Review article

Beyond assembly: the increasing flexibility of single-molecule sequencing technology

Paul W. Hook ⓘ & Winston Timp ⓘ ✉

Abstract

The maturation of high-throughput short-read sequencing technology over the past two decades has shaped the way genomes are studied. Recently, single-molecule, long-read sequencing has emerged as an essential tool in deciphering genome structure and function, including filling gaps in the human reference genome, measuring the epigenome and characterizing splicing variants in the transcriptome. With recent technological developments, these single-molecule technologies have moved beyond genome assembly and are being used in a variety of ways, including to selectively sequence specific loci with long reads, measure chromatin state and protein–DNA binding in order to investigate the dynamics of gene regulation, and rapidly determine copy number variation. These increasingly flexible uses of single-molecule technologies highlight a young and fast-moving part of the field that is leading to a more accessible era of nucleic acid sequencing.

Sections

Department of Biomedical Engineering, Molecular Biology and Genetics, and Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA. ✉e-mail: wtimp@jhu.edu

Review article

Introduction

Since the beginning of the Human Genome Project in 1990, there has been a close pairing between technological innovation driving science and science demanding technological innovation. This drive led to next-generation, short-read sequencing methods dominating the field of nucleic acid sequencing (reviewed in ref. [1]). However, short-read sequencing is fundamentally limited in read length (<1000 bp reported[1]) owing to cycle dephasing and the resulting drops in read quality over length[2,3]. By contrast, single-molecule sequencing methods, especially platforms from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), are not subject to this limitation and allow for the sequencing of long reads (>10 kb). Perhaps the most important difference between these platforms is that PacBio performs sequencing-by-synthesis whereas ONT uses a protein nanopore to characterize the molecule through electrolytic current modulation[4]. Though both technologies had initial issues with read accuracy (PacBio continuous long read accuracy 85–89%[5]; ONT R6 accuracy 67%[6]) and yield (PacBio RS II ~500–1000 Mb; ONT R6 yield ~250 Mb), these features have improved substantially over the past eight years. Both technologies can now achieve impressive accuracies — ~98% for ONT and 99% for PacBio[4,7] — and an ONT PromethION device can generate in excess of 100 Gb per flow cell, whereas a PacBio Sequel II HiFi run can generate over 30 Gb[4]. These output levels put the cost per Gb of PacBio (US$65) and ONT (US$17) sequencing closer to that of short-read instruments such as the Illumina NovaSeq 6000 (US$6) (Supplementary Note).

Long reads have already changed the landscape of genomics, expanding our knowledge by exploring areas that were previously unattainable with short reads. Long reads allow for more complete genome assemblies[8], highlighted by their use in the assembly of the first telomere-to-telomere human genome[9]. Many more structural variants and repetitive areas can be probed with long reads because of their ability to map through the variant[10,11], leading to the use of long-read sequencing for surveying structural variants in human populations[12,13]. Single-molecule sequencing even allows for native measurement of DNA methylation[14], including in previously inaccessible regions such as centromeres[15,16]. Aside from DNA, long reads have also been used to explore RNA, providing information about full-length transcript isoforms including allele-specific expression, poly(A) tail length and RNA modifications[17–19].

The increasing accuracy and affordability of single-molecule, long-read sequencing has resulted in the accelerated development of methods that apply it to new problems in biology. Here, we review a selection of emerging methods and applications using commercially available single-molecule platforms. First, we review methods used for targeted sequencing of long reads, which harness the advantages of long-read sequencing without the need for whole-genome sequencing, thereby improving coverage and affordability. Next, we focus on assays for mapping protein–DNA interactions, which in addition to ascertaining information already revealed by short reads also provide previously unknown insights into genome organization. Last, we cover the sequencing of short reads with single-molecule platforms, a suite of methods that seek to increase the accessibility of sequencing and the amount of information that can be gained from a single sequencing run.

Insights without whole-genome sequencing

Costs for whole-genome sequencing have dropped substantially during the past decade, but even with the lower cost there are biological questions for which focused, high-depth sequencing is needed. For example,

somatic variant calling and epigenetic sequencing of heterogeneous samples requires high sequencing depth to enable low-frequency variants or rare epigenetic states to be measured with confidence. Alternatively, when sequencing large sample sets such as complex disease cohorts, cost per sample becomes an important factor. In these scenarios, depth or sample number may be more important than unbiased genome-wide analysis, so targeting specific regions can drive down cost. Specific regions of interest — for example promoters or exons of protein-coding genes — can be selectively targeted for sequencing. Such targeted sequencing methods, including PCR amplicon sequencing and hybridization capture, have been extensively used in concert with short-read sequencing. These same methods have been adapted for long-read sequencing, in addition to the emergence of novel methods taking advantage of the PacBio and ONT platforms.

PCR enrichment

PCR enrichment, also known as amplicon sequencing, allows for targeted sequencing by simply designing primers flanking regions of interest. PCR enrichment is a mature method with low DNA input requirements and low hands-on time, which enables multiplexing of as many as 24,000 amplicons in one reaction with carefully designed commercial primer panels (Ion AmpliSeq assays[20]). Overlapping amplicons can be tiled across regions much longer than the amplicon length, with a recent example targeting genomic regions >40 kb[21]. PCR enrichment can be adapted to long-read sequencing (Fig. 1a) owing in part to the commercial availability of DNA polymerases that can amplify amplicons greater than 10 kb[22,23]. However, as the length of an amplicon increases, PCR becomes less efficient and requires optimization for each new reaction[24]. Amplicons greater than 7 kb and long amplicons with high GC content are difficult to consistently amplify[25]. PCR can also introduce errors (mainly substitutions)[25], which can be an issue when probing rare mutations[26]. Amplifying DNA with PCR erases native DNA modifications, eliminating one of the key advantages of single-molecule platforms (Table 1). Notably, amplicon approaches often require sets of primers to be split into multiple pools owing to possible interactions between primer pairs, thus requiring multiple, optimized PCRs. This makes scaling PCR amplicons to multiple regions difficult. This is especially true for schemes that attempt to tile overlapping amplicons across large regions, as demonstrated in peer-reviewed and preprint studies[21,27]. Despite these caveats, amplicon sequencing has been used with ONT to detect structural variant frequency in genes frequently mutated in pancreatic cancer (*CDKN2A* and *SMAD4*)[28] and with PacBio to identify disease-causing variants in a gene frequently mutated in autosomal-dominant polycystic kidney disease (*PKD1*)[29]. Outside human genetics, as demonstrated in both peer-reviewed[30] and preprint[27] articles, tiled amplicons have been used for low-cost, portable, infectious disease outbreak monitoring with ONT for a host of viruses including Zika[30], Ebola[31] and SARS-CoV-2[27], underscoring the utility of this method (Table 1).

Hybridization capture sequencing

Hybridization capture sequencing uses tagged, antisense oligonucleotide probes against regions of interest. Genomic DNA is denatured using a combination of heat and chemical methods, probes are hybridized against it, probe-bound DNA is captured and unbound DNA is washed away[32] (Fig. 1b). This method can be more easily scaled than PCR amplicons and often only requires one reaction, though probes are expensive and the resulting on-target rate tends to be lower (Table 1). Hybridization capture probes can also be used to enrich across large, contiguous
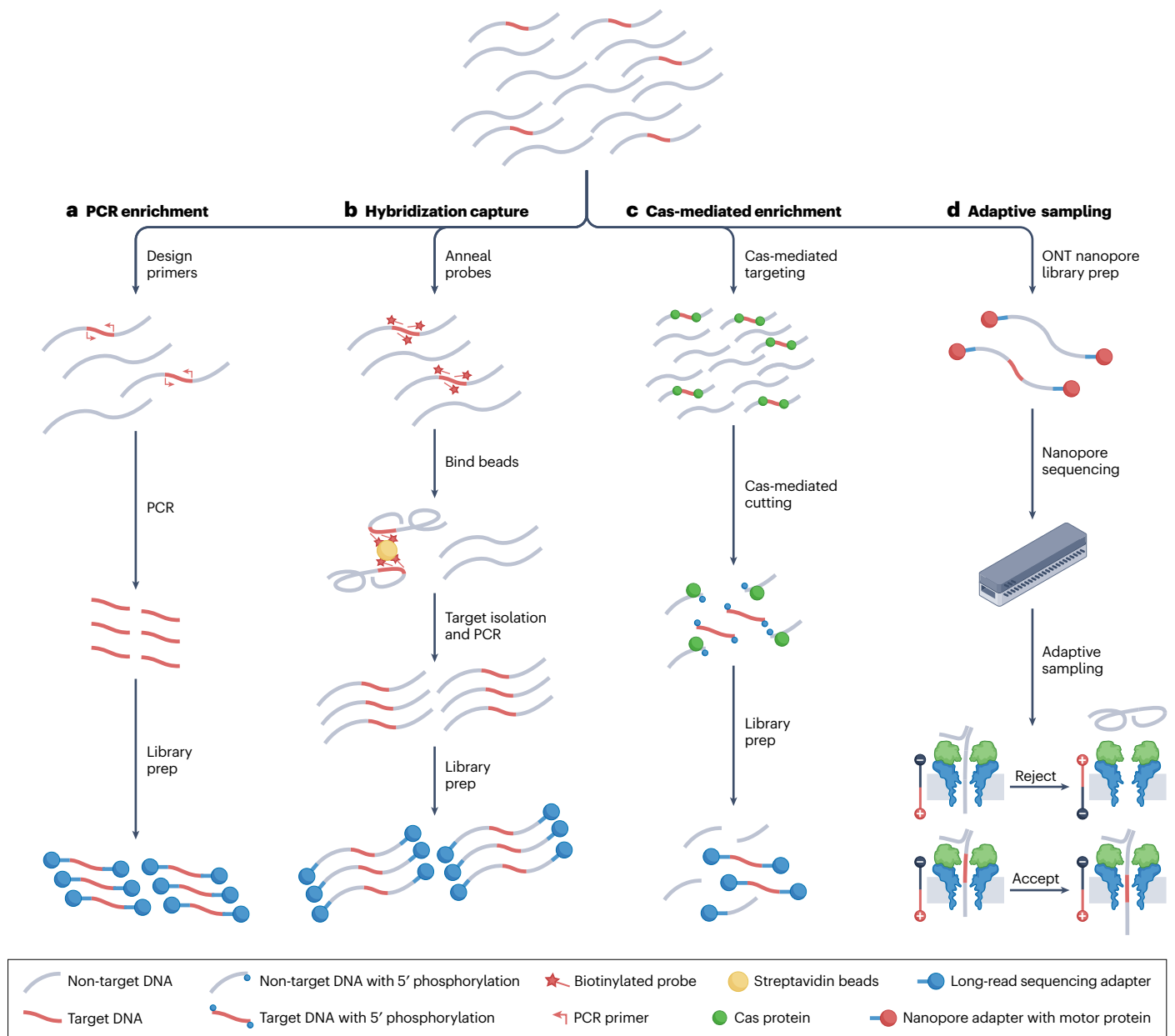
Fig. 1 | Long-read targeted sequencing methods. Long-read targeted enrichment methods fall within broad categories including PCR enrichment, hybridization capture, Cas-mediated enrichment and adaptive sampling. **a**, PCR enrichment uses specific primers to amplify regions of interest before library preparation. **b**, Hybridization capture uses biotinylated antisense probes designed against regions of interest to isolate DNA fragments containing the targets. PCR and hybridization capture enrichment methods are both commonly used with short-read sequencing and have been adapted to long-read sequencing. **c**, Cas-mediated enrichment uses Cas ribonuclear complexes (most commonly Cas9) to cut on either side of regions of interest. Cut fragments are selectively sequenced owing to preferential adapter ligation to the freshly cut ends[55]. Targeted fragments can be further enriched through depletion of off-target fragments[56–58]. **d**, Enrichment using adaptive sampling is a nanopore sequencing method in which regions of interest are selectively sequenced by controlling the voltage at individual pores to eject unwanted fragments. ONT, Oxford Nanopore Technologies.

