conLSH[295], QAlign[296]	
	Splice-aware	Minimap2, GraphMap2 (ref. [102]), GMAP[100], STAR[101], deSALT[103], Magic-BLAST[297], Deep-Long[298], uLTRA[299]
Genome assembly	Canu, Miniasm[107], Flye[110], Redbean/wtdbg2 (ref. [111]), Falcon-Unzip[300], Shasta[164], Raven[301], NextDenovo (https://github.com/Nextomics/NextDenovo), Peregrine[302], HINGE[303], TULIP[304], NECAT[305]	
	Metagenome tailored	metaFlye[306], OPERA-MS (hybrid)[106]
	Haplotype-aware	Hifiasm[307]
Genome polishing	Nanopolish, Racon[308], Medaka (https://github.com/nanoporetech/medaka), NeuralPolish[309], PEPPER-Margin-DeepVariant[310], NextPolish (https://github.com/Nextomics/NextPolish), POLCA[311], HomoPolish[38]	
SV detection	Sniffles[98], SVIM[312], NanoSV[112], Picky[33], NanoVar[113], Dysgu[313], SENSV[314], cuteSV[315]	
SNV detection	LongShot[116], DeepVariant[310], iGDA[316], Nanopanel2 (ref. [317]), Clair[318]	
Haplotyping	WhatsHap[117], Medaka, HapCUT2 (ref. [319]), flopp[320], DR2S[321], Nanopanel2, iGDA, Clair	
Repetitive element analysis	Non-reference transposable element detection	TLDR[118], PALMER[322], TELR (https://github.com/bergmanlab/TELR)
	Tandem repeat	TRiCoLOR[119], STRique[323], NanoSatellite[207]
Transcriptome construction and quantification	De novo	RATTLE[129], CARNAC-LR[324], isONclust[325], IDP-denovo (hybrid)[128]
	Reference genome guide	IDP (hybrid)[127], TALON[126], FLAIR[123], StringTie2 (ref. [125]), FLAMES[326]
	Quantification only	LIQA[327], AERON[328], Mili (https://github.com/Augroup/Mili)
Transcriptome characterization	Alternative splicing	FLAIR
	Gene fusion	IDP-fusion (hybrid)[329], JAFFAL[330], AERON, LongGF[331]
	Circular RNA	CIRI-long[170]
	Poly(A) tail length	Nanopolish
	Allele-specific expression	IDP-ASE (hybrid)[332], LORALS[333]

Continued

Table 1 | Computational tools and experimental assays for ONT data analysis and applications (continued)

Computational tools for ONT data analysis		
Experimental assays for ONT applications		
Application		**Experimental assay**
Genomics	Amplicon sequencing	NanoAmpli-seq[334], Dual-UMI-tagging[335]
	Targeted genome sequencing	nCATS[336], CATCH[197]
Epigenomics	DNA methylation and chromatin accessibility	MeSMLR-seq[72], SMAC-seq[175], nanoNOMe[176]
	Nucleosome occupancy	MeSMLR-seq
	Histone modification	DiMeLo-seq[181], BIND&MODIFY[182]
	3D genome structure	Pore-C[180]
Protein–DNA interaction mapping		DiMeLo-seq, Nanopore-DamID[337]
DNA replication (replication fork detection)		D-Nascent[179], FORK-seq[338]
Single-cell transcriptome		ScCOLOR-seq[339], ScISOr-Seq[340], scNaUmi-seq[341], FLT-seq[326]
Epitranscriptomics (RNA secondary structure)		PORE-cupine[184], SMS-seq[342]
RNA metabolism (nascent RNA detection)		Nano-ID[186], nano-COP[343]

3D, three dimensional; SNV, single nucleotide variation.

Error correction. Although the average accuracy of ONT sequencing is improving, certain subsets of reads or read fragments have very low accuracy, and the error rates of both 1D reads and 2D/1D^2 reads are still much higher than those of short reads generated by next-generation sequencing technologies. Thus, error correction is widely applied before many downstream analyses (for example, genome assembly and gene isoform identification), which can rescue reads for higher sensitivity (for example, mappability[85]) and improve the quality of the results (for example, break point determination at single-nucleotide resolution[86]). Two types of error correction algorithms are used[85,87] (Fig. 4, middle right, and Table 1): 'self-correction' uses graph-based approaches to produce consensus sequences among different molecules from the same origins (for example, Canu[88] and LoRMA[89]) in contrast to 2D and 1D^2 reads generated from the same molecules, and 'hybrid correction' uses high-accuracy short reads to correct long reads by alignment-based (for example, LSC[90] and Nanocorr[45]), graph-based (for example, LorDEC[91]) and dual alignment/graph-based algorithms (for example, HALC[92]). Recently, two benchmark studies demonstrated that the existing hybrid error correction tools (for example, FMLRC[93], LSC and LorDEC) together with sufficient short-read coverage can reduce the long-read error rate to a level (~1–4%) similar to that of short reads[85,87], whereas self-correction reduces the error rate to ~3–6% (ref. [87]), which may be due to non-random systematic errors in ONT data.

Aligners for error-prone long reads. Alignment tools have been developed to tackle the specific characteristics of error-prone long reads (Table 1). Very early aligners (for example, BLAST[94]) were developed for small numbers of long reads (for example, Sanger sequencing data). More recently, there has been considerable growth in alignment methods for high-throughput accurate short reads (for example, Illumina sequencing data) in response to the growth in next-generation sequencing. Development of several error-prone long-read aligners was initially motivated by PacBio data, and they were also tested on ONT data. In 2016, the first aligner specifically for ONT reads, GraphMap, was developed[95]. GraphMap progressively refines candidate alignments to handle high error rates and uses fast graph transversal to align long reads with high speed and precision. Using a seed–chain–align procedure, minimap2 was developed to match increases in ONT read length beyond 100 kb (ref. [96]). A recent benchmark paper revealed that minimap2 ran much faster than other long-read aligners (that is, LAST[97], NGMLR[98] and GraphMap) without sacrificing the accuracy[99]. In addition, minimap2 can perform splice-aware alignment for ONT cDNA or direct RNA-sequencing reads.

In addition to minimap2, GMAP, published in 2005 (ref. [100]), and a new mode of STAR, which was originally developed for short reads[101], have been widely used in splice-aware alignment of error-prone transcriptome long reads to genomes. Other aligners have also been developed, such as Graphmap2 (ref. [102]) and deSALT[103], for ONT transcriptome data. Especially for ONT direct RNA-sequencing reads with dense base modifications, Graphmap2 has a higher alignment rate than minimap2 (ref. [104]).

Hybrid sequencing. Many applications combine long reads and short reads in the bioinformatics analyses, termed hybrid sequencing. In contrast to hybrid correction of long reads for general purposes, many hybrid sequencing-based methods integrate long reads and short reads into the algorithms and pipeline designs to harness the strengths of both types of reads to address specific biological problems. The long-read length is well suited to identifying large-range genomic complexity with unambiguous alignments, whereas the high accuracy and high throughput of short reads is useful for characterizing local details (for example, splice site detection with single-nucleotide resolution) and improving quantitative analyses. For example, genome[105], transcriptome[42] and metagenome[106] assemblies have shown superior performance with hybrid sequencing data compared to either error-prone long reads alone or high-accuracy short reads alone.

De novo genome assembly. Error-prone long reads have been used for de novo genome assembly. Assemblers (Table 1) such as Canu[88] and Miniasm[107] are based on the overlap–layout–consensus algorithm, which builds a graph by overlapping similar sequences and is robust to sequencing error[58,67,108] (Fig. 4, middle center). To further remove errors, error correction of long reads and polishing of assembled draft genomes (that is, improving accuracy of consensus sequences using raw current data) are often performed before and after assembly, respectively. In addition to the genome-polishing software Nanopolish[109], ONT released Medaka, a neural network-based method, aiming for improved accuracy and speed compared to Nanopolish (Table 1).

These approaches take into account not only general assembly performance but also certain specific aspects, such as complex genomic regions and computational intensity. For example, Flye improves genome assembly at long and highly repetitive regions by constructing an assembly graph from concatenated disjoint

