Nanopore DNA sequencing technologies and their applications towards single-molecule proteomics

Adam Dorey[1,*] and Stefan Howorka[1,*]

[1]*Department of Chemistry & Institute of Structural Molecular Biology, University College London, UK*

*Corresponding authors, a.dorey@ucl.ac.uk, s.howorka@ucl.ac.uk

Abstract

Sequencing of nucleic acids with nanopores has emerged as a powerful tool offering rapid read-out, high accuracy, low cost, and portability. This label-free method for sequencing at the single-molecule level is an achievement on its own. Yet nanopores also show promise for the technologically even more challenging sequencing of polypeptides, something which could considerably benefit biological discovery, clinical diagnostics, and homeland security, as current techniques lack portability and speed. In this Review, we survey the bio-chemical innovations underpinning commercial and academic nanopore DNA/RNA sequencing techniques, and explore how these advances can fuel developments in future protein sequencing with nanopores.

– Alt-text summary

Sequencing of DNA and RNA with nanopores offers rapid read-out, high accuracy, low cost, and portability. This Review surveys technologies underpinning commercial and academic nanopore sequencing and explore how these advances can fuel developments in future protein sequencing with nanopores.

Introduction

Nucleic acid sequencing is relevant in many fields including genomic[1,2], metagenomic[3,4] and proteomic research[5–7], as well as clinical diagnostics[8–10], forensic analysis[11,12], and food safety[13,14]. To make sequencing more accessible, while increasing throughput and lowering costs[15–21], next-generation technologies have been developed. Among these, sequencing by nanopores stands out by achieving portability[22], long read-length[2,23], fast read times[24] and high throughput[24] in a label-free fashion. Recently, nanopore sequencing was named *Nature Methods'* "Method of the Year"[25] reflecting the technology's ability for long reads. Nanopore sequencing was first hypothesized in the 1980s[26–30] and relies on the simple concept of threading individual molecules of single-stranded DNA (ssDNA) through a protein pore's nanoscale passageway known as a lumen. Upon passing the lumen and a tight constriction, termed reading head, the strands are deciphered (Figure 1a and b). The read-out signatures unique to each base are measured as electrical perturbations of ionic current flowing through the membrane-embedded nanopore (Figure 1b). To sequentially read the bases, the negatively charged DNA strand is electrophoretically pulled across the pore lumen while a nanopore-associated motor proteins acts as brake on DNA by ratcheting the bases one-by-one into the reading head. As other molecular component, membrane-anchored adapters guide target DNA strands towards the membrane nanopore for enhanced sensitivity, while arrays of nanopores increase throughput. The nanopore method also directly reads methylated cytosine and other chemically altered bases of biomedical relevance[31,32]. The nanopore read-out is converted into the final sequence information using base-calling algorithms and software[24,33]. Nanopore sequencing is part of a larger and dynamic next-generation sequencing market[18] (Box 1).

Nanopore DNA and RNA sequencing has inspired the analogous label-free analysis of proteins[34–38]. In reflection of the expected wide-ranging benefits for research, biomedicine, and diagnostics[39], nanopore and other single-molecule protein sequencing approaches have been named *Nature*'s seven technologies to watch in 2023[40]. Nanopore sequencing of polypeptides is, however, more challenging than reading DNA and RNA strands. Firstly, protein sequencing is analytically more complex with twenty canonical as well as hundreds of non-proteinogenic amino acids, including many post-translational modifications to distinguish; DNA sequencing usually reads just four nucleobases. Furthermore, complex folding of proteins hinders the controlled ratcheting of polypeptide chains through nanopores while threading is more straight-forward for the structurally simple DNA strands. Another issue is the lack of protein amplification methods analogous to PCR for nucleic acids. Finally, the dynamic range of protein copy numbers can span

from 1 to 10^7 copies per cell[41] which is much higher than for DNA and RNA.