target regions (for example, ~750,000 bp[33]) by tiling probes across the region in one reaction. Multiple separate locations are easily targeted — exemplified by a study targeting 4800 genes simultaneously with nanopore sequencing (Table 1), even though reads were only ~1,000 bp[34]. Though long-read hybridization capture methods have been applied successfully even in human cohorts to resolve complex structural variants leading to disease[35–38], they have key limitations (Table 1). The lengths of sequenced fragments are typically shorter than those in the original library, suggesting bias towards shorter fragments[38]. This observation has been consistent across long-read hybridization capture

Table 1 | Summary of long-read DNA enrichment methods

Method	Typical read length	Typical coverage	Example of percent reads on-target	Example of number of targets	Platform	Advantages	Disadvantages
PCR	7 kb	50–1,000×	>96%[141]	19 amplicons[31]	ONT, PacBio	Simple design High enrichment Low hands-on time Easy to multiplex samples Easy to multiplex targets Sensitive Low input	Erases native DNA modifications Limited fragment length Lengthy optimization may be required Introduction of PCR errors Multiple reactions often needed
Hybridization capture	5 kb	200–1,000×	66.3%[42]	4,800 genes[34]	ONT, PacBio	Simple design High enrichment Easy to multiplex targets Easy to multiplex samples High scalability Single reaction needed	Erases native DNA modifications Limited fragment length Laborious protocols with high hands-on time Includes multiple PCR steps
Cas-mediated enrichment	Up to 100 kb	50–1,000×	4.6%[55]	10 genes[55]	ONT, PacBio	Preserves DNA modifications Long read lengths No PCR	High DNA input Less ability to multiplex
Adaptive sampling	10–20 kb	15–40×	~5%[68]	717 genes[68]	ONT	Simple design Preserves DNA modifications Long read lengths No PCR No additional molecular biology steps besides library prep	Low enrichment Limited to ONT nanopore sequencing Multiple sequencing runs to obtain maximum output Computationally intensive

ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences.

experiments[37,39–42] and is attributed to the hybridization capture step[41]. We and others have found large fragments more difficult to capture, with the most efficient capture size found to be about 5 kb[43–45]. As with PCR amplicons, amplification (pre-capture or post-capture) can lead to errors in reads; for example, errors in AT-rich regions led to gaps in assembled haplotypes of a complex genomic region containing the natural killer-cell immunoglobulin-like receptor (KIR) gene family[46]. Hybridization capture is often a lengthy protocol (often >3 days[42]) independent of the long-read platform used — though automation and high throughput (96 samples) are possible with liquid-handling robotics. Despite these limitations, hybridization capture can produce deep on-target coverage with one study reporting 1099-fold enrichment from a single run on an ONT MinION device[37].

Cas-mediated enrichment

Though powerful, amplicon and hybridization capture have key limitations in read length and maintenance of modification state: to fully capitalize on the potential of single-molecule targeted sequencing, methods need to be designed from the ground up with this in mind. A bacterial defence system, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins, though primarily used for genome editing[47], can be adapted to enrich long fragments (Table 1). In Cas-mediated enrichment, the CRISPR–Cas system is used to induce double-stranded breaks flanking the regions of interest, which produces long fragments with ends amenable to downstream applications. Initially used to clone large fragments[48,49], Cas9-assisted targeting of chromosome segments (CATCH) was adapted so that the

cut fragments were instead gel-isolated by size and sequenced on an ONT MinION flow cell, achieving ~25–70× mean coverage tiled across a 200-kb region encompassing the hereditary cancer gene *BRCA1* (ref. 50). Unfortunately, so little DNA was recovered after gel isolation that amplification was required, removing native DNA modifications and resulting in read lengths less than 5 kb[50].

Subsequent methods have instead used preferential ligation at freshly cut sites flanking the regions of interest to remove the size selection step and have been used with both PacBio[51,52] and ONT sequencing[53–58]. Typically in these approaches, Cas cleavage occurs before library preparation and the first step is to passivate existing DNA ends by dephosphorylating them, which prevents random ligation. DNA is then cut by a Cas protein–guide RNA complex, either on one side or flanking a region of interest, to create 5′ phosphorylated ends. Sequencing adapters are then ligated to the freshly cut and phosphorylated sites to enable selective sequencing of fragments containing the area of interest (Fig. 1c). Exemplifying this strategy is nanopore Cas9-targeted sequencing (nCATS), which achieved up to 1,000× coverage at loci on an ONT MinION sequencer[55]. However, without multiplexing, only a fraction of the flow cell capacity is used in this method because of the low molarity of resulting library molecules[55] (Table 1). Furthermore, this method seems to work best when two cut sites are generated. Additionally, obtaining read lengths greater than 50 kb was difficult, which may be attributed to the isolation of fragmented DNA during purification[55]. This affects the ability to obtain single reads that span larger regions.

Additional methods have been developed in an attempt to improve upon these caveats. For example, the affinity-based Cas9-mediated

enrichment method (ACME) removes non-target fragments (increasing the molarity of library molecules) via bead-based pulldown of a His-tagged Cas9, which remains bound to non-target fragments after cutting[56]. Data presented in a preprint article demonstrated that ACME excelled in enriching for single reads spanning the entire length of large target regions (~100 kb)[56]. Cas-mediated enrichment has also been demonstrated on a completed PacBio sequencing library. As presented in a preprint article from 2017, a special capture adapter can be ligated to cut sites after Cas-mediated digestion[51], allowing for a bead-based pull-down enrichment approach similar to ACME. This optimized PacBio approach was able to achieve 9% on-target reads, greater than reported with ACME (<1%)[51,56]. Alternatively, exonucleases can be used to digest off-target fragments, as in Cas9-based background elimination (CaBagE)[57], Negative Enrichment[58] and PacBio No-Amp[52]. These exonuclease-based methods can produce high coverage at target loci (~400× for small targets) with a high percentage of reads spanning the entire target region[57]. Furthermore, as shown in both published and preprint work, the size of target regions can be increased by tiling guide RNAs across a region[59,60], similar to tiling methods used with PCR amplicons or hybridization capture. By using a pool of in vitro transcribed guide RNAs tiled across the region, a recent preprint study demonstrated the ability to enrich reads across a region as large as 9 Mb[60].

Adaptive sampling

All the methods mentioned above include additional molecular biology steps involving targeted probes, primers or guide RNAs, which can add time and cost. An enrichment approach that does not include additional manipulations makes single-molecule targeted sequencing more accessible. Nanopore sequencing offers a unique opportunity in this regard — as the molecule is sequenced, a decision can be made to eject the molecule by flipping the voltage if the data do not match a database of targets, a process called adaptive sampling (Fig. 1d). Initially, adaptive sampling was implemented by matching the real-time electrical signal to a reference genome using dynamic time warping with the 'Read Until' approach[61], but was limited to small reference genomes. As a result, improved algorithms for mapping electrical signal were developed[62–67], exemplified by UNCALLED, which demonstrated real-time enrichment of 148 human cancer genes with an average coverage of ~30× (5.5-fold enrichment over non-enriched) using an ONT MinION flow cell[62] (Fig. 1d). Alternatively, improvements to the speed of the basecaller enabled the development of tools that align basecalled reads against a reference to decide whether or not a molecule should be sequenced[68–70]. These tools are exemplified by readfish, which demonstrated enrichment of the genomic sequence of ~700 genes associated with human cancer (~30× mean coverage)[68]. A version of these sequence-based methods has been directly incorporated into the ONT sequencing software (MinKNOW), making it easy for end-users to employ.

Compared to other methods, adaptive sampling can target large regions of interest without additional expense or optimization of primers, probes or guide RNAs. Even entire human chromosomes can be targeted[68], which can be ideal for biological questions such as exploring putative X chromosome-linked disorders. However, in order to achieve enrichment, sequenced fragments must be a sufficient length (>5 kb)[71]; the longer the 'rejected' molecule, the more time is saved by not sequencing it and hence the higher the enrichment of 'accepted' sequences. Best results are typically achieved for fragment sizes >10 kb[62,68,72]. Samples with damaged DNA (for example, formalin-fixed, paraffin-embedded tissue) typically have DNA lengths below this threshold, which may hinder their use with adaptive sampling. Finally, targeting either too low a percentage (<1%) or too high a percentage (>10%) of the genome will also lead to less enrichment: if too much time or not enough time is spent rejecting molecules, the resulting on-target sequence yield will not be sufficient.

Though easy to use, adaptive sampling methods result in lower coverage and a lower percentage of on-target reads than other enrichment methods (Table 1). Encouragingly, data presented in a recent preprint article demonstrated that readfish multiplexed sequencing on the ONT PromethION flow cell yielded 25–50× coverage for three human samples (5–6× enrichment over theoretical whole-genome sequencing), further reducing cost and indicating that higher depth is achievable[72]. Currently, adaptive sequencing requires relatively substantial computational resources, including access to graphical processing units (NVIDIA 2060 series or better with CUDA capability) or powerful central processing units to achieve the analysis speed needed for enrichment. Finally, pores become inactive more quickly during adaptive sampling than during standard nanopore sequencing runs, possibly owing to DNA blockages[62]. Maximum output can be achieved by performing a nuclease flush of the flow cells to remove blockages and a reload of the flow cell with fresh library[62,68,72], but this increases the amount of DNA, reagents and hands-on time required for these experiments.