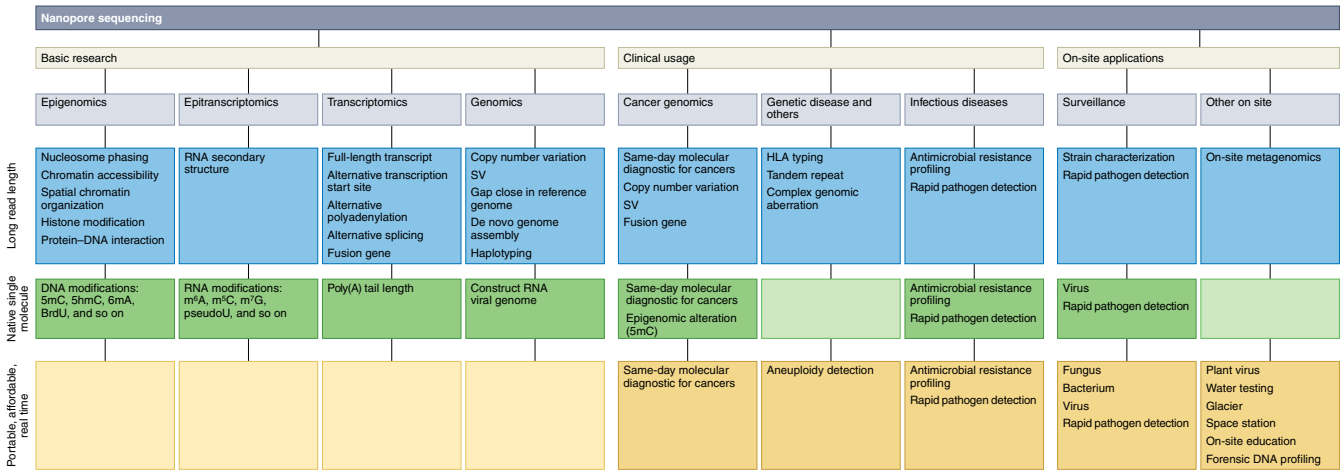

Fig. 5 | Applications of ONT sequencing. ONT sequencing applications are classified into three major groups (basic research, clinical usage and on-site applications) and are shown as a pie chart. The classifications are further categorized by specific topics, and the slice area is proportional to the number of publications (in $\log_2$ scale). Some applications span two categories, such as SV detection and rapid pathogen detection. The applications are also organized by the corresponding strengths of ONT sequencing as three layers of the pie chart: (1) long read length, (2) native single molecule and (3) portable, affordable and real time. The width of each layer is proportional to the number of publications (in $\log_2$ scale). Some applications that use all three strengths span all three layers (for example, antimicrobial resistance profiling). 'Fungus' includes *Candida auris*, 'bacterium' includes *Salmonella*, *Neisseria meningitidis* and *Klebsiella pneumoniae* and 'virus' includes severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, Zika, Venezuelan equine encephalitis, yellow fever, Lassa fever and dengue; HLA, human leukocyte antigens.

genomic segments[110]; Miniasm uses all-versus-all read self-mapping for ultrafast assembly[107], although postassembly polishing is necessary for higher accuracy. The recently developed assembler wtdbg2 runs much faster than other tools without sacrificing contiguity and accuracy[111].

SVs and repetitive regions. When a reference genome is available, ONT data can be used to study sample-specific genomic details, including SVs and haplotypes, with much higher precision than other techniques. A few SV detection tools have been developed (for example, NanoSV[112], Sniffles[98], Picky[33] and NanoVar[113]) (Fig. 4, bottom center, and Table 1). Picky, in addition to detecting regular SVs, also reveals enriched short-span SVs (~300 bp) in repetitive regions, as long reads cover the entire region including the variations. Given that single long reads can encompass multiple variants, including both SNVs and SVs, it is possible to perform phasing of multiploid genomes as well as other haplotype-resolved analyses[112,114,115] with appropriate bioinformatics software, such as LongShot[116] for SNV detection and WhatsHap[117] for haplotyping/phasing.

Several tools have also been developed to investigate highly repetitive genomic regions by ONT sequencing, such as TLDR for identifying non-reference transposable elements[118] and TRiCoLOR for characterizing tandem repeats[119] (Table 1).

Transcriptome complexity. When used in transcriptome analyses, ONT reads can be clustered and assembled to reconstruct full-length gene isoforms or aligned to a reference genome to characterize complex transcriptional events[42,120–123] (Fig. 4, bottom right). In particular, several transcript assemblers have been developed specifically for error-prone long reads, such as Traphlor[124], FLAIR[123], StringTie2 (ref. [125]) and TALON[126] as well as several based on hybrid sequencing data (for example, IDP[127]). In particular, IDP-denovo[128] and RATTLE[129] can perform de novo transcript assembly by long reads without a reference genome. More recently, ONT direct RNA sequencing has made transcriptome-wide investigation of native RNA molecules feasible[52,130,131]. However, development of corresponding bioinformatics tools, especially for quantitative analyses, remains inadequate.

Applications of nanopore sequencing

The long read length, portability and direct RNA sequencing capability of ONT devices have supported a diverse range of applications (Fig. 5). We review 11 applications that are the subject of the most publications since 2015.

Closing gaps in reference genomes. Genome assembly is one of the main uses of ONT sequencing (~30% of published ONT applications; Fig. 5). For species with available reference genomes, ONT long reads are useful for closing genome gaps, especially in the human genome. For example, ONT reads have been used to close 12 gaps (>50 kb for each gap) in the human reference genome and to measure the length of telomeric repeats[132] and also to assemble the centromeric region of the human Y chromosome[133]. Moreover, ONT enabled the first gapless telomere-to-telomere assembly of the human X chromosome, including reconstruction of a ~2.8 Mb centromeric satellite DNA array and closing of all remaining 29 gaps (totaling 1.1 Mb)[134]. The Telomere-to-Telomere Consortium reported the first complete human genome (T2T-CHM13) of the size 3.055 Gb (ref. [135]).

The *Caenorhabditis elegans* reference genome has also been expanded by >2 Mb through accurate identification of repetitive regions using ONT long reads[136]. Similar progress has been achieved in other model organisms and closely related species (for example, *Escherichia coli*[109], *Saccharomyces cerevisiae*[137], *Arabidopsis thaliana*[138] and 15 *Drosophila* species[139]) as well as in non-model organisms, including characterizing large tandem repeats in the bread wheat genome[140] and improving the continuity and completeness of the genome of *Trypanosoma cruzi* (the parasite causing Chagas disease)[141].

Building new reference genomes. ONT long reads have been used extensively to assemble the initial reference genomes of many non-model organisms. For instance, ONT data alone were used to assemble the first genome of *Rhizoctonia solani* (a pathogenic fungal species that causes damping-off diseases in a wide range of crops)[142], and hybrid sequencing data (ONT plus Illumina) were used to assemble the first draft genomes of *Maccullochella peelii*

(Australia's largest freshwater fish)[143] and *Amphiprion ocellaris* (the common clown fish)[144]. In more complicated cases, ONT long reads have been integrated with one or more other techniques (for example, Illumina short reads, PacBio long reads, 10x Genomics linked reads, optical mapping by Bionano Genomics and spatial distance by Hi-C) to assemble the initial reference genomes of many species, such as *Maniola jurtina* (the meadow brown butterfly, a model for ecological genetics)[145], *Varanus komodoensis* (the largest extant monitor lizard)[146], *Pavo cristatus* (the national bird of India)[147], *Panthera leo* (the lion)[148] and *Eumeta variegate* (a bagworm moth that produces silk with potential in biomaterial design)[149]. In addition, ONT direct RNA sequencing has been used to construct RNA viral genomes while eliminating the need for the conventional reverse transcription step, including Mayaro virus[150], Venezuelan equine encephalitis virus[150], chikungunya virus[150], Zika virus[150], vesicular stomatitis Indiana virus[150], Oropouche virus[150], influenza A[53] and human coronavirus[86]. For small DNA/RNA viral genomes (for example, the 27-kb human coronavirus genome[86]), the assembly process is not required given the long read length.

In the SARS-CoV-2 pandemic[151], ONT sequencing was used to reconstruct full-length SARS-CoV-2 genome sequences via cDNA and direct RNA sequencing[152–155], providing valuable information regarding the biology, evolution and pathogenicity of the virus.

The increasing yield, read length and accuracy of ONT data enable much more time- and cost-efficient genome assembly of all sizes of genomes, from bacteria of several megabases[109], fruit fly[139,156], fish[143,144,157], blood clam[158], banana[159], cabbage[159] and walnut[160,161], all of whose genomes are in the hundreds of megabases, as well as the Komodo dragon[146], Steller sea lion[162], lettuce (https://nanoporetech.com/resource-centre/tip-iceberg-sequencing-lettuce-genome) and giant sequoia[163], with genomes of a few gigabases, to coast redwood (https://www.savetheredwoods.org/project/redwood-genome-project/) and tulip (https://nanoporetech.com/resource-centre/beauty-and-beast), with genomes of 27–34 Gb. Only three PromethION flow cells were required to sequence the human genome, requiring <6 h for the computational assembly[164].

Identifying large SVs. A powerful application of ONT long reads is to identify large SVs (especially from humans) in biomedical contexts, such as the breast cancer cell line HCC1187 (ref. [33]), individuals with acute myeloid leukemia[113], the construction of the first haplotype-resolved SV spectra for two individuals with congenital abnormalities[112] and the identification of 29,436 SVs from a Yoruban individual NA19240 (ref. [165]).

Characterizing full-length transcriptomes and complex transcriptional events. A comprehensive examination of the feasibility of ONT cDNA sequencing (with R7 and R9 nanopores) in transcriptome analyses demonstrated its similar performance in gene isoform identification to PacBio long reads, both of which are superior to Illumina short reads[42]. With ONT data alone, there remain drawbacks in estimating gene/isoform abundance, detecting splice sites and mapping alternative polyadenylation sites, although recent improvements in accuracy and throughput have advanced these analyses. Nevertheless, ONT cDNA sequencing was also tested in individual B cells from mice[120] and humans[122,166]. Furthermore, ONT direct RNA sequencing has been used to measure the poly(A) tail length of native RNA molecules in humans[131], *C. elegans*[167], *A. thaliana*[168] and *Locusta migratoria*[169], corroborating a negative correlation between poly(A) tail length and gene expression[167,168]. In addition, the full-length isoforms of human circular RNAs have been characterized by ONT sequencing following rolling circle amplification[170,171].

Characterizing epigenetic marks. As early as 2013, independent reports demonstrated that methylated cytosines (5mC and 5hmC) in DNA could be distinguished from native cytosine by the characteristic current signals measured using the MspA nanopore[172,173]. Later, bioinformatics tools were developed to identify three kinds of DNA modifications (6mA, 5mC and 5hmC) from ONT data[71,75]. Recently, ONT was applied to characterize the methylomes from different biological samples, such as 6mA in a microbial reference community[174] as well as 5mC and 6mA in *E. coli*, *Chlamydomonas reinhardtii* and human genomes[76].