Several of the issues faced by polypeptide sequencing may be addressed by exploiting solutions developed for DNA nanopore sequencing. However, comprehensive molecular information on the most widely used forms of DNA nanopore sequencing is not accessible, even though excellent reviews focus on the separate aspects of sequencing applications[24,42] and base-calling algorithms and software[24,33]. Understanding the molecular underpinnings of nanopore DNA sequencing is non-trivial as this technology integrates various research areas such as nanopore design, motor protein engineering, membrane polymer chemistry, and nucleic acids chemistry. The interdisciplinary nature behind DNA sequencing and the fabrication of sequencing devices into commercial products is where companies, which are often well-resourced, have a lead over academic research groups who contributed the important preliminary discoveries. Many of the important advancements regarding the sequencing and device fabrication are not made publicly available as sequencing companies do not often publish in peer-reviewed journals[43]. However, companies reveal aspects of their technology through patents, as well as on websites and at conferences.

In this Review, we describe the molecular underpinnings of commercial nanopore DNA sequencing and ask how this relates to advances in nanopore protein sequencing[34,36,37,44] by taking into account principal similarities and differences between the two. Specifically, we describe the molecular components and detailed chemistry of the most widely used nanopore-based DNA sequencing device, the MinION from Oxford Nanopore Technologies (ONT). We then systematically compare this to other commercially available nucleic acid sequencers as well as related academic research, which may inspire protein nanopore sequencing. We finally discuss the state-of-the-art in nanopore-based peptide sequencing and possible advances based on learning from DNA sequencing.

Nanopores and motor proteins for DNA sequencing

Protein Nanopores

Protein pores form the basis of nanopore-based polynucleotide sequencing[43] as their biological function lends them to threading DNA. In biology, protein pores span lipid bilayer membranes with a hollow lumen to permit controlled passage of ions, water and small molecules. The narrow lumen of several pores at around 1 nm wide is wide for a single but not two DNA strands to pass. Protein pores also have a

structurally stable and atomistic defined scaffolds which helps obtain a high-quality signal in DNA sequencing. The protein scaffold does not vary greatly from pore to pore given their biogenic fabrication via reproducible protein folding[45] which is an advantages over synthetic pores fabricated in inorganic membranes[46–53]. Protein engineering can furthermore precisely tune size, charge, hydrophobicity, polarity, and other properties important for high pore stability and sequencing signal-to-noise ratio[45,54–56]. Among many different pores, bacterial membrane-puncturing cytotoxins are well suited for sequencing due to their constitutively open lumen structures. Two cytotoxins, α-haemolysin from *Staphylococcus aureus*[57]and MspA from *Mycobacterium smegmatis*[58], have paved the way for commercial nanopore-based DNA sequencing[59–61] by demonstrating single-nucleotide resolution after targeted protein engineering[59–63].

Commercial nanopore-sequencer from ONT use another bacterial protein, the Curlin sigma S-dependent growth subunit G (CsgG) pore from *Escherichia coli*. In biology, CsgG shuttles curlin subunits out of the cells to form extracellular fibres[64]. CsgG is a β-barrel pore made up of 36 strands from nine subunits[65] (Figure 1b). The up to 4 nm wide channel features a centrally located constriction of ~1 nm in diameter[65,66], which is used as the reading head for sequencing (Figure 1b). The reading head of CsG is shorter compared to α-haemolysin. For readout by CsgG, up to five consecutive nucleotides within or close to the reading head contribute to ionic current blocks whereby the base closest to the head has the biggest influence (Figure 1b and c). This complicates read-out compared to current blockades of single nucleotides[59], but the complex signals are deciphered by software algorithms based on machine learning[24,33,67,68]. To improve the sequencing performance of CsgG, ONT have screened over 4000 mutations[69]. The key pore regions are provided in Box 2 and Figure 2 (ref.[69]). ONT refers to the CsgG pore as R9 technology.