Additional methods

There are other approaches for long-read enrichment that do not fit into the above categories. For example, Xdrop partitions long DNA molecules into droplets with locus-specific primers, followed by droplet digital PCR. Droplets containing the loci of interest are isolated with flow sorting, and DNA is amplified[73]. This amplified DNA can then be sequenced with short-read or long-read platforms. This method requires a specialized microfluidic apparatus whereas the methods described above need only standard molecular biology tools.
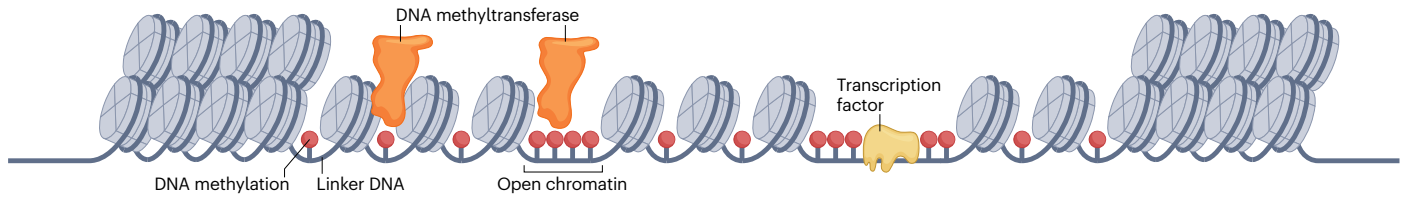
Mapping protein–DNA interactions

For decades, researchers have tried to understand not just the sequence of DNA, but how DNA is organized within the nucleus and how that organization affects cellular function, development, gene regulation and disease (reviewed in ref. 74). State-of-the-art genomics methods including microarrays and next-generation sequencing have been leveraged to study chromatin state and protein–DNA binding (reviewed in[75–77]), even down to the single-cell level (reviewed in ref. 78). Most of these assays rely on PCR enrichment for states of interest (such as open chromatin or bound protein), requiring input controls to correct for PCR bias and thereby making quantification difficult. These methods also typically fragment the DNA to small sizes to provide resolution, making it impossible to study the coordination of chromatin states at adjacent loci on the same single molecule of DNA. Short reads also make it difficult to assign reads to haplotypes given the infrequency of variants on short fragments. As emphasized above, PCR erases native DNA modifications, making additional steps necessary in order to measure methylation and protein–DNA interactions or chromatin state simultaneously[79–81].
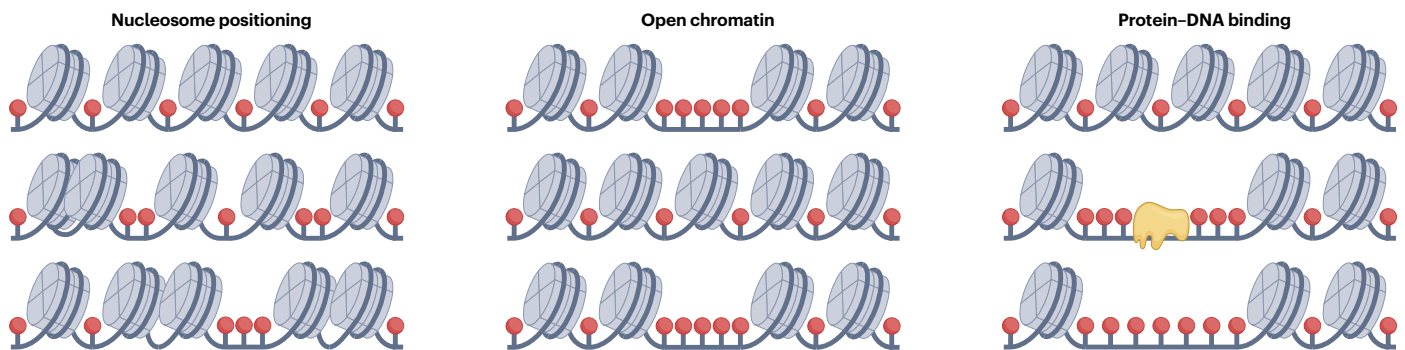
Specific short-read methods using methyltransferase footprinting have set the stage for long-read approaches to explore protein–DNA binding. Emerging from the observation that methyltransferase enzymes preferentially label accessible DNA[82], methyltransferase footprinting assays were developed to measure nucleosome positioning and protein–DNA interactions[83–86]. Such assays can even determine protein binding through the protection from labelling; though

a Methyltransferase footprinting

b Heterogeneity
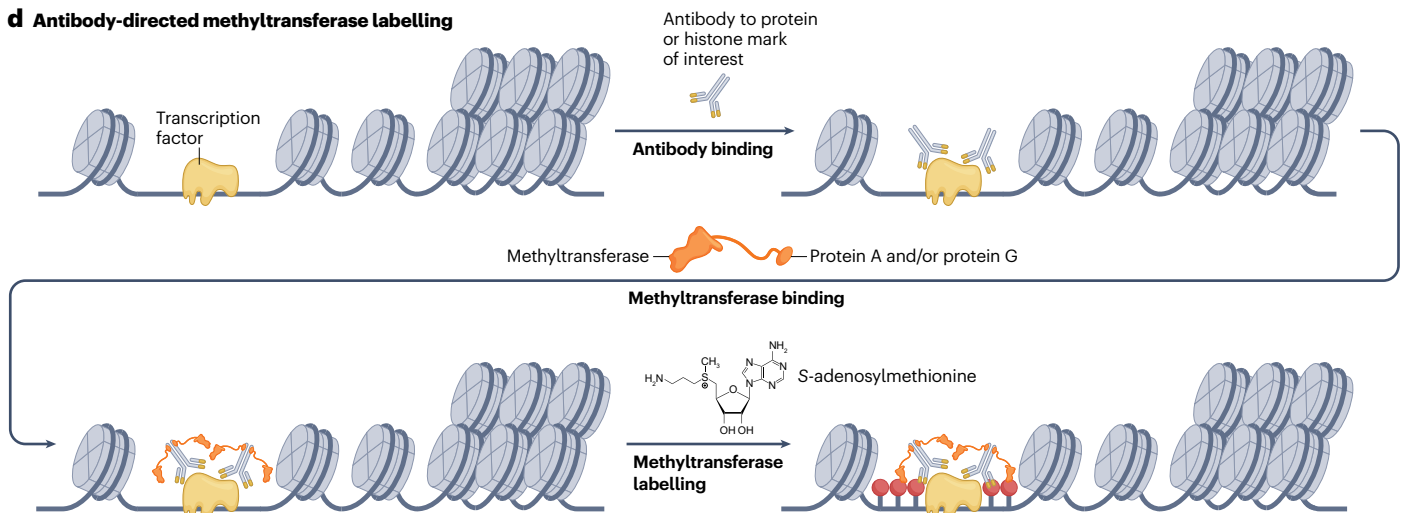
Nucleosome positioning   Open chromatin   Protein–DNA binding

c Coordination

Regulatory element 1   Regulatory element 2   Binding site 1   Binding site 2
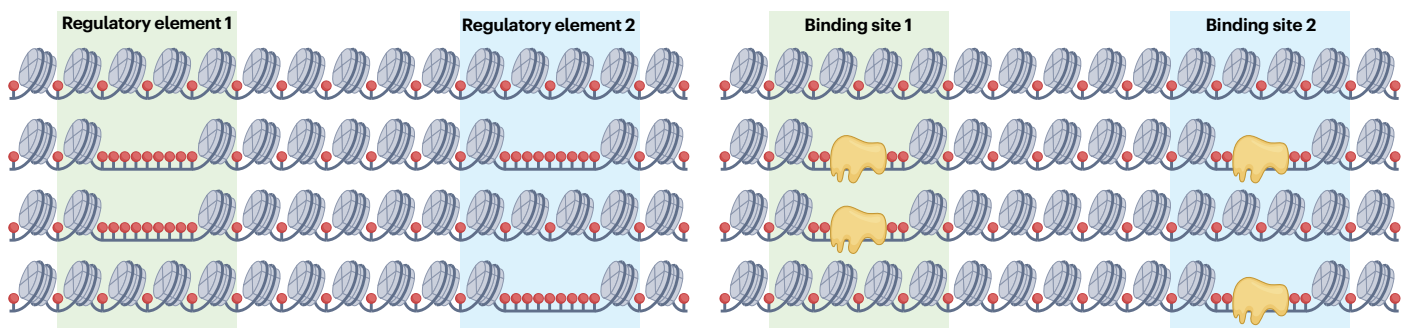
d Antibody-directed methyltransferase labelling

Review article

the identity of the protein is not known, it can be inferred from the size of the protected areas (nucleosomes) or motifs in the protected areas[87]. Chemical bisulfite conversion of unmethylated bases followed by next-generation sequencing allowed these footprinting assays to be applied to panels of promoters[88], to genome-wide footprinting[89], and down to single molecules with short reads[90]. These methods have now been combined with single-molecule platforms to begin to probe unknown aspects of gene regulation (Fig. 2).

Measuring chromatin accessibility with methyltransferase footprinting

Three methods have been developed that combine 5-methylcytosine (5mC) labelling with ONT sequencing to assay nucleosome positioning and open chromatin (Table 2). Two methods focused on yeast: one measured nucleosome positioning, with methyltransferase treatment followed by single-molecule long-read sequencing (MeSMLR-seq) using the GpC methyltransferase M.CviPI[91]; the other measured nucleosome occupancy via DNA methylation and high-throughput sequencing (ODM-seq) using both M.CviPI and the CpG methyltransferase M.SssI[92]. These methods were shown to correlate well with micrococcal nuclease (MNase) digestion sequencing (MNase-seq), a classic method for measuring nucleosome positioning. Using MeSMLR-seq data, over 300 inferred nucleosomes were phased on a single read and it was found that the number of molecules with open chromatin at a given promoter correlates with the expression of its corresponding gene[91]. ODM-seq estimated the number of nucleosomes across the entire genome in a yeast cell and quantified protein binding in nucleosome-free regions[92]. Methyltransferase footprinting has also been applied to human samples. Nanopore sequencing of nucleosome occupancy and methylome (nanoNOMe), adapted from NOMe-seq[89], used M.CviPI to simultaneously call accessible chromatin (GC 5mC) and native CpG methylation, allowing for footprinting of proteins bound to DNA in bulk and on single reads[93]. NanoNOMe made use of the advantages of long reads by exploring chromatin state in repetitive elements and phasing reads to measure allele-specific chromatin accessibility and CpG methylation[93]. In particular, nanoNOMe was able to quantitatively examine protein binding at known motifs, such as CTCF sites, by examining the inferred footprint at these locations. Unsurprisingly, this revealed that traditional chromatin immunoprecipitation followed by sequencing (ChIP–seq) methods are semi-quantitative and that a ChIP–seq peak can represent a large range of fractional binding states. Later work combining nanoNOMe with Cas-mediated enrichment for higher depth found that different CTCF-binding sites have very different percentages of reads (5–70%) supporting CTCF binding[94].