Mapping DNA modifications using ONT sequencing in combination with exogenous methyltransferase treatment (inducing 5mC at GpC sites) led to the development of an experimental and bioinformatics approach, MeSMLR-seq, that maps nucleosome occupancy and chromatin accessibility at the single-molecule level and at long-range scale in *S. cerevisiae*[72] (Table 1). Later, another method, SMAC-seq adopted the same strategy with the additional exogenous modification 6mA to improve the resolution of mapping nucleosome occupancy and chromatin accessibility[175]. Similarly, multiple epigenetic features, including the endogenous 5mC methylome (at CpG sites), nucleosome occupancy and chromatin accessibility, can be simultaneously characterized on single long human DNA molecules by MeSMLR-seq (K.F.A., unpublished data, and ref. [176]). Such epigenome analyses can be performed in a haplotype-resolved manner and thus will be informative for discovering allele-specific methylation linked to imprinted genes as well as for phasing genomic variants and chromatin states, even in heterogeneous cancer samples.

Similarly, several other methods have combined various biochemical techniques with ONT sequencing (Table 1). For example, the movement of DNA replication forks on single DNA molecules has been measured by detection of nucleotide analogs (for example, 5-bromodeoxyuridine (5-BrdU)) using ONT sequencing[177–179], and the 3D chromatin organization in human cells has been analyzed by integrating a chromatin conformation capture technique and ONT sequencing to capture multiple loci in close spatial proximity by single reads[180]. Two other experimental assays, DiMeLo-seq[181] and BIND&MODIFY[182], use ONT sequencing to map histone modifications (H3K9me3 and H3K27me3), a histone variant (CENP-A) and other specific protein–DNA interactions (for example, CTCF binding profile). They both construct a fusion protein of the adenosine methyltransferase and protein A to convert specific protein–DNA interactions to an artificial 6mA profile, which is subsequently detected by ONT sequencing.

Detecting RNA modifications. Compared to existing antibody-based approaches (which are usually followed by short-read sequencing), ONT direct RNA sequencing opens opportunities to directly identify RNA modifications (for example, m^6A) and RNA editing (for example, inosine), which have critical biological functions. In 2018, distinct ionic current signals for unmodified and modified bases (for example, m^6A and m^5C) in ONT direct RNA-sequencing data were reported[52]. Since then, epitranscriptome analyses using ONT sequencing have progressed rapidly, including detection of 7-methylguanosine (m^7G) and pseudouridine in 16S rRNAs of *E. coli*[183], m^6A in mRNAs of *S. cerevisiae*[73] and *A. thaliana*[168] and m^6A[130] and pseudouridine[104] in human RNAs. Recent independent research (K.F.A., unpublished data, and refs. [184,185]) has revealed that it is possible to probe RNA secondary structure using a combination of ONT direct RNA sequencing and artificial chemical modifications (Table 1). The dynamics of RNA metabolism were also analyzed by labeling nascent RNAs with base analogs (for example, 5-ethynyluridine[186] and 4-thiouridine[187]) followed by ONT direct RNA sequencing (Table 1).

Cancer. ONT sequencing has been applied to many cancer types, including leukemia[188–192], breast[33,176,193], brain[193], colorectal[194], pancreatic[195] and lung[196] cancers, to identify genomic variants of interests,

especially large and complex ones. For example, ONT amplicon sequencing was used to identify *TP53* mutations in 12 individuals with chronic lymphoblastic leukemia[188]. Likewise, MinION sequencing data revealed *BCR-ABL1* kinase domain mutations in 19 individuals with chronic myeloid leukemia and 5 individuals with acute lymphoblastic leukemia with superior sensitivity and time efficiency compared to Sanger sequencing[189]. Additionally, ONT whole-genome sequencing was used to rapidly detect chromosomal translocations and precisely determine the breakpoints in an individual with acute myeloid leukemia[192].

A combination of Cas9-assisted target enrichment and ONT sequencing has characterized a 200-kb region spanning the breast cancer susceptibility gene *BRCA1* and its flanking regions despite a high repetitive sequence fraction (>50%) and large gene size (~80 kb)[197]. This study provided a template for the analysis of full variant profiles of disease-related genes.

The ability to directly detect DNA modifications using ONT data has enabled the simultaneous capture of genomic (that is, copy number variation) and epigenomic (that is, 5mC) alterations using only ONT data from brain tumor samples[193]. The whole workflow (from sample collection to bioinformatics results) was completed in a single day, delivering a multimodal and rapid molecular diagnostic for cancers. In addition, same-day detection of fusion genes in clinical specimens has also been demonstrated by MinION cDNA sequencing[198].

Infectious disease. Because of its fast real-time sequencing capabilities and small size, MinION has been used for rapid pathogen detection, including diagnosis of bacterial meningitis[199], bacterial lower respiratory tract infection[200], infective endocarditis[201], pneumonia[202] and infection in prosthetic joints[203]. In the example of bacterial meningitis, 16S amplicon sequencing took only 10 min using MinION to identify pathogenic bacteria in all six retrospective cases, making MinION particularly useful for the early administration of antibiotics through timely detection of bacterial infections[199]. Likewise, clinical diagnosis of bacterial lower respiratory tract infection using MinION was faster (6 h versus >2 d) and had higher sensitivity than existing culture-based 'gold standard' methods[200].

In addition to pathogen detection, ONT sequencing can accelerate profiling antibiotic/antimicrobial resistance in bacteria and other microbes. For example, MinION was used to identify 51 acquired resistance genes directly from clinical urine samples (without culture) of 55 that were detected from cultivated bacteria using Illumina sequencing[204], and a recent survey of resistance to colistin in 12,053 *Salmonella* strains used a combination of ONT, PacBio and Illumina data[205]. Indeed, ONT sequencing is useful for detecting specific species and strains (for example, virulent ones) from microbiome samples given the unambiguous mappability of longer reads, which provides accurate estimates of microbiome composition compared to the conventional studies relying on 16S rRNA and DNA amplicons[57,206].

Genetic disease. ONT long reads have been applied to characterize complex genomic rearrangements in individuals with genetic disorders. For example, ONT sequencing of human genomes revealed that an expansion of tandem repeats in the *ABCA7* gene was associated with an increased risk of Alzheimer's disease[207]. ONT sequencing was also used to discover a new 3.8-Mb duplication in the intronic region of the *F8* gene in an individual with hemophilia A[208]. Other examples cover a large range of diseases and conditions, including autism spectrum disorder[209], Temple syndrome[210], congenital abnormalities[112], glycogen storage disease type Ia (ref. [211]), intellectual disability and seizures[212], epilepsy[213,214], Parkinson's disease[215], Gaucher disease[215], ataxia-pancytopenia syndrome and severe immune dysregulation[114].

In another clinical application, human leukocyte antigen genotyping benefited from the improved accuracy of the R9.5 nanopore[216–218]. MinION enabled the detection of aneuploidy in prenatal and miscarriage samples in 4 h compared to 1–3 weeks with conventional techniques[219].

Outbreak surveillance. The portable MinION device allows in-field and real-time genomic surveillance of emerging infectious diseases, aiding in phylogenetic and epidemiological investigations such as characterization of evolution rate, diagnostic targets, response to treatment and transmission rate. In April 2015, MinION devices were shipped to Guinea for real-time genomic surveillance of the ongoing Ebola outbreak. Only 15–60 min of sequencing per sample was required[220]. Likewise, a hospital outbreak of *Salmonella* was monitored with MinION, with positive cases identified within 2 h (ref. [221]). MinION was also used to conduct genomic surveillance for Zika virus[222], yellow fever virus[223] and dengue virus[224] outbreaks in Brazil.

With the increasing throughput of ONT sequencing, real-time surveillance has been applied to pathogens with larger genomes over the years, ranging from viruses of a few kilobases (for example, Ebola virus[220], 18–19 kb; Zika virus[222], 11 kb; Venezuelan equine encephalitis virus[225], 11.4 kb; Lassa fever virus[226], 10.4 kb and SARS-CoV-2 coronavirus[151], 29.8 kb) to bacteria of several megabases (for example, *Salmonella*[221], 5 Mb; *N. meningitidis*[227], 2 Mb and *K. pneumoniae*[228], 5.4 Mb) and to human fungal pathogens with genomes of >10 Mb (for example, *Candida auris*[229], 12 Mb).

Other on-site applications. Portable ONT devices have also been used for on-site metagenomics research. MinION characterized pathogenic microbes, virulence genes and antimicrobial resistance markers in the polluted Little Bighorn River, Montana, United States[230]. MinION and MinIT devices were brought to farms in sub-Saharan Africa for early and rapid diagnosis (<3 h) of plant viruses and pests in cassava[231]. In forensic research, a portable strategy known as 'MinION sketching' was developed to identify human DNA with only 3 min of sequencing[232], offering a rapid solution to cell authentication or contamination identification during cell or tissue culture.

The portability of the MinION system, which consists of the palm-sized MinION, mobile DNA extraction devices (for example, VolTRAX and Bento Lab) and real-time onboard base calling with Guppy and other offline bioinformatics tools, enables field research in scenarios where samples are hard to culture or store or where rapid genomic information is needed[233]. Examples include the International Space Station, future exploration of Mars and the Moon involving microgravity and high levels of ionizing radiation[69,234,235], ships[236], Greenland glaciers at subzero temperatures[237], conservation work in the Madagascar forest[238] and educational outreach[238].