The pore version used in current ONT devices, R10, improves upon R9 by significantly higher sequencing accuracies of up to 99.3% (ref.[70]). Little information has been made available regarding the specific structure of the R10 pore by the company. However, a concurrently published research paper presents a CsgG-CsgF fusion pore with a dual-constriction[70] (Figure 1b). The second constrictions is formed by the accessory protein CsgF (ref.[70,71]) and has an inner diameter of 15 Å (ref.[70]) (Figure 1b). The two spatially distinct constrictions allow for reading the same nucleotide sequence twice, which greatly improves read accuracy in comparison to the single-constriction CsgG pore. This is particularly important for homopolymeric polynucleotide stretches that are difficult to decipher with a single reading head[70]. The read-out signals using the CsgG-CsgF pore are more complex than the single-reading-head CsgG, hence

adapted algorithms for signal deciphering are used[72].

Protein Helicases

Nanopore-associated motor proteins are pivotal in nanopore sequencing by ratcheting ssDNA across the reading head one nucleobase at a time[60,61]. By doing so, they act as molecular breaks to slow down DNA passage. Without breaks, the electrophoresis would drive strands at 1.5 – 10 µs per nucleotide through the pore[73,74], which is too fast for accurate reading. Initial experiments on motor proteins used the DNA polymerase phi29[60,61]. Sequencing is nowadays conducted with other motor proteins, helicases. Unlike polymerases, helicases reversibly unwind DNA duplexes into two separate component single strands without polymerisation of nucleotides. Furthermore, helicases lack exonuclease activity which can cause backstepping of the DNA in the pore[75]. As other advantage, helicases bind at multiple positions along the nucleic acid strands[76] (Figure 2d,e). This is important for long sequencing reads where dissociation of the active helicase can be back-up by another, already bound helicase.

The MinION from ONT features mutated versions of the ATP-driven helicase motor protein Hel308 (ref.[77]). Hel308 contains a motor core composed of two RecA-like domains which bind to DNA along five consecutive nucleotides[76] (Figure 2d). ATP binding and hydrolysis between the RecA domains changes their relative distances and thereby causes the helicase to move in two steps along the DNA by one nucleotide[78]. The movement is repeated upon binding fresh ATP. The motor-aided DNA movement aids sequencing with nanopores, but can also be used to explore the biophysical properties of the motor itself[78]. By measuring the miniscule motor forces via the electrophoretic pull on DNA, Hel308 substates were resolved at millisecond resolution. This informed how helicases of superfamily 1 and 2 convert ATP hydrolysis into motion[78].

Electrical potential drives DNA translocation

In nanopore sequencing, the voltage set up across the membrane is key for several reasons. First, the electric field set up between the two membrane sides electrophoretically threads the negatively charged DNA strands into the pore, until the bound helicase stalls at the pore entrance[79]. The potential secondly drives the transport of DNA through the pore to allow sequencing sequential bases in the current MinION read-out mode. Voltage thirdly controls the electrophoretic movement of small electrolyte ions such as K^+ and Cl^-, which carry the ionic read-out pore current. As a fourth factor, the membrane potential's force

stretches the DNA during pore translocation between the helicase DNA-binding sites and the nanopore reading-head[79–81]. For example, increasing the applied voltage from 100 to 200 mV straightens DNA by one nucleotide in length per nine-nucleotides[81]. Hence, high membrane potentials can pull the DNA from the helicase's grip whilst low voltages can cause back-slipping of the DNA strand leading to undesired re-reading of bases as well as low electrolyte transport and poor signals. To balance between these effects, the optimum voltage is carefully chosen depending on the motor protein and typically lies at 180 mV for the R9 technology[77].

Polynucleotide adapters for DNA sequencing

Polynucleotide adapters are short oligonucleotide structures that ligate to analyte DNA[82] and improve sequencing in two ways[83]. As adapters carry membrane anchors, analyte strands are concentrated at the membrane interface[77] (Figure 3a-c) to increase capturing and threading of nucleic acid into nanopores more than a thousand-fold[82]. Adapters are also preloaded with motor protein and hence facilitate sequencing further[82] (Figure 3a). The polynucleotide adaptors have been used by ONT in three sequencing versions, termed 1D, 2D and $1D^2$. In 1D sequencing, each strand is ligated with an adapter and sequenced independently (Figure 3a). 2D sequencing links both strands of DNA duplex with a hairpin loop, such that both template and complement strands are sequenced sequentially. $1D^2$ does not use hairpin loops but subsequently sequences template and complement of a membrane-tethered duplex. 2D and $1D^2$ sequencing are no longer supported by ONT and instead accuracy for 1D sequencing has been improved by base-calling algorithms.