The absence of recognition motifs for these 5mC methyltransferases can limit their ability to label some parts of the genome, such as AT-rich regions. Thus, other methods have leveraged N^6-methyladenine (m6dA, also known as 6mA) methyltransferases (Table 2) for labelling, as m6dA is either absent from or present only at low levels in the genomes of eukaryotes[95]. The single-molecule long-read accessible

chromatin mapping sequencing assay (SMAC-seq) uses a combination of methyltransferases (including M.CviPI, M.SssI and EcoGII (m6dA on all adenines)) to achieve high-resolution (<5 bp) mapping in order to study chromatin states and the coordination of regulatory elements on single molecules using ONT nanopore sequencing[96] (Fig. 2). Fiber-seq used the Hia5 methyltransferase (m6dA on all adenines) with readout from PacBio sequencing[97]. Both methods were developed using model organisms with small genomes: SMAC-seq was developed using yeast and Fiber-seq using the *Drosophila melanogaster* S2 cell line. Both showed high correlation with existing open-chromatin data and the ability to study the coordination of chromatin state between adjacent regulatory sites (Fig. 2). More recently, a preprint article has described the use of Fiber-seq in human samples, leveraging improvements in single-molecule yield to profile the chromatin state of telomeres[98].

Methyltransferase labelling has been further extended by combining it with other methods that can reveal protein–DNA interactions (Table 2). The single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA) combines EcoGII-mediated m6dA labelling with MNase digestion[99], which targets reads to accessible regions. Footprinting information can be obtained both from the molecule ends and from m6dA labelling. A recent preprint described tagmentation-assisted SAMOSA (SAMOSA-Tag)[100] in which the MNase is replaced with Tn5 transposase, commonly used in the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and cleavage under targets and tagmentation (CUT&Tag)[101]. Importantly, the authors demonstrate identification of m6dA labelling and native 5mC CpG modifications, showing that SAMOSA-Tag can assay protein–DNA interactions, epigenetic modifications and primary DNA sequence simultaneously with PacBio sequencing.

Directly mapping protein–DNA interactions

In an extension of footprinting, m6dA labelling has been used within the framework of cleavage under targets & release using nuclease (CUT&RUN) and CUT&Tag methods[101,102] to directly measure interactions between specific proteins and DNA (Table 2). In these approaches, a protein of interest is bound by specific antibodies (Fig. 2d). These antibodies are bound by bacterial proteins that bind tightly to IgG (protein A, protein G or both)[103] fused to methyltransferases, thereby concentrating methyltransferase activity — and m6dA labelling with *S*-adenosylmethionine — around protein binding sites (Fig. 2d). This approach has been implemented for Hia5 (ref.[104]) and EcoGII[105] and can map protein–DNA binding with a resolution of 100–200 bp. Directed methylation with long-read sequencing (DiMeLo-seq) uses Hia5 and is the most extensively tested and optimized approach: it has been used to measure protein–DNA interactions across repetitive regions of the genome, study the coordination and heterogeneity of adjacent binding sites and phase read to study allele-specific protein–DNA binding[104].

Although single-molecule approaches for measuring protein–DNA binding unlock the ability to explore previously intractable biological questions, the way the interactions are measured is fundamentally

Table 2 | Summary of long-read footprinting assays

Method	Enzymes used	Theoretical resolution in humans[a]	Sequencing platform	Demonstrated uses
Methyltransferase treatment followed by single-molecule long-read sequencing (MeSMLR-seq)[91]	M.CviPI (5mC GC)	Medium (>20 bp)	ONT	Nucleosome positioning, chromatin accessibility, chromatin coordination on single molecules
Occupancy measurement via DNA methylation and high-throughput sequencing (ODM-seq)[92]	M.CviPI (5mC GC), M.SssI (5mC CG)	Medium (>20 bp)	ONT	Nucleosome positioning, chromatin accessibility, transcription factor footprinting
Nanopore sequencing of nucleosome occupancy and methylome (nanoNOMe)[93]	M.CviPI (5mC GC)	Medium (>20 bp)	ONT	Nucleosome positioning, chromatin accessibility, transcription factor footprinting
Single-molecule long-read accessible chromatin mapping sequencing assay (SMAC-seq)[96]	M.CviPI (5mC GC), M.SssI (5mC CG), EcoGII (m6dA on all A)	High (<5 bp)	ONT	Nucleosome positioning, chromatin accessibility, transcription factor footprinting, chromatin coordination on single molecules
Fiber-seq[97]	Hia5 (m6dA on all A)	High (<5 bp)	PacBio	Nucleosome positioning, chromatin accessibility, transcription factor footprinting, chromatin coordination on single molecules
Single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA)[99]	EcoGII (m6dA on all A), MNase	High (<5 bp)	PacBio	Oligonucleosome patterns, chromatin accessibility, transcription factor footprinting
Tagmentation-assisted single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA-Tag)[100]	EcoGII (m6dA on all A), Tn5 transposase	High (<5 bp)	PacBio	Nucleosome positioning, chromatin accessibility, transcription factor footprinting, PacBio library preparation
Directed methylation with long-read sequencing (DiMeLo-seq)[104]	pA–Hia5 or pAG–Hia5	Low (within 200 bp)	PacBio, ONT	Protein–DNA interactions, coordination between binding sites on single molecules
BIND&MODIFY[105]	pA–EcoGII (m6dA on all A)	Low (within 100 bp)	ONT	Measuring protein–DNA interactions

5mC, 5-methylcytosine; m6dA, N^6-methyladenine; MNase, micrococcal nuclease; ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences; pA–EcoGII, fusion of protein A and EcoGII; pAG–Hia5, fusion of proteins A and G with Hia5. [a]Estimates of theoretical resolution are based on analysis described previously[96].

different from established short-read methods (such as ChIP-seq, CUT&RUN and CUT&Tag). Short-read methods enrich bound regions, producing peaks of enrichment that cover a small percentage of the genome (<10%[106,107]) but often contain >50% of sequenced reads (the so-called fraction of reads in peaks)[108]. By contrast, the single-molecule methods discussed above have no built-in enrichment step, and although this makes them more quantitative and removes bias, it also requires whole-genome sequencing in order to obtain the same genome-wide signal. Fortunately, recent efforts have shown that these labelling techniques can be combined with enrichment methods for long reads[94,104], allowing cost-effective profiling.

Measuring chromosome conformation
Moving to a larger scale, there is an interplay between DNA methylation, chromatin state, protein–DNA interactions and DNA organization in the nucleus. The three-dimensional organization of the genome plays a critical role in gene regulation, development and human disease (reviewed in refs. 109,110). Primary methods used to measure three-dimensional organization rely on proximity ligation and are known as chromatin conformation capture (3C) assays (reviewed in ref. 111). Most of these methods measure pairwise interactions with short-read sequencing and fail to capture information about potential cooperation between multiple loci[112]. Although methods that do not rely on proximity ligation make it possible to measure multi-way contacts[113], long-read sequencing platforms have the potential to read long fragments from 3C-based experiments that represent multi-way interactions and have been employed in a variety of methods. PacBio sequencing was initially employed by a method measuring chromosomal walks in which

3C DNA was directly sequenced[114]. However, the long-read data were mostly used to validate short-read data, the reads were not very long (<8 kb) and the data produced represented <0.5× coverage of the mouse and human genomes, limiting what information could be gleaned[114]. Multi-contact circular chromosome conformation capture (MC-4C) employed circular chromosome conformation capture combined with Cas9 targeting to measure all interactions at one locus (a so-called 'one versus all approach') with ONT sequencing[115,116]. Again, the average sequenced read size was not very long (~2 kb), owing in part to the use of PCR, with most reads measuring three-way or four-way contacts and some measuring ten contacts[115]. Genome-wide methods such as multi-contact 3C (MC-3C)[117] and Pore-C[118] do not employ PCR and are 'all versus all' methods (that is, all contacts at all loci are measured) like Hi-C and chromosomal walks. MC-3C used PacBio, whereas Pore-C used ONT. Of these two methods, the data from Pore-C best demonstrate the potential of these approaches owing to extremely deep sequencing (up to >132× genome coverage)[118]. With high-depth data, the authors were able to explore CpG methylation on haplotype-specific, multi-way interactions on single molecules. In a good example of how quickly this area is moving, Pore-C has already been modified to reduce cost and improve throughput with a method termed high-throughput Pore-C (HiPore-C)[119].

Short reads on single-molecule platforms
Although single-molecule sequencing typically emphasizes read length, both PacBio and ONT technologies can sequence short nucleic acid fragments. Despite Illumina (and other short-read sequencers) dominating the short-read sequencing field, approaches that sequence

short reads on ONT and PacBio have gained traction. The portability, low physical footprint and ability to analyse sequencing data in real-time make ONT sequencing devices ripe for use with short reads directly at the bench or in the field, without the need for a sequencing core. Single-molecule sequencing can reduce cost as multiple types of -omics data (for example, methylation and genetic variation) can be gleaned from a single sequencing run. The increases in throughput and accuracy of these single-molecule platforms provide advantages that have made them even more attractive for short-read sequencing. These advantages fall into the 'iron triangle' of project management: fast, good or cheap.

Fast: portability and speed

Recent attempts to detect chromosomal abnormalities by optimizing short-read sequencing on ONT highlight the advantage of the low cost and small size of the ONT sequencing devices, especially the ONT Flongle flow cells and ONT MinION flow cells. These aspects could make sequencing more accessible for environments with limited resources and bring these assays from centralized cores to the laboratory benchtop. Additionally, real-time sequencing with ONT enables rapid turnaround times compared to waiting for a completed sequencing run[120,121]. Chromosomal abnormalities, including aneuploidies and copy number variants (CNVs), play a role in human disease and are commonly screened for during pregnancy and in cancer (reviewed in refs. 122,123). Multiple studies have shown that short-read sequencing can be optimized for the portable ONT MinION device to detect aneuploidies[124,125] and CNVs[126,127]. These approaches showed that sequencing libraries could be multiplexed, detected abnormalities were concordant with Illumina sequencing, only 0.5–2 million reads were required and sufficient reads could be obtained in under 3 hours (Fig. 3a). Additionally, similar CNV estimates were observed on the same sequencing device with short or long reads, underscoring the flexibility of these devices[126].