Outlook

Nanopore sequencing has enabled many biomedical studies by providing ultralong reads from single DNA/RNA molecules in real time. Nonetheless, current ONT sequencing techniques have several limitations, including relatively high error rates and the requirement for relatively high amounts of nucleic acid material. Overcoming these challenges will require further breakthroughs in nanopore technology, molecular experiments and bioinformatics software.

The principal concern in many applications is the error rate, which, at 6–15% for the R9.4 nanopore, is still much higher than that of Illumina short-read sequencing (0.1–1%). Despite substantial improvements in data accuracy over the past 7 years, there may be an intrinsic limit to 1D read accuracy. The sequencing of single molecules has a low signal-to-noise ratio, in contrast to bulk sequencing of molecules as in Illumina sequencing. Indeed, the same issue arises in the other single-molecule measurement techniques, such as Helicos, PacBio and BioNano Genomics. There

is currently no theoretical estimation of this limit, but for reference, Helicos managed to reduce error rates to 4% (ref. [239]). Future improvements in accuracy can be expected through optimization of molecule translocation ratcheting and, in particular, through engineering existing nanopores or discovering new ones. Indeed, many studies have been exploring new biological or non-biological nanopores with shorter sensing regions to achieve context-independent and high-quality raw signals. For example, graphene-based nanopores are capable of DNA sensing and have high durability and insulating capability in high ionic strength solutions[240–242], where their thickness (~0.35 nm) is ideal for capturing single nucleotides[243]. Because such context-independent signals minimize the complex signal interference between adjacent modified bases, they could also make it possible to detect base modifications at single-molecule and single-nucleotide resolutions. Another approach for improving 1D read accuracy is to develop base-calling methods based on advanced computational techniques, such as deep learning.

Repetitive sequencing of the same molecule, for example, using 2D and $1D^2$ reads, was helpful in improving accuracy. However, both of these approaches were limited in that each molecule could only be measured twice. By contrast, the R2C2 protocol involves the generation and sequencing of multiple copies of target molecules[122]. It may also be possible to increase data accuracy by recapturing DNA molecules into the same nanopore[244] or by using multilayer nanopores for multiple sequencing of each molecule.

Improved data accuracy would advance single-molecule omics studies. Haplotype-resolved genome assembly has been demonstrated for PacBio data[245], which could likely be achieved using ONT sequencing. Methods are being developed to characterize epigenomic and epitranscriptomic events beyond base modifications at the single-molecule level, such as nucleosome occupancy and chromatin accessibility[72,175,176] and RNA secondary structure[184,185]. These approaches would allow investigation of the heterogeneity and dynamics of the epigenome and epitranscriptome as well as analysis of allele-specific and/or strand-specific epigenomic and epitranscriptomic phenomena. They would require specific experimental protocols (for example, identifying chromatin accessibility by detecting artificial 5mC footprints[72,175,176]) rather than the simple generation of long reads.

Although the ultralong read length of ONT data remains its principal strength, further increases in read length would be beneficial, further facilitating genome assembly and the sequencing of difficult to analyze genomic regions (for example, eukaryotic centromeres and telomeres). Once read lengths reach a certain range, or even cover entire chromosomes, genome assembly would become trivial, requiring little computation and having superior completeness and accuracy. Personalized genome assembly would become widely available, and it would be possible to assemble the genomes of millions of species across the many Earth ecosystems. Obtaining megabase-scale or longer reads will require the development of HMW DNA extraction and size selection methods as well as protocols to maintain ultralong DNA fragments intact.

The other key experimental barrier to be addressed is the large amount of input DNA and RNA required for ONT sequencing, which is up to a few micrograms of DNA and hundreds of nanograms of RNA. PCR amplification of DNA is impractical for very long reads or impermissible for native DNA/RNA sequencing. Reducing the sample size requirement would make ONT sequencing useful for the many biomedical studies in which genetic material is limited. In parallel, ONT sequencing will benefit from the development of an end-to-end system. For example, the integration and automation of DNA/RNA extraction systems, sequencing library preparation and loading systems would allow users without specific training to generate ONT sequencing data. More robust and user-friendly bioinformatics software, such as cloud storage and computing and real-time analysis, will provide a further boost to ONT sequencing applications, ultimately moving the technology beyond the lab and into daily life.

Received: 9 December 2019; Accepted: 22 September 2021;
Published online: 8 November 2021

References

1. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
2. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
3. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).
4. Yang, Y. et al. Advances in nanopore sequencing technology. *J. Nanosci. Nanotechnol.* **13**, 4521–4538 (2013).
5. Maitra, R. D., Kim, J. & Dunbar, W. B. Recent advances in nanopore sequencing. *Electrophoresis* **33**, 3418–3428 (2012).
6. Leggett, R. M. & Clark, M. D. A world of opportunities with nanopore sequencing. *J. Exp. Bot.* **68**, 5419–5429 (2017).
7. Noakes, M. T. et al. Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nat. Biotechnol.* **37**, 651–656 (2019).
8. Branton, D. et al. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).
9. Song, L. et al. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1866 (1996).
10. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
11. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* **77**, 3227–3233 (1999).
12. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. & Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl Acad. Sci. USA* **97**, 1079–1084 (2000).
13. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl Acad. Sci. USA* **106**, 7702–7707 (2009).
14. Stoddart, D. et al. Nucleobase recognition in ssDNA at the central constriction of the α-hemolysin pore. *Nano Lett.* **10**, 3633–3637 (2010).
15. Stoddart, D., Maglia, G., Mikhailova, E., Heron, A. J. & Bayley, H. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew. Chem. Int. Ed. Engl.* **49**, 556–559 (2010).
16. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl Acad. Sci. USA* **105**, 20647–20652 (2008).
17. Derrington, I. M. et al. Nanopore DNA sequencing with MspA. *Proc. Natl Acad. Sci. USA* **107**, 16060–16065 (2010).
18. Niederweis, M. et al. Cloning of the mspA gene encoding a porin from *Mycobacterium smegmatis*. *Mol. Microbiol.* **33**, 933–945 (1999).
19. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189–1192 (2004).
20. Benner, S. et al. Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nanotechnol.* **2**, 718–724 (2007).
21. Hornblower, B. et al. Single-molecule analysis of DNA–protein complexes using nanopores. *Nat. Methods* **4**, 315–317 (2007).
22. Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J. Am. Chem. Soc.* **130**, 818–820 (2008).
23. Lieberman, K. R. et al. Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J. Am. Chem. Soc.* **132**, 17961–17972 (2010).
24. Cherf, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
25. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
26. Mason, C. E. & Elemento, O. Faster sequencers, larger datasets, new challenges. *Genome Biol.* **13**, 314 (2012).
27. Wang, Y., Yang, Q. & Wang, Z. The evolution of nanopore sequencing. *Front. Genet.* **5**, 449 (2015).
28. Shi, W., Friedman, A. K. & Baker, L. A. Nanopore sensing. *Anal. Chem.* **89**, 157–188 (2017).
29. Minei, R., Hoshina, R. & Ogura, A. De novo assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genomics* **19**, 700 (2018).