Polynucleotide adapters for RNA sequencing

Nanopore technology has also been tailored to sequence RNA and complementary DNA (cDNA). For the latter, reverse transcription of RNA leads to full-length cDNAs that are modified with adapters[84] (Figure 3b). This allows for sequencing the transcriptome of single cells[85]. The alterative sequencing of native RNA sequencing avoids biases from cDNA synthesis or PCR amplification[86,87] while retaining important epigenetic information[84]. To prepare the sequencing library, native RNA strands are first annealed and ligated at their the polyA tail to a duplex adapter with a complementary oligo(dT) overhang. Subsequent ligation to a second adapter pre-loading with motor proteins initiates sequencing (Figure 3b). Direct sequencing of native RNA is typically less accurate than DNA sequencing, achieving average accuracies of 83-86% (ref.[88,89]). This is likely due to the propensity of RNA to form secondary structures, and a less

efficient motor protein for controlled ratcheting of native RNA.

Nucleic acid movement schemes

A combination of membrane potential and motor protein-actuated movement determines the directional threading of nucleic acids during nanopore sequencing. In general, single-stranded DNA and RNA can enter the pore with the 5′ or 3′ terminus. Current kits from ONT use 5′ threading of DNA and 3′ of RNA for electrophoretic threading through the pore (Figure 3c, Inny; 3d) whereby the motor protein acts as a molecular brake. In the other mode, sequencing occurs when a motor protein type pulls the strand against the opposing electrophoretic force out of the pore (Figure 3c, Outy; 3e). In this mode, the speed of DNA translocation can be tightly controlled by adjusting the membrane potential, and the DNA strand can be kept at the nanopore until ejected. This helps size the strands before sequencing, as the electrical readout duration relates to the length of the DNA strand. Furthermore, native DNA strands can be re-read multiple times to improve read accuracy.

Resilient polymer membranes

The membrane patches are another key component of sequencing platforms as they function as an embedding layer for nanopores and adaptors, and as steric and electric barrier to set up electrophoretic transport of DNA and ionic transport for current read-out. Planar lipid membranes are widely used in nanopore recordings[90,91], but have inherent issues relating to oxidation, hydrolysis and mechanical collapse[92]. This makes bilayers unsuitable for manufacturing commercial nanopore sequencing devices where membrane stability is key for distribution and ease-of-use. The stability of membranes can be increased[93] by reducing the lateral bilayer size[94–96], using hydrogels or inorganic supports[97,98], or by replacing biological lipids with polymerizable lipids[99,100] or synthetic amphiphilic non-lipid molecules[101,102]. MinION devices make use of the synthetic polymer membranes of higher mechanical and thermal stablility[103]. Although ONT does not disclose the exact chemical composition, polymer membranes have previously been formed of synthetic amphiphilic block copolymers[103,104]. In block copolymers, two or more oligomeric units are linked into a single polymer chain[104] as in triblock copolymer $PMOXA_7$-$PDMS_{60}$-$PMOXA_7$ with two poly(2-methyloxazoline) (PMOXA) units flanking the central poly(dimethylsiloxane) (PDMS) unit[104]. In this membrane-forming copolymer, two terminal hydrophilic PMOXA units and a hydrophobic PDMS core mimic the amphiphilic structure of phospholipid bilayers featuring two water-exposed polar headgroups between the hydrophobic fatty acids. As advantage over biology, the exact

composition of synthetic membranes can be carefully tuned with regard to chain length and other properties important for pore interaction. Copolymer membranes also have a higher stability over an extended pH range, as well as against mechanical stress and high temperatures, as noted[103,104].