Good: multimodal measurements

An important advantage for single-molecule platforms is that base modification information is acquired for free (not counting computational requirements) alongside the primary sequence. Specifically, short-read single-molecule assays can take advantage of modification data to measure cell-free DNA (cfDNA), which is fragmented DNA found in plasma that is usually the same length as DNA wrapped around a nucleosome (~150 bp). cfDNA has become a popular diagnostic tool owing to the relative ease of collection (via blood draws or 'liquid biopsies') and has been used to analyse fetal DNA during pregnancy, circulating tumour DNA and donor-derived DNA in transplant patients (reviewed in refs. 128,129). As reported in both published and preprint articles, cfDNA has been sequenced with PacBio and ONT to detect fetal DNA in maternal blood[130,131] and assay circulating tumour DNA[132–135]. The ability to measure native CpG methylation and patterns from fragment ends (known as 'fragmentomics'[129]) has been used to classify placental and maternal DNA[130], show that tumour-derived DNA had lower methylation than non-tumour-derived DNA[132], estimate tissue-of-origin and cell-type proportions (Fig. 3b), footprint transcription factor binding sites and measure nucleosome positioning[133]. ONT and PacBio platforms can also capture any longer fragments in these liquid biopsies, revealing previously unknown biology. For example, long reads (>1 kb) can constitute a large proportion (up to ~41%) of cfDNA reads in maternal plasma and the percentage of long reads increases as pregnancy progresses[130].

Though exogenous labelling methods are a focus of single-molecule chromatin assay development (see 'Mapping protein–DNA interactions'), methods sequencing short fragments from chromatin assays have also emerged. For example, Array-seq simply sequences the typical MNase digestion ladder to measure nucleosome positioning with ONT[136] and short fragments from native ChIP-seq without amplification have been sequenced with PacBio[137], allowing for both protein binding and native DNA modifications to be measured simultaneously. Another example is DamID, which uses exogenous DNA adenine methyltransferase (Dam) labelling and methylation-sensitive restriction enzyme digestion to probe protein–DNA interactions[138]. DamID output has been directly sequenced with ONT both with amplification (RNA Pol DamID (RAPID))[139] and without amplification (nanopore-DamID)[140], the latter reported in a recent preprint. These approaches have been shown to benefit from the single-molecule platforms that can sequence longer reads, measuring binding sites in repetitive sequences and segmental duplications as well as simultaneously investigating protein–DNA binding and native methylation[140].

Good: accuracy

Two primary methods have been used to improve the accuracy of reads on single-molecule platforms: consensus methods and molecular indexing methods. Consensus methods have received the most attention with various approaches existing for both ONT and PacBio. PacBio sequencing natively supports consensus sequencing ('circular consensus sequencing' (CCS) with PacBio HiFi) and has been used on both short fragments (<1,000 bp)[141] and long fragments (>13 kb)[142] to generate highly accurate (99.8%)[142] consensus reads. As ONT does not sequence circular molecules, a variety of methods have been developed using rolling circle amplification to generate linear molecules composed of concatemers of the original molecule (Fig. 3c). These methods usually begin with linear fragments of DNA that are circularized by intramolecular ligation[143], molecular inversion probes[144], ligation into a backbone[145], or by using Gibson assembly and a common DNA splint[146]. The circular molecules are then amplified using the phi29 polymerase to create long concatemerized molecules. After sequencing, concatemers are identified and a consensus sequence of the original molecule is constructed (Fig. 3c). Even though long reads could be used with these methods, during development these methods have focused on short reads (<1000 bp) down to 52 bp[144]. All of these methods show increased accuracy (for example, improving from 74% to >95% accuracy[144]) when consensus molecules are constructed, with a recent publication reporting the added benefit of increasing the sequencing yield compared to sequencing the short fragments directly[146].

In addition to consensus sequencing, unique molecular identifiers (UMIs) have been developed for single-molecule platforms and incorporated into amplicon sequencing[147]. UMIs were shown to improve the error rate of both ONT and PacBio (all >99.5% accuracy) and remove PCR chimeras that may arise during amplification. Although the UMIs were shown to work with long amplicons (>4,000 bp), they have the potential to be used in short-read methodologies as well.

Regardless of the approach used to improve accuracy, systematic errors in sequencing reads from these single-molecule platforms will prevent all errors from being corrected. For example, nanopore sequencing is error-prone in low-complexity sequences[148] and homopolymer sequences, even with the latest commercially available pores[7]. PacBio is more accurate than ONT in general, but also shows systematic errors in homopolymer regions[147,149]. That said, further improvement is possible as indicated by recent efforts combining PacBio CCS with UMIs
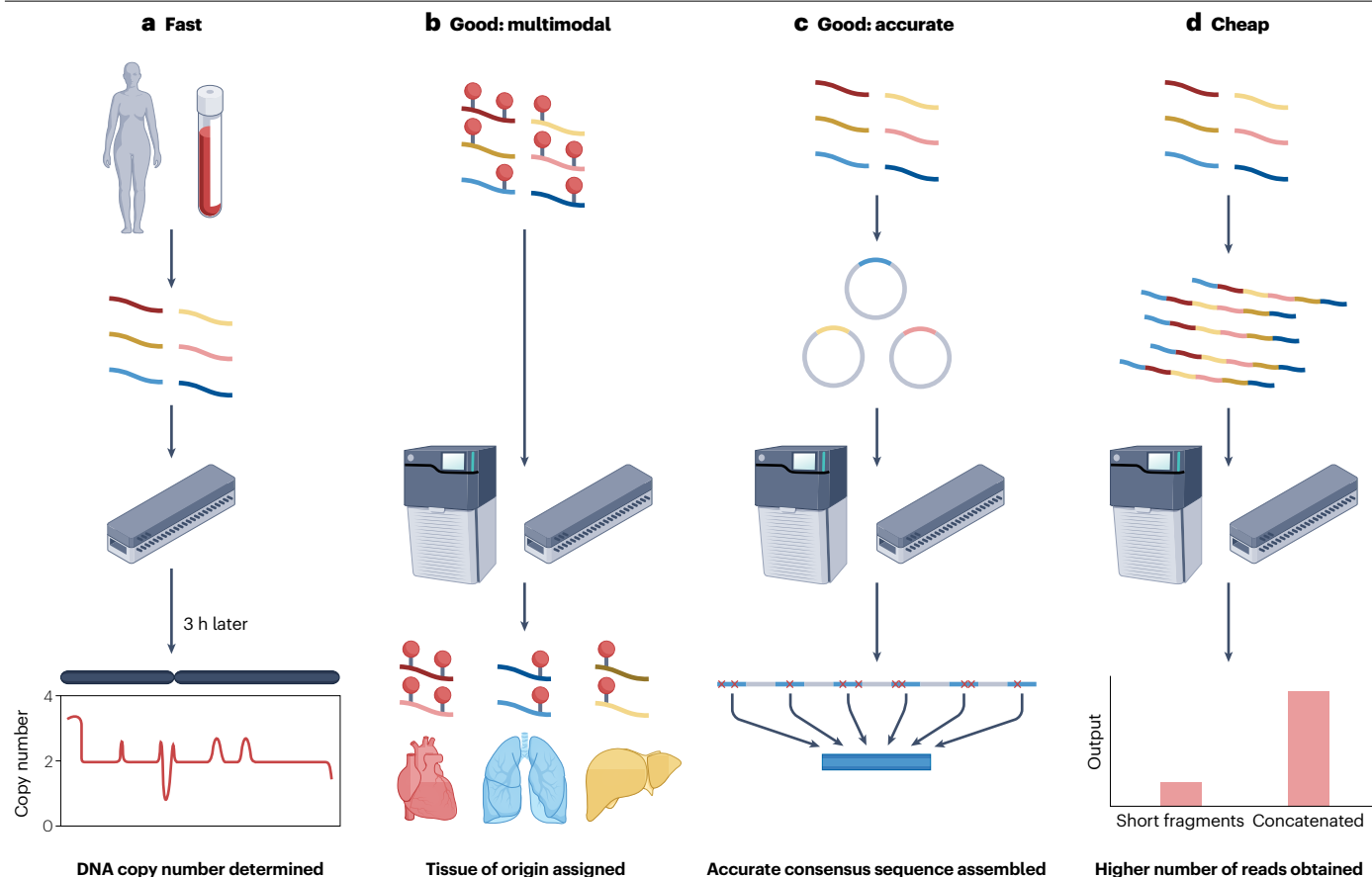
Fig. 3 | Applications of shorter-read sequencing (<5 kb) on single-molecule platforms. a, Short reads can be quickly sequenced on portable Oxford Nanopore sequencing devices, returning real-time information about copy number variants and aneuploidy in 3 h or less. **b**, Primary sequence, fragment patterns and endogenous methylation can be measured simultaneously with single-molecule platforms, and that information can be used to assign reads to tissues of origin. **c**, Accuracy of short reads on single-molecule platforms can be improved by correcting for errors by reading the same molecule multiple times. **d**, The cost of sequencing short fragments on single-molecule platforms can be decreased by combining multiple different short molecules into a single, long molecule.

that resulted in very few errors[147] and the improvement of accuracy seen by retraining nanopore basecallers with troublesome sequences[150].

Cheap: increasing throughput

Both PacBio and ONT typically produce fewer reads per sequencing run than an Illumina device, affecting the cost of these platforms for read-counting applications such as assaying CNVs and RNA-seq. Because of this, a set of methods have been developed to increase the yield of short reads on single-molecule platforms. The methods are similar to approaches used to increase Sanger sequencing throughput in the 1990s[151,152] and rely on concatenating short fragments into artificial, long fragments to increase throughput using either Gibson assembly[153] or sticky-end ligation[154–156] (Fig. 3d). For example, a method published in a recent preprint article, multiplexed arrays sequencing of isoforms (MAS-ISO-seq), shows ~15–25× increase in throughput with PacBio[156] and sampling molecules using re-ligated fragments (SMURF-seq) achieves a ~3× increase on ONT[155]. Based on the gain in sequencing output, both methods can reduce the cost per million reads or full-length transcripts from >US$883 (PacBio) and >US$415 (ONT) to <US$56 (PacBio) and <US$146 (ONT) (see Supplementary Note and Supplementary Data).

These approaches have been used in a variety of ways including identifying cancer variants[153,155], measuring CNVs[154] and sequencing RNA isoforms[156,157].

It is currently unclear if any biases are introduced during these concatemerization methods and how they may affect the resulting data. Two of the methods recently described in preprints, MAS-ISO-seq[156] and HIT-scISOseq[157], both show relative depletion of longer spike-in RNA variants compared to shorter transcripts when compared to PacBio Iso-Seq. This could be due to any step in those protocols, including PCR, uracil digestion or ligation. Furthermore, the ligases used in these assays may have some GC bias, as was shown for serial analysis of gene expression (SAGE)[151,158,159]. Finally, these concatemerization methods rely on being able to accurately identify the junction sites between molecules in order to split them into individual fragments. Although most of these methods are paired with software for resolving concatemers, the base pair accuracy of these methods has not been fully elucidated. For example, ConcatSeq showed a small distribution of fragments deviating from the expected fragment length[153]. We expect that benchmarking and further exploration of these data will elucidate any sources of bias.