30. Ashton, P. M. et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* **33**, 296–300 (2015).
31. Carter, J. M. & Hussain, S. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res* **2**, 23 (2017).
32. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
33. Gong, L. et al. Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nat. Methods* **15**, 455–460 (2018).
34. Brickwedde, A. et al. Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357^{T}. *Front. Microbiol.* **9**, 1786 (2018).
35. Zeng, J. et al. Causalcall: nanopore basecalling using a temporal convolutional network. *Front. Genet.* **10**, 1332 (2020).
36. Helmersen, K. & Aamot, H. V. DNA extraction of microbial DNA directly from infected tissue: an optimized protocol for use in nanopore sequencing. *Sci. Rep.* **10**, 2985 (2020).
37. Tytgat, O. et al. Nanopore sequencing of a forensic STR multiplex reveals loci suitable for single-contributor STR profiling. *Genes* **11**, 381 (2020).
38. Huang, Y. T., Liu, P. Y. & Shih, P. W. Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing. *Genome Biol.* **22**, 95 (2021).
39. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
40. Ip, C. L. C. et al. MinION Analysis and Reference Consortium: phase 1 data release and analysis. *F1000Res* **4**, 1075 (2015).
41. Jain, M. et al. MinION Analysis and Reference Consortium: phase 2 data release and analysis of R9.0 chemistry. *F1000Res* **6**, 760 (2017).
42. Weirather, J. L. et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* **6**, 100 (2017).
43. Seki, M. et al. Evaluation and application of RNA-seq by MinION. *DNA Res.* **26**, 55–65 (2019).
44. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**, 90 (2018).
45. Goodwin, S. et al. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1750–1756 (2015).
46. David, M., Dursi, L. J., Yao, D., Boutros, P. C. & Simpson, J. T. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics* **33**, 49–55 (2017).
47. Boza, V., Brejova, B. & Vinar, T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS ONE* **12**, e0178751 (2017).
48. Gong, L., Wong, C. H., Idol, J., Ngan, C. Y. & Wei, C. L. Ultra-long read sequencing for whole genomic DNA analysis. *J. Vis. Exp.* https://doi.org/10.3791/58954 (2019).
49. Payne, A., Holmes, N., Rakyan, V. & Loose, M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* **35**, 2193–2198 (2019).
50. Quick, J. & Loman, N. J. in *Nanopore Sequencing: An Introduction* Ch. 7 (World Scientific Press, 2019).
51. Deschamps, S. et al. A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *Nat. Commun.* **9**, 4844 (2018).
52. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
53. Keller, M. W. et al. Direct RNA sequencing of the coding complete influenza A virus genome. *Sci. Rep.* **8**, 14408 (2018).
54. Pitt, M. E. et al. Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA Nanopore sequencing. *Gigascience* **9**, giaa002 (2020).
55. Cartolano, M., Huettel, B., Hartwig, B., Reinhardt, R. & Schneeberger, K. cDNA library enrichment of full length transcripts for SMRT long read sequencing. *PLoS ONE* **11**, e0157779 (2016).
56. Chen, Y. et al. A systematic benchmark of Nanopore long read RNA sequencing for transcript level analysis in human cell lines. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.21.440736 (2021).
57. Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* **8**, giz043 (2019).
58. Magi, A., Semeraro, R., Mingrino, A., Giusti, B. & D'Aurizio, R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief. Bioinform.* **19**, 1256–1272 (2018).
59. Cao, M. D., Ganesamoorthy, D., Cooper, M. A. & Coin, L. J. Realtime analysis and visualization of MinION sequencing data with npReader. *Bioinformatics* **32**, 764–766 (2016).
60. Watson, M. et al. poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics* **31**, 114–115 (2015).
61. Leggett, R. M., Heavens, D., Caccamo, M., Clark, M. D. & Davey, R. P. NanoOK: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics* **32**, 142–144 (2016).
62. Tarraga, J., Gallego, A., Arnau, V., Medina, I. & Dopazo, J. HPG pore: an efficient and scalable framework for nanopore sequencing data. *BMC Bioinformatics* **17**, 107 (2016).
63. Bolognini, D., Bartalucci, N., Mingrino, A., Vannucchi, A. M. & Magi, A. NanoR: a user-friendly R package to analyze and compare nanopore sequencing data. *PLoS ONE* **14**, e0216471 (2019).
64. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30**, 3399–3401 (2014).
65. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
66. Semeraro, R. & Magi, A. PyPore: a python toolbox for nanopore sequencing data handling. *Bioinformatics* **35**, 4445–4447 (2019).
67. Senol Cali, D., Kim, J. S., Ghose, S., Alkan, C. & Mutlu, O. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief. Bioinform.* **20**, 1542–1559 (2019).
68. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
69. McIntyre, A. B. R. et al. Nanopore sequencing in microgravity. *NPJ Microgravity* **2**, 16035 (2016).
70. Teng, H. et al. Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. *Gigascience* **7**, giy037 (2018).
71. Rand, A. C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).
72. Wang, Y. et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).
73. Liu, H. et al. Accurate detection of m^{6}A RNA modifications in native RNA sequences. *Nat. Commun.* **10**, 4079 (2019).
74. Stoiber, M. H. et al. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. Preprint at *bioRxiv* https://doi.org/10.1101/094672 (2016).
75. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
76. Liu, Q. et al. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).
77. Ni, P. et al. DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics* **35**, 4586–4595 (2019).
78. Liu, Q., Georgieva, D. C., Egli, D. & Wang, K. NanoMod: a computational tool to detect DNA modifications using Nanopore long-read sequencing data. *BMC Genomics* **20**, 78 (2019).
79. Yuen, Z. W. et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat. Commun.* **12**, 3438 (2021).
80. Liu, Y. et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 (2021).
81. Fang, G. et al. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* **30**, 1232–1239 (2012).
82. Saletore, Y. et al. The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13**, 175 (2012).
83. Jenjaroenpun, P. et al. Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res.* **49**, e7 (2020).
84. Lorenz, D. A., Sathe, S., Einstein, J. M. & Yeo, G. W. Direct RNA sequencing enables m^{6}A detection in endogenous transcript isoforms at base-specific resolution. *RNA* **26**, 19–28 (2020).
85. Fu, S., Wang, A. & Au, K. F. A comparative evaluation of hybrid error correction methods for error-prone long reads. *Genome Biol.* **20**, 26 (2019).
86. Viehweger, A. et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.* **29**, 1545–1554 (2019).
87. Lima, L. et al. Comparative assessment of long-read error correction software applied to Nanopore RNA-sequencing data. *Brief. Bioinform.* **21**, 1164–1181 (2019).
88. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
89. Salmela, L., Walve, R., Rivals, E. & Ukkonen, E. Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* **33**, 799–806 (2017).
90. Au, K. F., Underwood, J. G., Lee, L. & Wong, W. H. Improving PacBio long read accuracy by short read alignment. *PLoS ONE* **7**, e46679 (2012).
91. Salmela, L. & Rivals, E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**, 3506–3514 (2014).
92. Bao, E. & Lan, L. HALC: high throughput algorithm for long read error correction. *BMC Bioinformatics* **18**, 204 (2017).

93. Wang, J. R., Holt, J., McMillan, L. & Jones, C. D. FMLRC: hybrid long read error correction using an FM-index. *BMC Bioinformatics* **19**, 50 (2018).
94. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
95. Sovic, I. et al. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat. Commun.* **7**, 11307 (2016).
96. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
97. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
98. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
99. Zhou, A., Lin, T. & Xing, J. Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biol.* **20**, 237 (2019).
100. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
101. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
102. Marić, J., Sović, I., Križanović, K., Nagarajan, N. & Šikić, M. Graphmap2—splice-aware RNA-seq mapper for long reads. Preprint at *bioRxiv* https://doi.org/10.1101/720458 (2019).
103. Liu, B. et al. deSALT: fast and accurate long transcriptomic read alignment with de Bruijn graph-based index. *Genome Biol.* **20**, 274 (2019).
104. Begik, O. et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **39**, 1278–1291 (2021).
105. Giordano, F. et al. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci. Rep.* **7**, 3935 (2017).
106. Bertrand, D. et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937–944 (2019).
107. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
108. de Lannoy, C., de Ridder, D. & Risse, J. The long reads ahead: de novo genome assembly using the MinION. *F1000Res* **6**, 1083 (2017).
109. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **12**, 733–735 (2015).
110. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
111. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
112. Cretu Stancu, M. et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **8**, 1326 (2017).
113. Tham, C. Y. et al. NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol.* **21**, 56 (2020).
114. Bowden, R. et al. Sequencing of human genomes with nanopore technology. *Nat. Commun.* **10**, 1869 (2019).
115. Chaisson, M. J. P. et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **10**, 1784 (2019).
116. Edge, P. & Bansal, V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* **10**, 4660 (2019).
117. Schrinner, S. D. et al. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biol.* **21**, 252 (2020).
118. Ewing, A. D. et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol. Cell* **80**, 915–928 (2020).
119. Bolognini, D., Magi, A., Benes, V., Korbel, J. O. & Rausch, T. TRiCoLOR: tandem repeat profiling using whole-genome long-read sequencing data. *Gigascience* **9**, giaa101 (2020).
120. Byrne, A. et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* **8**, 16027 (2017).
121. Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D. & Ragoussis, J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci. Rep.* **6**, 31602 (2016).
122. Volden, R. et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc. Natl Acad. Sci. USA* **115**, 9726–9731 (2018).
123. Tang, A. D. et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat. Commun.* **11**, 1438 (2020).
124. Kuosmanen, A., Sobih, A., Rizzi, R., Mäkinen, V. & Tomescu, A. I. On using longer RNA-seq reads to improve transcript prediction accuracy. In *Proc. 9th International Joint Conference on Biomedical Engineering Systems and Technologies* 272–277 (BIOSTEC, 2016).
125. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
126. Wyman, D. et al. A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. Preprint at *bioRxiv* https://doi.org/10.1101/672931 (2020).
127. Au, K. F. et al. Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl Acad. Sci. USA* **110**, E4821–E4830 (2013).
128. Fu, S. et al. IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. *Bioinformatics* **34**, 2168–2176 (2018).
129. de la Rubia, I. et al. Reference-free reconstruction and quantification of transcriptomes from Nanopore long-read sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2020.02.08.939942 (2021).
130. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
131. Soneson, C. et al. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat. Commun.* **10**, 3359 (2019).
132. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).
133. Jain, M. et al. Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* **36**, 321–323 (2018).
134. Miga, K. H. et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79–84 (2020).
135. Nurk, S. et al. The complete sequence of a human genome. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.26.445798 (2021).
136. Tyson, J. R. et al. MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Res.* **28**, 266–274 (2018).
137. Salazar, A. N. et al. Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.* **17**, fox074 (2017).
138. Michael, T. P. et al. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* **9**, 541 (2018).
139. Miller, D. E., Staber, C., Zeitlinger, J. & Hawley, R. S. Highly contiguous genome assemblies of 15 *Drosophila* species generated using nanopore sequencing. *G3* **8**, 3131–3141 (2018).
140. Kapustova, V. et al. The dark matter of large cereal genomes: long tandem repeats. *Int. J. Mol. Sci.* **20**, 2483 (2019).
141. Diaz-Viraque, F. et al. Nanopore sequencing significantly improves genome assembly of the protozoan parasite *Trypanosoma cruzi*. *Genome Biol. Evol.* **11**, 1952–1957 (2019).
142. Datema, E. et al. The megabase-sized fungal genome of *Rhizoctonia solani* assembled from nanopore reads only. Preprint at *bioRxiv* https://doi.org/10.1101/084772 (2016).
143. Austin, C. M. et al. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *Gigascience* **6**, 1–6 (2017).
144. Tan, M. H. et al. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience* **7**, 1–6 (2018).
145. Singh, K. S. et al. De novo genome assembly of the Meadow Brown Butterfly, *Maniola jurtina*. *G3* **10**, 1477–1484 (2020).
146. Lind, A. L. et al. Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat. Ecol. Evol.* **3**, 1241–1252 (2019).
147. Dhar, R. et al. De novo assembly of the Indian blue peacock (*Pavo cristatus*) genome using Oxford Nanopore technology and Illumina sequencing. *Gigascience* **8**, giz038 (2019).
148. Armstrong, E. E. et al. Long live the king: chromosome-level assembly of the lion (*Panthera leo*) using linked-read, Hi-C, and long-read data. *BMC Biol.* **18**, 3 (2020).
149. Kono, N. et al. The bagworm genome reveals a unique fibroin gene that provides high tensile strength. *Commun. Biol.* **2**, 148 (2019).
150. Wongsurawat, T. et al. Rapid sequencing of multiple RNA viruses in their native form. *Front. Microbiol.* **10**, 260 (2019).
151. Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
152. Kim, D. et al. The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921 (2020).
153. Moore, S. C. et al. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. Preprint at *medRxiv* https://doi.org/10.1101/2020.03.05.20032011 (2020).
154. Taiaroa, G. et al. Direct RNA sequencing and early evolution of SARS-CoV-2. Preprint at *bioRxiv* https://doi.org/10.1101/2020.03.05.976167 (2020).
155. Wang, M. et al. Nanopore targeted sequencing for the accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses. *Small* **16**, e2002169 (2020).