Device hardware

The sequencing components in the MinION device are contained within membrane-covered and electrolyte-filled nanowells (Figure 1a). A MinION flow-cell features a total of 2048 nanowells, each capable of running a separate nanopore DNA sequencing reaction. As not all of the membrane patches carry an nanopore due to their stochastic insertion, the current blockades stemming from up to four nanowells are recorded[33]. During DNA sequencing runs, fluctuations in nanopore currents are measured several thousand times per second by sensors in the base of the flow cell. Data streams are subsequently sent to a microchip developed by ONT known as the Application Specific Integrated Circuits (ASICs), which are optimized for nanopore sequencing[33]. The ASICs perform real-time signal processing, base-calling, and data compression, enabling the MinION device to produce high-quality sequence data with minimal hardware requirements. The ASIC is specifically designed to handle the complex data processing required for nanopore sequencing. Each nanopore is measured and controlled individually by the ASIC, meaning multiple sequencing experiments can be performed in parallel. The MinION also contains components relating to temperature sensing and regulation, field-programmable gate arrays, and electrical shielding.

Data analysis software

In addition to hardware components, bioinformatical software tools for the analysis of DNA readouts are an important for commercial nanopore sequencing. Commercial sequencing devices do not require high computing power for data analysis, and users are able to collect and analyse data themselves[24]. As well as many company software tools, third-party programmes have been developed to expand the scope of data analysis. Current software packages are available for quality control[105–107], processing and visualisation[108–110], improved analysis times[111], detection of DNA[112–114] and RNA[115–117] modifications, error corrections[118–120], genome alignment and assembly[121–123] and analysis of repetitive elements[124,125]. An in-depth survey of the software is beyond the scope of this review[126].

Academic lead and technological maturation

The success of nanopore DNA sequencing highlights the importance of initial academic breakthroughs and their technological maturation into a commercial device that synergistically integrates other technologies. The first breakthrough came in the 1980s by proving that nanopores can characterise individual DNA and RNA strands, even though at high strand translocation speeds prohibiting single base resolution[73]. The second breakthrough was to engineer a reading head into a pore to distinguish all four individual nucleotides yet not in strands[59]. The third breakthrough was to use motor proteins to control the speed and orientation of DNA movement, thereby increasing the resolution of oligonucleotide sequencing[60,61]. These lab-based techniques were turned into a commercial product by crucial hardware advances including high-precision-low-noise amplifiers for the sensitive detection of electrical base signals, and improved techniques for fabricating uniform membrane-embedded nanopores and consistent blockade levels, as well as software developments to analyse the electrical pore blockade signals for the rapid and accurate sequencing of DNA.

Alternative nanopore-based sequencing techniques

Sequencing by ONT is one important but not the only technique to analyse DNA and RNA with nanopores. Like ONT, the other techniques exploit the small opening of nanopores to capture and detect individual strands. Yet they use other biogenic or synthetic nanopore, and obtain sequence information via routes different to the directional DNA threading and electronic base recognition. Here we briefly describe the principles of some alternative nanopore sequencing technologies, which have been developed by industrial and academic research teams (Figure 4). Several of the technologies are commercially available while others are still at the concept stage. While our overview is comprehensive, other companies and technologies may exist.

Quantapore

Nanopore sequencing from Quantapore claims long kilobase reads, with easy, amplification-free sample preparation (Figure 4a). This technology is expected to be scalable, with cartridges containing up to hundreds of thousands of pores, whilst maintaining low costs. Early versions of this technology were based on fluorescently labelled polynucleotides which are obtained by template-direct synthesis and subsequently forced into a constraint conformation inside the nanopore to self-quench fluorescence[127–

[129]. Upon exiting the nanopore, the fluorescence of the dye-labelled nucleotides is unquenched base-by-base and optically detected to reveal the DNA sequences. A recent patent[130] describes a different route with arrays of solid-state nanopores to isolate single fluorescently labelled DNA tethered to carrier particles (Figure 4a). An exonuclease cleaves off from the distal end of the DNA strand stepwise bases, which are detected as they diffuse through the fluorescence excitation zone, one nucleobase at a time. The temporal single-molecule fluorescent readout enables to infer the polynucleotide sequence for each pore in an array[130].