Glossary

Basecaller

An algorithm that converts the raw signal from nucleic acid sequencing into the bases that the signal represents.

Centromeres

The region of a chromosome where the kinetochore attaches during cell division, typically an extremely repetitive region.

Chemical bisulfite conversion

A method used to measure the DNA modifications 5-methylcytosine and 5-hydroxymethylcytosine. DNA is treated with sodium bisulfite, which converts unmodified cytosine to uracil, whereas modified cytosines are protected from conversion. Following conversion and PCR, unmodified cytosines are read as thymines when sequenced, whereas modified cytosines remain cytosines.

Chromatin immunoprecipitation followed by sequencing

(ChIP–seq). A method for directly measuring protein–DNA binding with antibody-mediated immunoprecipitation of protein–DNA complexes.

Cleavage under targets & release using nuclease

(CUT&RUN). A method for directly measuring protein–DNA binding with antibody-guided DNA digestion with a micrococcal nuclease.

Cleavage under targets and tagmentation

(CUT&Tag). A method for directly measuring protein–DNA binding with antibody-guided transposition and fragmentation (tagmentation) with Tn5 transposase.

Cycle dephasing

Mechanism of error that affects sequencing devices using polymerase colonies (polonies). This occurs when clonal molecules within the same cluster are not all elongated in a given extension step, diluting the sequencing signal during subsequent cycles as the molecules become out of phase. More molecules become 'dephased' with each additional sequencing cycle, leading to increasingly lower sequencing quality as different positions on the template contribute to the signal.

Dynamic time warping

An algorithm for measuring similarity between two time series. In this context it refers to matching experimental nanopore data to a modelled electrical signal from a reference DNA sequence to identify the correct sequence from a database.

Human Genome Project

An international effort launched in 1990 with the primary goal of assembling the human genome. The project was completed in 2003.

ONT Flongle flow cell

Low-throughput flow cell (<1 Gb) from Oxford Nanopore Technologies. This flow cell can be sequenced on MinION or GridION sequencing devices.

ONT MinION

Hand-held sequencing device from Oxford Nanopore Technologies that can perform sequencing with MinION or Flongle flow cells.

ONT MinION flow cell

Medium-throughput (2–20 Gb) flow cell from Oxford Nanopore Technologies. This flow cell can be sequenced on MinION or GridION sequencing devices.

ONT PromethION

High-throughput sequencing device from Oxford Nanopore Technologies that can perform sequencing with PromethION flow cells.

ONT PromethION flow cell

High-throughput (50–100+ Gb) flow cell from Oxford Nanopore Technologies. This flow cell can be sequenced on PromethION sequencing devices.

PacBio RS II

Sequencing device released by Pacific Biosciences in 2013 that can perform single-molecule, real-time sequencing.

PacBio Sequel II

Sequencing device released by Pacific Biosciences in 2019 that can perform single-molecule, real-time sequencing.

Sequencing depth

The number of reads that map to a given locus, also known as sequencing coverage. This is usually represented as an average, and a locus can refer to a single nucleotide, region(s) of interest, entire chromosome(s) or entire genomes. We would consider 'high' coverage or depth as >100× for most assays.

Telomeres

Repetitive regions at the end of chromosomes.

Tn5 transposase

A bacterial protein that facilitates the movement of DNA sequences through a 'cut and paste' mechanism. This protein has become a valuable molecular biology tool with its uses ranging from efficient library preparation to probing chromatin state.

Unique molecular identifiers

(UMIs). Short sequences of random nucleotides that tags an individual nucleic acid molecule. UMIs can be used to identify subsequently amplified fragments that arose from the same original molecule, mitigating bias introduced during PCR and allowing for more accurate quantification.

Whole-genome sequencing

A sequencing approach that attempts to obtain reads that map to all bases in the genome.

Conclusions and future perspectives

The increasing use of single-molecule sequencing platforms in genomics has led to an increase in applications beyond typical use cases. As they enter the mainstream, the number of creative uses of these platforms will increase and the methods detailed in this Review will be optimized, refined and expanded. If anything, development will be accelerated in coming years owing to the massive increase in the use of ONT sequencing to monitor the SARS-CoV-2 pandemic, as illustrated by ~50% of COVID-19 sequencing across the African continent being performed with ONT[160]. This increase will give an expanded population of researchers ready access to single-molecule sequencing technology.

Targeted sequencing methods will be improved to capture longer reads to take full advantage of these platforms. The optimization of these methods will lead to greater read depths and lengths, enabling applications that need ultra-high-depth sequencing such as identifying somatic mosaic variants or intratumoural heterogeneity. Further developments in combining methods, such as Cas-mediated enrichment with adaptive sampling[161], will improve on-target rates and drive costs even lower. Targeted long reads are likely to generate new insights into the direct molecular impact of mutations and alterations as their single-molecule nature is a proxy for cellular heterogeneity in complex clinical samples.

Since their inception, short-read assays measuring protein–DNA binding have been developed to reduce input even to the single-cell level (reviewed in ref. 78) and to measure multiple protein–DNA interactions simultaneously[162,163]. We expect single-molecule methods to follow the same trajectory as they offer an appealing route to quantitative methods for measuring these interactions. Early work on the coordination

Review article

of epigenetic marks on long, single reads − in some cases as long as 100 kb − offers tantalizing views into exploring epigenetic heterogeneity, such as examining the temporal dynamics of T cell activation[94]. However, determining whether exogenous labelling variation is biological or technical requires careful molecular controls. Potential confounding technical aspects include the extent to which both protein and antibody penetrate cells and/or nuclei and their binding efficiencies, fidelity of modification calling and enzyme labelling efficiencies.

Although the throughput of short reads on single-molecule platforms is improving, it still remains at a relatively high cost per million reads for counting applications, such as RNA-seq, CNV analysis and CUT&RUN. Improvements increasing the number of short reads obtained in a single sequencing run will enable sample multiplexing, driving down the cost of sequencing. With increasing throughput, we expect more short reads from a variety of assays to be sequenced on these long-read platforms owing to decreasing cost, increased speed and portability, and the ability to gain multimodal information.

Although we focus on DNA-based methods in this Review, we believe the ability to sequence RNA directly will also have an important role in a variety of methods going forward. However, at this time, direct RNA sequencing lags behind DNA sequencing and will require improvement in many aspects, including accuracy, to spur further use[164]. Similarly, we expect the young field of protein sequencing on nanopores to continue to advance[165], eventually completing our ability to measure the central dogma in its entirety.

Finally, we imagine these advances could be combined with parallel advances in the portability and flexibility of sample collection[166] and data analysis[167,168]. This is an especially exciting prospect when considering their use with portable ONT sequencing, which could lead to sequencing assays leaving core facilities for use directly at the bench or even the field. Improvements and future developments in these methods set the stage for a more flexible and accessible field of genomics, pushing it into a new and exciting era.

References

1. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
2. Erlich, Y., Mitra, P. P., delaBastide, M., McCombie, W. R. & Hannon, G. J. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat. Methods* **5**, 679–682 (2008).
3. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
4. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614 (2020).
   This comprehensive review goes into great detail about long-read sequencing technologies. It is a good resource for further information about the sequencing technologies that are the focus of this manuscript.
5. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
6. Timp, W. et al. Think small: nanopores for sensing and synthesis. *IEEE Access.* **2**, 1396–1408 (2014).
7. Sereika, M. et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* **19**, 823–826 (2022).
8. Rhie, A. et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
9. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
10. Chaisson, M. J. P. et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2015).
11. Aganezov, S. et al. Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Res.* **30**, 1258–1273 (2020).
12. Ebert, P. et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021).
13. Beyter, D. et al. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat. Genet.* **53**, 779–786 (2021).
14. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
15. Altemose, N. et al. Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).
16. Gershman, A. et al. Epigenetic patterns in a complete human genome. *Science* **376**, eabj5089 (2022).
17. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
18. Glinos, D. A. et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* **608**, 353–359 (2022).
19. Pratanwanich, P. N. et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **39**, 1394–1402 (2021).
20. Gampawar, P. et al. Evaluation of the performance of AmpliSeq and SureSelect exome sequencing libraries for ion proton. *Front. Genet.* **10**, 856 (2019).
21. Togi, S., Ura, H. & Niida, Y. Optimization and validation of multimodular, long-range PCR-based next-generation sequencing assays for comprehensive detection of mutation in tuberous sclerosis complex. *J. Mol. Diagn.* **23**, 424–446 (2021).
22. Barnes, W. M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl Acad. Sci. USA* **91**, 2216–2220 (1994).
23. Jia, H., Guo, Y., Zhao, W. & Wang, K. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Sci. Rep.* **4**, 5737 (2014).
24. Walczak, M. et al. Long-range PCR libraries and next-generation sequencing for pharmacogenetic studies of patients treated with anti-TNF drugs. *Pharmacogenomics J.* **19**, 358–367 (2019).
25. Brait, N., Külekçi, B. & Goerzer, I. Long range PCR-based deep sequencing for haplotype determination in mixed HCMV infections. *BMC Genomics* **23**, 31 (2022).
26. Potapov, V. & Ong, J. L. Examining sources of error in PCR by single-molecule sequencing. *PLoS ONE* **12**, e0169774 (2017).
27. Tyson, J. R. et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.04.283077v1 (2020).
28. 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).
29. Borràs, D. M. et al. Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum. Mutat.* **38**, 870–879 (2017).
30. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
31. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
    This landmark study developed an amplicon-based assay for sequencing the Ebola genome from infected individuals using the ONT MinION portable sequencer. It has served as the template for portable disease monitoring efforts during Zika, COVID-19 and monkeypox outbreaks.
32. Turner, E. H., Ng, S. B., Nickerson, D. A. & Shendure, J. Methods for genomic partitioning. *Annu. Rev. Genomics Hum. Genet.* **10**, 263–284 (2009).
33. Gnirke, A. et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
34. Leung, A. W.-S. et al. ECNano: a cost-effective workflow for target enrichment sequencing and accurate variant calling on 4800 clinically significant genes using a single MinION flowcell. *BMC Med. Genomics* **15**, 43 (2022).
35. Zhang, L. et al. Efficient CNV breakpoint analysis reveals unexpected structural complexity and correlation of dosage-sensitive genes with clinical severity in genomic disorders. *Hum. Mol. Genet.* **26**, 1927–1941 (2017).
36. Beck, C. R. et al. Megabase length hypermutation accompanies human structural variation at 17p11.2. *Cell* **176**, 1310–1324.e10 (2019).
    This study is a great example of the use of hybridization capture followed by PacBio sequencing to help to characterize structural variation in a disease cohort. PacBio sequencing increased the ability to determine base pair-level breakpoints in this cohort, particularly when they occurred in repeat regions, emphasizing the utility of this method of long-read enrichment.
37. Yamaguchi, K. et al. Application of targeted nanopore sequencing for the screening and determination of structural variants in patients with Lynch syndrome. *J. Hum. Genet.* **66**, 1053–1060 (2021).
38. Wang, M. et al. PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics* **16**, 214 (2015).
39. Giolai, M. et al. Targeted capture and sequencing of gene-sized DNA molecules. *Biotechniques* **61**, 315–322 (2016).
40. Bethune, K. et al. Long-fragment targeted capture for long-read sequencing of plastomes. *Appl. Plant Sci.* **7**, e1243 (2019).
41. Lefoulon, E. et al. Large enriched fragment targeted sequencing (LEFT-SEQ) applied to capture of Wolbachia genomes. *Sci. Rep.* **9**, 5939 (2019).
42. Steiert, T. A. et al. High-throughput method for the hybridisation-based targeted enrichment of long genomic fragments for PacBio third-generation sequencing. *NAR Genom. Bioinform.* **4**, lqac051 (2022).
43. Karamitros, T. & Magiorkinis, G. A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res.* **43**, e152 (2015).