156. Bayega, A. et al. De novo assembly of the olive fruit fly (*Bactrocera oleae*) genome with linked-reads and long-read technologies minimizes gaps and provides exceptional Y chromosome assembly. *BMC Genomics* **21**, 259 (2020).
157. Kadobianskyi, M., Schulze, L., Schuelke, M. & Judkewitz, B. Hybrid genome assembly and annotation of *Danionella translucida*. *Sci. Data* **6**, 156 (2019).
158. Bai, C. M. et al. Chromosomal-level assembly of the blood clam, *Scapharca* (*Anadara*) *broughtonii*, using long sequence reads and Hi-C. *Gigascience* **8**, giz067 (2019).
159. Belser, C. et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* **4**, 879–887 (2018).
160. Marrano, A. et al. High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. *Gigascience* **9**, giaa050 (2020).
161. Ning, D. L. et al. Chromosomal-level assembly of *Juglans sigillata* genome using Nanopore, BioNano, and Hi-C analysis. *Gigascience* **9**, giaa006 (2020).
162. Kwan, H. H. et al. The genome of the Steller Sea Lion (*Eumetopias jubatus*). *Genes* **10**, 486 (2019).
163. Scott, A. D. et al. The giant sequoia genome and proliferation of disease resistance genes. Preprint at *bioRxiv* https://doi.org/10.1101/2020.03.17.995944 (2020).
164. Shafin, K. et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat. Biotechnol.* **38**, 1044–1053 (2020).
165. De Coster, W. et al. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res.* **29**, 1178–1187 (2019).
166. Singh, M. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **10**, 3120 (2019).
167. Roach, N. P. et al. The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome Res.* **30**, 299–312 (2020).
168. Parker, M. T. et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m^6A modification. *eLife* **9**, e49658 (2020).
169. Jiang, F. et al. Long-read direct RNA sequencing by 5′-cap capturing reveals the impact of Piwi on the widespread exonization of transposable elements in locusts. *RNA Biol.* **16**, 950–959 (2019).
170. Zhang, J. et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat. Biotechnol.* **39**, 836–845 (2021).
171. Xin, R. et al. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat. Commun.* **12**, 266 (2021).
172. Laszlo, A. H. et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl Acad. Sci. USA* **110**, 18904–18909 (2013).
173. Schreiber, J. et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc. Natl Acad. Sci. USA* **110**, 18910–18915 (2013).
174. McIntyre, A. B. R. et al. Single-molecule sequencing detection of N^6-methyladenine in microbial reference materials. *Nat. Commun.* **10**, 579 (2019).
175. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
176. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
177. Georgieva, D., Liu, Q., Wang, K. & Egli, D. Detection of base analogs incorporated during DNA replication by nanopore sequencing. *Nucleic Acids Res.* **48**, e88 (2020).
178. Hennion, M. et al. Mapping DNA replication with nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/426858 (2018).
179. Muller, C. A. et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat. Methods* **16**, 429–436 (2019).
180. Ulahannan, N. et al. Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. Preprint at *bioRxiv* https://doi.org/10.1101/833590 (2019).
181. Altemose, N. et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome-wide. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.06.451383 (2021).
182. Weng, Z. et al. Long-range single-molecule mapping of chromatin modification in eukaryotes. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.08.451578 (2021).
183. Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R. & Akeson, M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE* **14**, e0216709 (2019).
184. Aw, J. G. A. et al. Determination of isoform-specific RNA structure with nanopore long reads. *Nat. Biotechnol.* **39**, 336–346 (2020).
185. Stephenson, W. et al. Direct detection of RNA modifications and structure using single molecule nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.31.126763 (2020).
186. Maier, K. C., Gressel, S., Cramer, P. & Schwalb, B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res.* **30**, 1332–1344 (2020).
187. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* **77**, 985–998 (2020).
188. Minervini, C. F. et al. *TP53* gene mutation analysis in chronic lymphocytic leukemia by nanopore MinION sequencing. *Diagn. Pathol.* **11**, 96 (2016).
189. Minervini, C. F. et al. Mutational analysis in *BCR-ABL1* positive leukemia by deep sequencing based on nanopore MinION technology. *Exp. Mol. Pathol.* **103**, 33–37 (2017).
190. Orsini, P. et al. Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia. *Sci. Rep.* **8**, 11798 (2018).
191. Cumbo, C. et al. Genomic *BCR-ABL1* breakpoint characterization by a multi-strategy approach for "personalized monitoring" of residual disease in chronic myeloid leukemia patients. *Oncotarget* **9**, 10978–10986 (2018).
192. Au, C. H. et al. Rapid detection of chromosomal translocation and precise breakpoint characterization in acute myeloid leukemia by nanopore long-read sequencing. *Cancer Genet.* **239**, 22–25 (2019).
193. Euskirchen, P. et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathol.* **134**, 691–703 (2017).
194. Pradhan, B. et al. Detection of subclonal L1 transductions in colorectal cancer by long-distance inverse-PCR and Nanopore sequencing. *Sci. Rep.* **7**, 14521 (2017).
195. Norris, A. L., Workman, R. E., Fan, Y., Eshleman, J. R. & Timp, W. Nanopore sequencing detects structural variants in cancer. *Cancer Biol. Ther.* **17**, 246–253 (2016).
196. Suzuki, A. et al. Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer. *DNA Res.* **24**, 585–596 (2017).
197. Gabrieli, T. et al. Selective nanopore sequencing of human *BRCA1* by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.* **46**, e87 (2018).
198. Jeck, W. R. et al. A nanopore sequencing-based assay for rapid detection of gene fusions. *J. Mol. Diagn.* **21**, 58–69 (2019).
199. Moon, J. et al. Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing: a pilot study. *Int. J. Med. Microbiol.* **309**, 151338 (2019).
200. Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **37**, 783–792 (2019).
201. Cheng, J. et al. Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. *Ann. Clin. Microbiol Antimicrob.* **17**, 43 (2018).
202. Gorrie, C. L. et al. Antimicrobial-resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. *Clin. Infect. Dis.* **67**, 161–170 (2018).
203. Sanderson, N. D. et al. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics* **19**, 714 (2018).
204. Schmidt, K. et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* **72**, 104–114 (2016).
205. Lu, X. et al. Epidemiologic and genomic insights on *mcr-1*-harbouring *Salmonella* from diarrhoeal outpatients in Shanghai, China, 2006–2016. *EBioMedicine* **42**, 133–144 (2019).
206. Hu, Y., Fang, L., Nicholson, C. & Wang, K. Implications of error-prone long-read whole-genome shotgun sequencing on characterizing reference microbiomes. *iScience* **23**, 101223 (2020).
207. De Roeck, A. et al. NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol.* **20**, 239 (2019).
208. Chatron, N. et al. Severe hemophilia A caused by an unbalanced chromosomal rearrangement identified using nanopore sequencing. *J. Thromb. Haemost.* **17**, 1097–1103 (2019).
209. Brandler, W. M. et al. Paternally inherited *cis*-regulatory structural variants are associated with autism. *Science* **360**, 327–331 (2018).
210. Carvalho, C. M. B. et al. Interchromosomal template-switching as a novel molecular mechanism for imprinting perturbations associated with Temple syndrome. *Genome Med.* **11**, 25 (2019).
211. Miao, H. et al. Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. *Hereditas* **155**, 32 (2018).
212. Dutta, U. R. et al. Breakpoint mapping of a novel de novo translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing. *Genomics* **111**, 1108–1114 (2019).
213. Ishiura, H. et al. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat. Genet.* **50**, 581–590 (2018).