NobleGen Biosciences

An alternative fluorescence-based nanopore sequencing technology has been proposed by NobleGen Biosciences, named the Optipore system[131] (Figure 4b). Here, each nucleotide of a target ssDNA is enzymatically converted into a binary code sequence within a designed DNA polymer. Each base is transcribed into two short DNA stretches, where each stretch relates to either '0' or '1' (Figure 4b). Each binary sequence is recognized by a molecular beacon that contains on one strand end a fluorophore specific for each binary sequence and at the other end a universal quencher. After hybridization, the neighbouring universal quenchers suppress fluorescence emission from all molecular beacons except the leading one. Upon threading single strands through a solid-state nanopore, beacons are stripped off, which leads to a reversal of quenching and a series of detectable photon bursts, characteristic of the binary sequence and the original DNA sequence. Although it is complicated to generate the final read-out strands from the initial DNA strand, the signals of multiple nanopores can be detected simultaneously, which would allow parallelization to larger nanopore arrays and high-throughput analysis[131,132].

Genia Technologies

The technology developed by Genia Technologies[133] and subsequently acquired by Roche is based on nanotag-based real-time sequencing by synthesis (NanoTag-SBS). In the route, a polymerase close to the entrance of a protein nanopore, typically α-haemolysin, catalyses the synthesis of a strand complementary to an individual single-stranded analyte DNA strand (Figure 4c) using triphosphate nucleotides carrying gamma-phosphate tagged poly(ethylene glycol) (PEG) chains of different size that encode for a base. Nucleotide incorporation cleaves off the diphosphate-PEG tags which block the pore lumen to a degree characteristic for each of the four tags and bases. The time-dependent occurrence of blockades yields sequence information[133]. Similar to all other nanopore sequencing methods, the read-out is conducted in a highly parallel fashion for arrays of nanopores.

Quantum Biosystems

The sequencing by electronic tunnelling (SBET) underpins the technology from Quantum Biosystems[134] (Figure 4d). This DNA sequencing technology is based on the principle that each of the four bases has a distinct, characteristic structure and electron distribution that will specifically perturb the tunnelling current when individual DNA strands are translocated between a pair of nanoelectrode tips separated by a gap of 1-2 nm[135–138]. The current fluctuations from electron tunnelling differ from ionic pore currents which are carried by electrophoretically driven solvated electrolytes in classical nanopore sequencing. SBET technology has been used to detect short ssDNA and RNA fragments[139].

Nooma Bio

The technology being developed by Nooma Bio uses a dual-nanopore system[140–144] (Figure 4e) where the motion of an individual DNA strand through the two pores can be tightly controlled by regulating the potential at each electrode to increase resolution. This method has been used to map a bacterial genome by estimating distances between tags incorporated at sequence motifs[144]. Single-nucleotide resolution has not been achieved, but simulations suggest this might be possible by repeatedly flossing DNA or RNA strands through the pores[140].

Nabsys

High-definition (HD) mapping from Nabsys uses electronic mapping with solid-state nanodetectors to analyse long-length DNA[145] (Figure 4e). Here, whole genome maps are created by detecting sequence-specific tags that are incorporated along the DNA strands. Tagged molecules are translocated through a nanodetector and the position of the tags is determined electronically as a change in the resistance of the detector. Tag positions are subsequently analysed by the software. HD-mapping is commercially available and used in genomics mapping[146], characterization[147] and screening[148], as well as in structural variant verification[149]. Another company, Cambridge Nucleomics, uses nanopores for direct quantification of native RNA. It uses nanostructured carriers for binding RNA strands[150,151]. The RNA-carrier complexes are threaded through solid-state nanopores for identification and quantification. INanoBio are another company that have received funding to produce a device based on CMOS semiconductor nanotechnology, that boasts 100 times faster sequencing than protein based nanopore devices[152]