Review article

44. Karamitros, T. & Magiorkinis, G. in *Next Generation Sequencing: Methods and Protocols* (eds Head, S. R., Ordoukhanian, P. & Salomon, D. R.) 43–51 (Springer New York, 2018).

45. Lee, I., Workman, R. E., Wang, J. Z. & Timp, W. Use of Agilent SureSelect to perform targeted long-read nanopore sequencing. *Agilent Application Note* (Agilent Technologies, 2017).

46. Roe, D. et al. Efficient sequencing, assembly, and annotation of human KIR haplotypes. *Front. Immunol.* **11**, 582927 (2020).

47. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

48. Lee, N. C. O., Larionov, V. & Kouprina, N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.* **43**, e55 (2015).

49. Jiang, W. et al. Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* **6**, 8101 (2015).

50. Gabrieli, T. et al. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.* **46**, e87 (2018).

51. Tsai, Y.-C. et al. Amplification-free, CRISPR-Cas9 targeted enrichment and SMRT sequencing of repeat-expansion disease causative genomic regions. Preprint at *bioRxiv* https://doi.org/10.1101/203919v1 (2017).

52. Tsai, Y.-C. et al. in *Genomic Structural Variants in Nervous System Disorders* (ed Proukakis, C.) 95–120 (Springer, 2022).

53. Watson, C. M. et al. Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Lab. Invest.* **100**, 135–146 (2020).

54. Giesselmann, P. et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat. Biotechnol.* **37**, 1478–1481 (2019).

55. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
    In this study, the authors develop a targeted Cas9 digestion approach to enrich sequencing reads in regions of interest. They show that designing multiple guide RNAs on each side of the target region increases coverage and that ten targets can be multiplexed in one experiment, achieving high coverage for all the targets.

56. Iyer, S. V., Kramer, M., Goodwin, S. & McCombie, W. R. ACME: an affinity-based Cas9 mediated enrichment method for targeted nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2022.02.03.478550v2 (2022).

57. Wallace, A. D. et al. CaBagE: a Cas9-based background elimination strategy for targeted, long-read DNA sequencing. *PLoS ONE* **16**, e0241253 (2021).

58. Stevens, R. C. et al. A novel CRISPR/Cas9 associated technology for sequence-specific nucleic acid enrichment. *PLoS ONE* **14**, e0215441 (2019).

59. Bruijnesteijn, J., van der Wiel, M., de Groot, N. G. & Bontrop, R. E. Rapid characterization of complex killer cell immunoglobulin-like receptor (KIR) regions using Cas9 enrichment and nanopore sequencing. *Front. Immunol.* **12**, 722181 (2021).

60. Gilpatrick, T. et al. IVT generation of guideRNAs for Cas9-enrichment nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.07.527484v1 (2023).

61. Loose, M., Malla, S. & Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods* **13**, 751–754 (2016).
    This landmark study is the first to demonstrate that DNA molecules being sequenced with a nanopore could be selectively sequenced using only computational methods. As part of this, the authors develop a method using dynamic time warping to match the electrical signal from the sequencing read, in real time, to a small reference genome.

62. Kovaka, S., Fan, Y., Ni, B., Timp, W. & Schatz, M. C. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotechnol.* **39**, 431–441 (2021).
    In this study, the authors develop a new algorithmic approach for aligning nanopore electrical signals to a reference sequence, making it possible to apply nanopore adaptive sampling to human-sized genomes. The authors use this approach to enrich a gene panel of 148 hereditary cancer genes.

63. Zhang, H. et al. Real-time mapping of nanopore raw signals. *Bioinformatics* **37**, i477–i483 (2021).

64. Bao, Y. et al. SquiggleNet: real-time, direct classification of nanopore signals. *Genome Biol.* **22**, 298 (2021).

65. Han, R., Wang, S. & Gao, X. Novel algorithms for efficient subsequence searching and mapping in nanopore raw signals towards targeted sequencing. *Bioinformatics* **36**, 1333–1343 (2020).

66. Masutani, B. & Morishita, S. A framework and an algorithm to detect low-abundance DNA by a handy sequencer and a palm-sized computer. *Bioinformatics* **35**, 584–592 (2019).

67. Dunn, T. et al. in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture* 535–549 (Association for Computing Machinery, 2021).

68. Payne, A. et al. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* **39**, 442–450 (2021).
    This article describes software to perform nanopore adaptive sampling sequencing by using basecalled sequencing reads, as opposed to nanopore electrical signals. The authors show that this approach can enrich entire chromosomes, half of the human exome and a 700+ cancer gene panel. This approach has been incorporated into ONT sequencing software.

69. Edwards, H. S. et al. Real-time selective sequencing with RUBRIC: read until with basecall and reference-informed criteria. *Sci. Rep.* **9**, 11475 (2019).

70. Ulrich, J.-U., Lutfi, A., Rutzen, K. & Renard, B. Y. ReadBouncer: precise and scalable adaptive sampling for nanopore sequencing. *Bioinformatics* **38**, i153–i160 (2022).

71. Martin, S. et al. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* **23**, 11 (2022).

72. Payne, A. et al. Barcode aware adaptive sampling for GridION and PromethION Oxford Nanopore sequencers. Preprint at *bioRxiv* https://doi.org/10.1101/2021.12.01.470722v2 (2022).

73. Madsen, E. B., Höijer, I., Kvist, T., Ameur, A. & Mikkelsen, M. J. Xdrop: targeted sequencing of long DNA molecules from low input samples using droplet sorting. *Hum. Mutat.* **41**, 1671–1679 (2020).

74. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).

75. Minnoye, L. et al. Chromatin accessibility profiling methods. *Nat. Rev. Methods Prim.* **1**, 10 (2021).

76. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).

77. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).

78. Preissl, S., Gaulton, K. J. & Ren, B. Characterizing *cis*-regulatory elements using single-cell epigenomics. *Nat. Rev. Genet.* https://doi.org/10.1038/s41576-022-00509-1 (2022).

79. Brinkman, A. B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).

80. Clark, S. J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).

81. Luo, C. et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom.* **2**, 100107 (2022).

82. Fehér, Z., Kiss, A. & Venetianer, P. Expression of a bacterial modification methylase gene in yeast. *Nature* **302**, 266–268 (1983).

83. Singh, J. & Klar, A. J. Active genes in budding yeast display enhanced in vivo accessibility to foreign DNA methylases: a novel in vivo probe for chromatin structure of yeast. *Genes Dev.* **6**, 186–196 (1992).

84. Kladde, M. P. & Simpson, R. T. Positioned nucleosomes inhibit Dam methylation in vivo. *Proc. Natl Acad. Sci. USA* **91**, 1361–1365 (1994).

85. Kladde, M. P., Xu, M. & Simpson, R. T. Direct study of DNA-protein interactions in repressed and active chromatin in living cells. *EMBO J.* **15**, 6290–6300 (1996).

86. Xu, M., Simpson, R. T. & Kladde, M. P. Gal4p-mediated chromatin remodeling depends on binding site position in nucleosomes but does not require DNA replication. *Mol. Cell. Biol.* **18**, 1201–1212 (1998).

87. Sönmezer, C. et al. Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. *Mol. Cell* **81**, 255–267.e6 (2021).

88. Nabilsi, N. H. et al. Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* **24**, 329–339 (2014).

89. Kelly, T. K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).

90. Kleinendorst, R. W. D., Barzaghi, G., Smith, M. L., Zaugg, J. B. & Krebs, A. R. Genome-wide quantification of transcription factor binding at single-DNA-molecule resolution using methyl-transferase footprinting. *Nat. Protoc.* **16**, 5673–5706 (2021).

91. Wang, Y. et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).

92. Oberbeckmann, E. et al. Absolute nucleosome occupancy map for the Saccharomyces cerevisiae genome. *Genome Res.* **29**, 1996–2009 (2019).
    This study develops ODM-seq, which is one of the first methods to combine methyltransferase labelling with ONT sequencing. The authors use this method to quantify nucleosome occupancy in the yeast genome and calculate the exact number of nucleosomes per yeast cell.

93. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
    This study develops a method that combines GpC methyltransferase labelling with ONT sequencing to assay open chromatin in human cells. As part of this, the authors develop a novel ONT model for calling exogenous GpC methylation (representing chromatin accessibility) and endogenous CpG methylation simultaneously on single, long molecules.

94. Battaglia, S. et al. Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nat. Genet.* **54**, 1504–1513 (2022).

95. Kong, Y. et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science* **375**, 515–522 (2022).

96. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).

97. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
    This study develops a method that combines adenine methyltransferase labelling (m6dA) with PacBio sequencing to probe chromatin accessibility in the *Drosophila melanogaster* S2 cell line. Adenine labelling provides higher-resolution data than CpG or GpC labelling techniques and the authors use this labelling to measure the coordination between adjacent regulatory elements.

98. Dubocanin, D. et al. Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin. Preprint at *bioRxiv* https://doi.org/10.1101/2022.05.09.491186v1 (2022).

Review article

99. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).

100. Nanda, A. S. et al. Sensitive multimodal profiling of native DNA by transposase-mediated single-molecule sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2022.08.07.502893v2 (2022).

101. Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S. & Ahmad, K. Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife* **9**, e63274 (2020).

102. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

103. Eliasson, M., Andersson, R., Olsson, A., Wigzell, H. & Uhlén, M. Differential IgG-binding characteristics of staphylococcal protein A, streptococcal protein G, and a chimeric protein AG. *J. Immunol.* **142**, 575–581 (1989).