214. Zeng, S. et al. Long-read sequencing identified intronic repeat expansions in *SAMD12* from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J. Med. Genet.* **56**, 265–270 (2019).
215. Leija-Salazar, M. et al. Evaluation of the detection of *GBA* missense mutations and other variants using the Oxford Nanopore MinION. *Mol. Genet Genom. Med* **7**, e564 (2019).
216. Lang, K. et al. Full-length HLA class I genotyping with the MinION nanopore sequencer. *Methods Mol. Biol.* **1802**, 155–162 (2018).
217. Liu, C. et al. Accurate typing of human leukocyte antigen class I genes by Oxford Nanopore sequencing. *J. Mol. Diagn.* **20**, 428–435 (2018).
218. Duke, J. L. et al. Resolving MiSeq-generated ambiguities in HLA-DPB1 typing by using the Oxford Nanopore Technology. *J. Mol. Diagn.* **21**, 852–861 (2019).
219. Wei, S. & Williams, Z. Rapid short-read sequencing and aneuploidy detection using MinION nanopore technology. *Genetics* **202**, 37–44 (2016).
220. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
221. Quick, J. et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 114 (2015).
222. Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406–410 (2017).
223. Faria, N. R. et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **361**, 894–899 (2018).
224. de Jesus, J. G. et al. Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019. *Mem. Inst. Oswaldo Cruz* **115**, e190423 (2020).
225. Russell, J. A. et al. Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. *Sci. Rep.* **8**, 5417 (2018).
226. Kafetzopoulou, L. E. et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* **363**, 74–77 (2019).
227. Brynildsrud, O. B. et al. Acquisition of virulence genes by a carrier strain gave rise to the ongoing epidemics of meningococcal disease in West Africa. *Proc. Natl Acad. Sci. USA* **115**, 5510–5515 (2018).
228. Dong, N., Yang, X., Zhang, R., Chan, E. W. & Chen, S. Tracking microevolution events among ST11 carbapenemase-producing hypervirulent *Klebsiella pneumoniae* outbreak strains. *Emerg. Microbes Infect.* **7**, 146 (2018).
229. Rhodes, J. et al. Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. *Emerg. Microbes Infect.* **7**, 43 (2018).
230. Hamner, S. et al. Metagenomic profiling of microbial pathogens in the Little Bighorn River, Montana. *Int. J. Environ. Res. Public Health* **16**, 1097 (2019).
231. Boykin, L. M. et al. Tree Lab: portable genomics for early detection of plant viruses and pests in sub-Saharan Africa. *Genes* **10**, 632 (2019).
232. Zaaijer, S. et al. Rapid re-identification of human samples using portable DNA sequencing. *eLife* **6**, e27798 (2017).
233. Runtuwene, L. R., Tuda, J. S. B., Mongan, A. E. & Suzuki, Y. On-site MinION sequencing. *Adv. Exp. Med. Biol.* **1129**, 143–150 (2019).
234. Sutton, M. A. et al. Radiation tolerance of nanopore sequencing technology for life detection on Mars and Europa. *Sci. Rep.* **9**, 5370 (2019).
235. Castro-Wallace, S. L. et al. Nanopore DNA sequencing and genome assembly on the International Space Station. *Sci. Rep.* **7**, 18022 (2017).
236. Ducluzeau, A., Lekanoff, R. M., Khalsa, N. S., Smith, H. H. & Drown, D. M. Introducing DNA sequencing to the next generation on a research vessel sailing the Bering Sea through a storm. Preprint at *Preprints* https://doi.org/10.20944/preprints201905.0113.v1 (2019).
237. Edwards, A. et al. In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. Preprint at *bioRxiv* https://doi.org/10.1101/073965 (2019).
238. Blanco, M. B. et al. Next-generation technologies applied to age-old challenges in Madagascar. *Conserv. Genet.* **21**, 785–793 (2020).
239. Pushkarev, D., Neff, N. F. & Quake, S. R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* **27**, 847–850 (2009).
240. Merchant, C. A. et al. DNA translocation through graphene nanopores. *Nano Lett.* **10**, 2915–2921 (2010).
241. Schneider, G. F. et al. DNA translocation through graphene nanopores. *Nano Lett.* **10**, 3163–3167 (2010).
242. Garaj, S. et al. Graphene as a subnanometre trans-electrode membrane. *Nature* **467**, 190–193 (2010).
243. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
244. Gershow, M. & Golovchenko, J. A. Recapturing and trapping single molecules with a solid-state nanopore. *Nat. Nanotechnol.* **2**, 775–779 (2007).
245. Seo, J. S. et al. De novo assembly and phasing of a Korean human genome. *Nature* **538**, 243–247 (2016).
246. Boza, V., Peresini, P., Brejova, B. & Vinar, T. DeepNano-blitz: a fast base caller for MinION nanopore sequencers. *Bioinformatics* **36**, 4191–4192 (2020).
247. Stoiber, M. & Brown, J. BasecRAWller: streaming nanopore basecalling directly from raw signal. Preprint at *bioRxiv* https://doi.org/10.1101/133058 (2017).
248. Wang, S., Li, Z., Yu, Y. & Gao, X. WaveNano: a signal-level nanopore base-caller via simultaneous prediction of nucleotide labels and move labels through bi-directional WaveNets. *Quant. Biol.* **6**, 359–368 (2018).
249. Miculinić, N., Ratković, M. & Šikić, M. MinCall-MinION end2end convolutional deep learning basecaller. Preprint at https://arxiv.org/abs/1904.10337 (2019).
250. Zhang, Y. et al. Nanopore basecalling from a perspective of instance segmentation. *BMC Bioinformatics* **21**, 136 (2020).
251. Lv, X., Chen, Z., Lu, Y. & Yang, Y. An end-to-end Oxford Nanopore basecaller using convolution-augmented transformer. *2020 IEEE Intl. Conf. Bioinformatics and Biomedicine (BIBM)* **1**, 337–342 (2020).
252. Huang, N., Nie, F., Ni, P., Luo, F. & Wang, J. SACall: a neural network basecaller for Oxford Nanopore sequencing data based on self-attention mechanism. *IEEE/ACM Trans. Comput. Biol. Bioinform.* https://doi.org/10.1109/TCBB.2020.3039244 (2020).
253. Konishi, H., Yamaguchi, R., Yamaguchi, K., Furukawa, Y. & Imoto, S. Halcyon: an accurate basecaller exploiting an encoder-decoder model with monotonic attention. *Bioinformatics* **37**, 1211–1217 (2021).
254. Xu, Z. et al. Fast-Bonito: a faster basecaller for nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.08.318535 (2020).
255. Fukasawa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M. S. LongQC: a quality control tool for third generation sequencing long read data. *G3* **10**, 1193–1196 (2020).
256. Leger, A. & Leonardi, T. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J. Open Source Softw.* **4**, 1236 (2019).
257. Lanfear, R., Schalamun, M., Kainer, D., Wang, W. & Schwessinger, B. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* **35**, 523–525 (2019).
258. Yin, Z. et al. RabbitQC: high-speed scalable quality control for sequencing data. *Bioinformatics* **37**, 573–574 (2021).
259. Tardaguila, M. et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* **28**, 396–411 (2018).
260. Ferguson, J. M. & Smith, M. A. SquiggleKit: a toolkit for manipulating nanopore signal data. *Bioinformatics* **35**, 5372–5373 (2019).
261. Cheetham, S. W., Kindlova, M. & Ewing, A. D. Methylartist: tools for visualising modified bases from nanopore sequence data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.22.453313 (2021).
262. Su, S. et al. NanoMethViz: an R/Bioconductor package for visualizing long-read methylation data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.18.426757 (2021).
263. De Coster, W., Stovner, E. B. & Strazisar, M. Methplotlib: analysis of modified nucleotides from nanopore sequencing. *Bioinformatics* **36**, 3236–3238 (2020).
264. Pratanwanich, P. N. et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00949-w (2021).
265. Leger, A. et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/843136 (2019).
266. Gao, Y. et al. Quantitative profiling of N^6-methyladenosine at single-base resolution in stem-differentiating xylem of *Populus trichocarpa* using Nanopore direct RNA sequencing. *Genome Biol.* **22**, 22 (2021).
267. Parker, M. T., Barton, G. J. & Simpson, G. G. Yanocomp: robust prediction of m^6A modifications in individual nanopore direct RNA reads. Preprint at *bioRxiv* https://doi.org/10.1101/2021.06.15.448494 (2021).
268. Price, A. M. et al. Direct RNA sequencing reveals m^6A modifications on adenovirus RNA are necessary for efficient splicing. *Nat. Commun.* **11**, 6016 (2020).
269. Miclotte, G. et al. Jabba: hybrid error correction for long sequencing reads. *Algorithms Mol. Biol.* **11**, 10 (2016).
270. Lee, H. et al. Error correction and assembly complexity of single molecule sequencing reads. Preprint at *bioRxiv* https://doi.org/10.1101/006395 (2014).
271. Morisse, P., Lecroq, T. & Lefebvre, A. Hybrid correction of highly noisy long reads using a variable-order de Bruijn graph. *Bioinformatics* **34**, 4213–4222 (2018).
272. Madoui, M. A. et al. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* **16**, 327 (2015).
273. Holley, G. et al. Ratatosk: hybrid error correction of long reads enables accurate variant calling and assembly. *Genome Biol.* **22**, 28 (2021).
274. Koren, S. et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).
275. Hackl, T., Hedrich, R., Schultz, J. & Forster, F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004–3011 (2014).
276. Firtina, C., Bar-Joseph, Z., Alkan, C. & Cicek, A. E. Hercules: a profile HMM-based hybrid error correction algorithm for long reads. *Nucleic Acids Res.* **46**, e125 (2018).