Comparison of techniques

Each of these methods holds unique advantages and disadvantages with regards to classical ONT nanopore sequencing such as with regards to labelling which usually leads to inefficiencies and sample loss. The Quantapore systems utilise simple-to-produce and stable arrays of solid-state nanopores, yet fluorescent tagging all four bases for kb-long DNA can be challenging to achieve, even though partial sequence information can be obtained when only two types of bases are labelled. Similarly, the Optipore system allows for highly parallel optical readouts from solid-state nanopore arrays, however, template DNA requires high degrees of processing to produce the binary sequences. In comparison, Genia uses a simpler nano-tag sequencing-by-synthesis route. But as each base-encoding tags is detected only once, readout accuracy is lower compared to direct strand sequencing where a single base can be read multiple times. Furthermore, base modifications such as methylation are lost in sequencing-by-synthesis. Strand sequencing by Quantum Biosystems' electron tunnelling directly reads nucleic acids including base modifications but the high sensitivity of electron tunnelling can be a limiting factor for commercialisation, as different DNA base configurations relative to the nanoelectrodes overlap in nucleobase readout[135]. The bases' degree of freedom may be restricting with tight solid-state nanopores[135]. Nooma Bio's two-pore system of strand flossing has been used for genome mapping[144] and maybe used for single-base sequencing of nucleic acids[140] but is currently not offered as product. Nabsys HD-mapping is a commercially available system for large-scale genome mapping but not single-base sequencing.

Several other routes are explored to sequence DNA. These include the use of atom-layer thin membranes composed of graphene, MoS_2 and related materials[46,153,154] which match the molecular dimension of nucleotides to allow better ionic current discrimination between different nucleotides[46] when compared to thicker membranes composed of SiN. The undesired adsorption of DNA to the hydrophobic pore membranes can be partly addressed by coating with passivating layers[155]. In another sequencing route, solid-state pores in thicker membranes carry proximal metallic nanostructures which plasmonically enhance the fluorescence readout of dye-tagged DNA strands[156,157]. Nanopores are also used to electrophoretically concentrate DNA strands into sequencing nanowells of the established PacBio's platform for fluorescence-based sequencing-by-synthesis[158]. In yet another approach, mass spectrometry is explored to detect individual nucleotides, which exit from a singulating nanopore cone[159,160], which would be compatible with exonuclease sequencing.

Clearly, nanopore-based DNA sequencing can be pursued with a wide variety of approaches. Different

pores of biogenic or synthetic origin can capture single strands, and sequencing can rely on directly reading strands base-by-base, by tracking the polymerization of a complementary strand via sequencing-by-synthesis, by deciphering of exonuclease-generated tags or nucleotides, or by mapping of tagged sections of DNA; the readout can be based on ionic current, tunnelling current, or fluorescence. Several of these approaches may guide future developments in protein sequencing.

Current advancements towards proteomic analysis and sequencing

The success of nanopore-based DNA sequencing has inspired research into the sequencing of peptides. If realized, portable, label-free, and fast protein sequencing could lead to considerable benefits in biological discovery, clinical diagnostics, homeland security, and food safety[44]. A crucial question is whether protein sequencing with nanopores can draw on the principles established for DNA sequencing (Figure 5a,b) and use the same MinION components such as polynucleotide adapters for efficient DNA capture, motor proteins and electrophoresis for controlled DNA threading, and designed reading heads for base recognition (Figure 5a). Compared to nucleic acids, sequencing of polypeptide chains via nanopores is a particularly challenging endeavour with three main issues relating to capturing intricately folded proteins to initiate unravelling, controlling polypeptide chain ratcheting through nanopores while unfolding the protein's secondary structures, and distinguishing the chemical complexity of 20 amino acids and many post-translational modifications (Figure 5b). In the following, we survey current research, which advances each of these key areas.