104. Altemose, N. et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat. Methods* **19**, 711–723 (2022).
This study develops a method using antibody-directed adenine methyltransferase labelling (m6dA) combined with ONT and PacBio sequencing to directly measure protein–DNA interactions of specific proteins. The authors perform extensive optimization of this method and also develop an approach to enrich for sequencing reads in centromeric regions.

105. Weng, Z. et al. BIND&MODIFY: a long-range method for single-molecule mapping of chromatin modifications in eukaryotes. *Genome Biol.* **24**, 61 (2023).

106. Gopi, L. K. & Kidder, B. L. Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. *Nat. Commun.* **12**, 1419 (2021).

107. Battle, S. L. et al. Enhancer chromatin and 3D genome architecture changes from naive to primed human embryonic stem cell states. *Stem Cell Rep.* **12**, 1129–1144 (2019).

108. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

109. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).

110. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

111. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).

112. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Mol. Cell* **77**, 688–708 (2020).

113. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).

114. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).

115. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).

116. Vermeulen, C. et al. Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat. Protoc.* **15**, 364–397 (2020).

117. Tavares-Cadete, F., Norouzi, D., Dekker, B., Liu, Y. & Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat. Struct. Mol. Biol.* **27**, 1105–1114 (2020).

118. Deshpande, A. S. et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).
This study develops a method adapting the 3C chromatin conformation assay to ONT sequencing as well as software for analysing this type of dataset. Combining 3C with long-read sequencing allows for multi-way contacts to be measured, revealing insights about gene regulation that cannot be ascertained through short-read, pairwise methods.

119. Zhong, J.-Y. et al. High-throughput Pore-C reveals the single-allele topology and cell type-specificity of 3D genome folding. *Nat. Commun.* **14**, 1250 (2023).

120. Magi, A. et al. Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data. *Bioinformatics* **35**, 4213–4221 (2019).

121. Munro, R. et al. MinoTour, real-time monitoring and analysis for nanopore sequencers. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btab780 (2021).

122. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).

123. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

124. Wei, S. & Williams, Z. Rapid short-read sequencing and aneuploidy detection using MinION nanopore technology. *Genetics* **202**, 37–44 (2016).

125. Wei, S., Weiss, Z. R. & Williams, Z. Rapid multiplex small DNA sequencing on the MinION nanopore sequencing platform. *G3* **8**, 1649–1657 (2018).
This pioneering study was one of the first to demonstrate that portable ONT sequencers could be used to sequence short reads. In this study, the authors optimize library preparation for short fragments and use their approach to measure chromosome copy number, correctly identifying aneuploidy in <4 hours of sequencing.

126. Baslan, T. et al. High resolution copy number inference in cancer using short-molecule nanopore sequencing. *Nucleic Acids Res.* **49**, e124 (2021).
In this study, short DNA fragments are sequenced with portable ONT sequencers to determine CNVs in cancer. The authors optimize the loading of short reads on these devices, show that profiles are equivalent to those from Illumina sequencing and demonstrate that accurate results can be determined in <3 hours of sequencing.

127. Martignano, F. et al. Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients. *Mol. Cancer* **20**, 32 (2021).

128. Corcoran, R. B. & Chabner, B. A. Application of cell-free DNA analysis to cancer treatment. *N. Engl. J. Med.* **379**, 1754–1765 (2018).

129. Lo, Y. M. D., Han, D. S. C., Jiang, P. & Chiu, R. W. K. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* **372**, eaaw3616 (2021).

130. Yu, S. C. Y. et al. Single-molecule sequencing reveals a large population of long cell-free DNA molecules in maternal plasma. *Proc. Natl Acad. Sci. USA* **118**, e2114937118 (2021).
In this study, PacBio sequencing is used to measure cfDNA from maternal plasma. The authors are among the first to show that long molecules (>1000 bp) are present in cfDNA and are missed when these samples are sequenced on short-read platforms. They also show that native CpG methylation called by PacBio sequencing can assign reads to a fetal or maternal origin.

131. Cheng, S. H. et al. Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: feasibility assessment. *Clin. Chem.* **61**, 1305–1306 (2015).

132. Choy, L. Y. L. et al. Single-molecule sequencing enables long cell-free DNA detection and direct methylation analysis for cancer patients. *Clin. Chem.* **68**, 1151–1163 (2022).

133. Katsman, E. et al. Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. *Genome Biol.* **23**, 158 (2022).
In this study, the authors explore the feasibility of sequencing circulating tumour DNA with ONT. The authors demonstrate that circulating tumour DNA fragments could be assigned to a tissue of origin based on CpG methylation, identify copy number alterations and measure nucleosome positioning, all in one assay.

134. Lau, B. T. et al. Single molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2022.06.22.497080v1 (2022).

135. Sampathi, S. et al. Nanopore sequencing of clonal IGH rearrangements in cell-free DNA as a biomarker for acute lymphoblastic leukemia. *Front. Oncol.* **12**, 958673 (2022).

136. Baldi, S., Krebs, S., Blum, H. & Becker, P. B. Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat. Struct. Mol. Biol.* **25**, 894–901 (2018).

137. Wu, T. P. et al. DNA methylation on N^6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).

138. Aughey, G. N. & Southall, T. D. Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 25–37 (2016).

139. Gómez-Saldivar, G. et al. Tissue-specific transcription footprinting using RNA Pol DamID (RAPID) in *Caenorhabditis elegans*. *Genetics* **216**, 931–945 (2020).

140. 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/10.1101/2021.08.09.455753v2 (2022).

141. Hebert, P. D. N. et al. A sequel to Sanger: amplicon sequencing that scales. *BMC Genomics* **19**, 219 (2018).

142. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
This landmark study reports the optimization of PacBio HiFi sequencing. It demonstrates that reading the same molecule multiple times increases the accuracy of individual reads.

143. Li, C. et al. INC-Seq: accurate single molecule reads using nanopore sequencing. *Gigascience* **5**, 34 (2016).

144. Wilson, B. D., Eisenstein, M. & Soh, H. T. High-fidelity nanopore sequencing of ultra-short DNA targets. *Anal. Chem.* **91**, 6783–6789 (2019).

145. Marcozzi, A. et al. Accurate detection of circulating tumor DNA using nanopore consensus sequencing. *NPJ Genom. Med.* **6**, 106 (2021).

146. Zee, A. et al. Sequencing Illumina libraries at high accuracy on the ONT MinION using R2C2. *Genome Res.* **32**, 2092–2106 (2022).

147. 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).
This study designs UMIs that work with both PacBio and ONT sequencing. The authors develop software to identify the UMIs and show that they increase read accuracy and reduce the presence of chimeric molecules when PCR is used.

148. Timp, W., Comer, J. & Aksimentiev, A. DNA base-calling from a nanopore using a Viterbi algorithm. *Biophys. J.* **102**, L37–L39 (2012).

149. Mikheenko, A., Prjibelski, A. D., Joglekar, A. & Tilgner, H. U. Sequencing of individual barcoded cDNAs using Pacific Biosciences and Oxford Nanopore Technologies reveals platform-specific error patterns. *Genome Res.* **32**, 726–737 (2022).

150. Tan, K.-T., Slevin, M. K., Meyerson, M. & Li, H. Identifying and correcting repeat-calling errors in nanopore sequencing of telomeres. *Genome Biol.* **23**, 180 (2022).

151. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).

152. Andersson, B. et al. Adaptor-based uracil DNA glycosylase cloning simplifies shotgun library construction for large-scale sequencing. *Anal. Biochem.* **218**, 300–308 (1994).

153. Schlecht, U., Mok, J., Dallett, C. & Berka, J. ConcatSeq: a method for increasing throughput of single molecule sequencing by concatenating short DNA fragments. *Sci. Rep.* **7**, 5252 (2017).

Review article

154. Prabakar, R. K., Xu, L., Hicks, J. & Smith, A. D. SMURF-seq: efficient copy number profiling on long-read sequencers. *Genome Biol.* **20**, 134 (2019).
This study introduces a method called SMURF-seq that concatenates short DNA fragments together for more efficient and cheaper sequencing on ONT. The authors show that concatenated molecules increase the number of reads recovered from a single sequencing run compared to non-concatenated short reads. They go on to show that this method allowed for accurate identification of CNV profiles.

155. Thirunavukarasu, D. et al. Oncogene concatenated enriched amplicon nanopore sequencing for rapid, accurate, and affordable somatic mutation detection. *Genome Biol.* **22**, 227 (2021).

156. Al'Khafaji, A. M. et al. High-throughput RNA isoform sequencing using programmable cDNA concatenation. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.01.462818v1 (2021).

157. Zheng, Y.-F. et al. HIT-scISOseq: high-throughput and high-accuracy single-cell full-length isoform sequencing for corneal epithelium. Preprint at *bioRxiv* https://doi.org/10.1101/2020.07.27.222349v1 (2020).

158. Margulies, E. H., Kardia, S. L. & Innis, J. W. Identification and prevention of a GC content bias in SAGE libraries. *Nucleic Acids Res.* **29**, E60–E60 (2001).

159. Bilotti, K. et al. Mismatch discrimination and sequence bias during end-joining by DNA ligases. *Nucleic Acids Res.* **50**, 4647–4658 (2022).

160. Tegally, H. et al. The evolving SARS-CoV-2 epidemic in Africa: insights from rapidly expanding genomic surveillance. *Science* **378**, eabq5358 (2022).

161. Rubben, K. et al. Cas9 targeted nanopore sequencing with enhanced variant calling improves CYP2D6–CYP2D7 hybrid allele genotyping. *PLoS Genet* **18**, e1010176 (2022).

162. Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazzio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol. Cell* **81**, 4736–4746.e5 (2021).

163. Stuart, T. et al. Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01588-5 (2022).

164. Jain, M., Abu-Shumays, R., Olsen, H. E. & Akeson, M. Advances in nanopore direct RNA sequencing. *Nat. Methods* **19**, 1160–1164 (2022).

165. Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A. & Dekker, C. Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science* **374**, 1509–1513 (2021).

166. Bhamla, M. S. et al. Hand-powered ultralow-cost paper centrifuge. *Nat. Biomed. Eng.* **1**, 0009 (2017).

167. Samarakoon, H. et al. Genopo: a nanopore sequencing analysis toolkit for portable Android devices. *Commun. Biol.* **3**, 538 (2020).

168. Palatnick, A., Zhou, B., Ghedin, E. & Schatz, M. C. iGenomics: comprehensive DNA sequence analysis on your smartphone. *Gigascience* **9**, giaa138 (2020).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
W.T. has two patents (8,748,091 and 8,394,584) licensed to ONT. W.T. has received travel funds to speak at symposia organized by ONT. P.H. declares no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41576-023-00600-1.

Peer review information *Nature Reviews Genetics* thanks Matthew Loose and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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