277. Haghshenas, E., Hach, F., Sahinalp, S. C. & Chauve, C. CoLoRMap: correcting long reads by mapping short reads. *Bioinformatics* **32**, i545–i551 (2016).
278. Tischler, G. & Myers, E. W. Non hybrid long read consensus using local de Bruijn graph assembly. Preprint at *bioRxiv* https://doi.org/10.1101/106252 (2017).
279. Xiao, C. L. et al. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* **14**, 1072–1074 (2017).
280. Chin, C. S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
281. Bao, E., Xie, F., Song, C. & Song, D. FLAS: fast and high-throughput algorithm for PacBio long-read self-correction. *Bioinformatics* **35**, 3953–3960 (2019).
282. Nowoshilow, S. et al. The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).
283. Wang, L., Qu, L., Yang, L., Wang, Y. & Zhu, H. NanoReviser: an error-correction tool for nanopore sequencing based on a deep learning algorithm. *Front. Genet.* **11**, 900 (2020).
284. Broseus, L. et al. TALC: transcript-level aware long-read correction. *Bioinformatics* **36**, 5000–5006 (2020).
285. Sahlin, K. & Medvedev, P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat. Commun.* **12**, 2 (2021).
286. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
287. Ren, J. & Chaisson, M. J. P. lra: a long read aligner for sequences and contigs. *PLoS Comput. Biol.* **17**, e1009078 (2021).
288. Jain, C., Rhie, A., Hansen, N. F., Koren, S. & Phillippy, A. M. A long read mapping method for highly repetitive reference sequences. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.01.363887 (2020).
289. Jain, C., Koren, S., Dilthey, A., Phillippy, A. M. & Aluru, S. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* **34**, i748–i756 (2018).
290. Amin, M. R., Skiena, S. & Schatz, M. C. NanoBLASTer: fast alignment and characterization of Oxford Nanopore single molecule sequencing reads. In *2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences* 1–6 (ICCABS, 2016).
291. Yang, W. & Wang, L. Fast and accurate algorithms for mapping and aligning long reads. *J. Comput. Biol.* **28**, 789–803 (2021).
292. Rautiainen, M. & Marschall, T. GraphAligner: rapid and versatile sequence-to-graph alignment. *Genome Biol.* **21**, 253 (2020).
293. Wei, Z. G., Zhang, S. W. & Liu, F. smsMap: mapping single molecule sequencing reads by locating the alignment starting positions. *BMC Bioinformatics* **21**, 341 (2020).
294. Haghshenas, E., Sahinalp, S. C. & Hach, F. lordFAST: sensitive and fast alignment search tool for long noisy read sequencing data. *Bioinformatics* **35**, 20–27 (2019).
295. Chakraborty, A., Morgenstern, B. & Bandyopadhyay, S. S-conLSH: alignment-free gapped mapping of noisy long reads. *BMC Bioinformatics* **22**, 64 (2021).
296. Joshi, D., Mao, S., Kannan, S. & Diggavi, S. QAlign: aligning nanopore reads accurately using current-level modeling. *Bioinformatics* **37**, 625–633 (2021).
297. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**, 405 (2019).
298. Hou, L. & Wang, Y. DEEP-LONG: a fast and accurate aligner for long RNA-seq. Preprint at *Research Square* https://doi.org/10.21203/rs.3.rs-79489/v1 (2020).
299. Sahlin, K. & Mäkinen, V. Accurate spliced alignment of long RNA sequencing reads. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btab540 (2021).
300. Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
301. Vaser, R. & Šikić, M. Time- and memory-efficient genome assembly with Raven. *Nat. Comput. Sci.* **1**, 332–336 (2021).
302. Chin, C. S. & Khalak, A. Human genome assembly in 100 minutes. Preprint at *bioRxiv* https://doi.org/10.1101/705616 (2019).
303. Kamath, G. M., Shomorony, I., Xia, F., Courtade, T. A. & Tse, D. N. HINGE: long-read assembly achieves optimal repeat resolution. *Genome Res.* **27**, 747–756 (2017).
304. Jansen, H. J. et al. Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Sci. Rep.* **7**, 7213 (2017).
305. Chen, Y. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* **12**, 60 (2021).
306. Kolmogorov, M. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **17**, 1103–1110 (2020).
307. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
308. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
309. Huang, N. et al. NeuralPolish: a novel Nanopore polishing method based on alignment matrix construction and orthogonal Bi-GRU Networks. *Bioinformatics* **11**, btab354 (2021).
310. Shafin, K. et al. Haplotype-aware variant calling enables high accuracy in nanopore long-reads using deep neural networks. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.04.433952 (2021).
311. Zimin, A. V. & Salzberg, S. L. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput. Biol.* **16**, e1007981 (2020).
312. Heller, D. & Vingron, M. SVIM: structural variant identification using mapped long reads. *Bioinformatics* **35**, 2907–2915 (2019).
313. Cleal, K. & Baird, D. M. Dysgu: efficient structural variant calling using short or long reads. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.28.446147 (2021).
314. Leung, H. C. et al. SENSV: detecting structural variations with precise breakpoints using low-depth WGS data from a single Oxford Nanopore MinION flowcell. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.20.440583 (2021).
315. Jiang, T. et al. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol.* **21**, 189 (2020).
316. Feng, Z., Clemente, J. C., Wong, B. & Schadt, E. E. Detecting and phasing minor single-nucleotide variants from long-read sequencing data. *Nat. Commun.* **12**, 3032 (2021).
317. Popitsch, N., Preuner, S. & Lion, T. Nanopanel2 calls phased low-frequency variants in Nanopore panel sequencing data. *Bioinformatics* **16**, btab526 (2021).
318. Luo, R. et al. Exploring the limit of using a deep neural network on pileup data for germline variant calling. *Nat. Mach. Intell.* **2**, 220–227 (2020).
319. Edge, P., Bafna, V. & Bansal, V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* **27**, 801–812 (2017).
320. Shaw, J. & Yu, Y. W. Practical probabilistic and graphical formulations of long-read polyploid haplotype phasing. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.06.371799 (2021).
321. Klasberg, S., Schmidt, A. H., Lange, V. & Schofl, G. DR2S: an integrated algorithm providing reference-grade haplotype sequences from heterozygous samples. *BMC Bioinformatics* **22**, 236 (2021).
322. Zhou, W. et al. Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. *Nucleic Acids Res.* **48**, 1146–1163 (2020).
323. Giesselmann, P. et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat. Biotechnol.* **37**, 1478–1481 (2019).
324. Marchet, C. et al. De novo clustering of long reads by gene from transcriptomics data. *Nucleic Acids Res.* **47**, e2 (2019).
325. Sahlin, K. & Medvedev, P. De novo clustering of long-read transcriptome data using a greedy, quality value-based algorithm. *J. Comput. Biol.* **27**, 472–484 (2020).
326. Tian, L. et al. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.10.243543 (2020).
327. Hu, Y. et al. LIQA: long-read isoform quantification and analysis. *Genome Biol.* **22**, 182 (2021).
328. Rautiainen, M. et al. AERON: transcript quantification and gene-fusion detection using long reads. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.27.921338 (2020).
329. Weirather, J. L. et al. Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. *Nucleic Acids Res.* **43**, e116 (2015).
330. Davidson, N. M. et al. JAFFAL: detecting fusion genes with long read transcriptome sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.26.441398 (2021).
331. Liu, Q. et al. LongGF: computational algorithm and software tool for fast and accurate detection of gene fusions by long-read transcriptome sequencing. *BMC Genomics* **21**, 793 (2020).
332. Deonovic, B., Wang, Y., Weirather, J., Wang, X. J. & Au, K. F. IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. *Nucleic Acids Res.* **45**, e32 (2017).
333. Glinos, D. A. et al. Transcriptome variation in human tissues revealed by long-read sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.22.427687 (2021).
334. Calus, S. T., Ijaz, U. Z. & Pinto, A. J. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *Gigascience* **7**, giy140 (2018).

335. Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods* **18**, 165–169 (2021).
336. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
337. Cheetham, S. W. et al. Single-molecule simultaneous profiling of DNA methylation and DNA–protein interactions with Nanopore-DamID. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.09.455753 (2021).
338. Hennion, M. et al. FORK-seq: replication landscape of the *Saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biol.* **21**, 125 (2020).
339. Philpott, M. et al. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00965-w (2021).
340. Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
341. Lebrigand, K., Magnone, V., Barbry, P. & Waldmann, R. High throughput error corrected Nanopore single cell transcriptome sequencing. *Nat. Commun.* **11**, 4025 (2020).
342. Bizuayehu, T. T., Labun, K., Jefimov, K. & Valen, E. Single molecule structure sequencing reveals RNA structural dependencies, breathing and ensembles. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.18.101402 (2020).
343. Drexler, H. L. et al. Revealing nascent RNA processing dynamics with nano-COP. *Nat. Protoc.* **16**, 1343–1375 (2021).

Acknowledgements

K.F.A., Yunhao Wang, Y.Z., A.B. and Yuru Wang are grateful for support from an institutional fund of the Department of Biomedical Informatics, The Ohio State University, and the National Institutes of Health (R01HG008759, R01HG011469 and R01GM136886). The authors apologize to colleagues whose studies were not cited due to length and reference constraints. The authors also apologize that the very latest research published during the publication process of this article was not included. We would like to thank K. Aschheim and G. Riddihough for critical reading and editing of the manuscript.

Author contributions

K.F.A. designed the outline of the article. Yunhao Wang and A.B. collected information and prepared the materials for the 'Technology development' and 'Data analysis' sections. Y.Z. collected information and prepared the materials for the 'Applications of nanopore sequencing' section. K.F.A., Yunhao Wang, Y.Z. and A.B. wrote and revised the main text. Yuru Wang collected the references for the 'Applications of nanopore sequencing' section and prepared Fig. 1.

Competing interests

K.F.A. was invited by ONT to present at the conference London Calling 2020.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01108-x.

Correspondence should be addressed to Kin Fai Au.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature America, Inc. 2021