Capturing folded proteins

Protein capturing reveals molecular processes prior to polypeptide unravelling and informs on the transport mode of electroosmosis relevant for folded proteins as well as polypeptide threading. Capturing and sensing of globular proteins usually uses wider pores than used for sequencing the approximately 1 nm-wide elongated DNA strands[161,162]. Examples are biological pores PlyAB[163] or ClyA[164] of around 5 nm width that accommodate 35 kD-big proteins[165–168] and synthetic bilayer-embedded DNA nanopores of tuneable shape and size up to 20 nm (Figure 5e)[169,170] constructed with DNA origami nanotechnology[171–173]. The DNA pores enable detecting folded 150 kD-proteins that pass the lumen[169,174,175] or are temporarily held via an optional bioaffinity recognition tag installed at the pore (Figure 5e)[169]. Even wider up to 100 nm-large synthetic pores can be fabricated into solid-state membrane materials[46–53] using e-beam or ion

drilling methods, or simpler dielectric breakthrough using kits such as from Northern Nanopore[176], and glass capillaries[52,53,177–179] (Figure 5c,d). These solid-state pores are thermally, chemically, and mechanically highly stable but their detailed structures are usually not reliably reproduced in high numbers due to limitations in fabrication, despite progress[47,50,53]. Nevertheless, the solid-sate pores have helped sense proteins in a mixture, as well as characterize protein folding/unfolding, protein conformation changes, enzyme binding and shape approximation, as reviewed in[47,53,180,181].

Independent of the nanopore type and size, capturing and transport of folded protein usually proceeds via electroosmosis[182,183], which is of minor importance to DNA threading (Figure 5f). In electroosmosis, electrophoretically driven small electrolyte ions drag along the surrounding water shell, which in turn leads to the co-flux of proteins. Electroosmosis becomes particularly prevalent for negatively charged SiOx[182], glass[184], and DNA origami pores that attract a surface layer of cations[169,174,175]. Electroosmosis can also be induced by docking a DNA-origami nanosphere into a solid-state nanopore[183]. Electroosmosis can even cause the transport of charged proteins against the electrophoretic force towards an electrode of the same charge[182]. If desired, electrophoresis can be instated to dominate electroosmosis by adhering charged tags to the protein[185] or by tuning buffer pH and solvent conditions as exemplarily shown for the analysis of hyaluronan chains[186]. Irrespective of transport mode, proteins pass large-diameter nanopores often too fast for detection with classical current amplifiers[187] even though new-generation devices address this limitation[188]. Alternatively, proteins can be slowed down using nanopores with macromolecular crowding[189], charged lumen walls[175,182], coatings for analyte adsorption[48], defined molecular recognition tags[169], DNA nanospheres that sterically block transport[183], elongated DNA nanocarriers that bind proteins and thread the complex through the solid-state pores for multiplex sensing[179,190,191], or macromolecular crowding[189]. Some of the approaches allow detecting the wide range of protein shapes and volumes as well as orientations in which proteins enter the pore[48,183].

Peptide threading

Strand threading via electrophoresis and motor proteins, as successfully used for DNA, can also be used for peptides. To activate the first transport mode, very high voltages can electronically drive unfolding of some native proteins[192]. Alternatively, polypeptides of heterogenous charge can be coated with charged denaturants[185] to achieve electrophoresis similar to homogeneously charged short peptides[193]. Electrophoresis also proceeds by coupling a negatively charged oligonucleotide to single-domain proteins to unfold and thread[194] even though individual amino acids were not resolved. Coupling a peptide to an

electrophoretically active DNA strands provides a basis for protein unfolding and nanopore threading[41,195–197] and has indeed been key to attain single-amino acid resolution[34] as discussed in the following section on high resolution.

The second transport mode via molecular motors can unfold and thread polypeptides through a nanopore. In pioneering work, molecular motor ClpX controlled the ATP-dependent pulling of a multidomain protein across the α-haemolysin pore[36,198]. The ensuing current block depended on which structural domains resisted unfolding and translocation. To gather more control, a hybrid nanopore capable of protein unfolding and cleavage was genetically engineered[37] (Figure 6a). The 900 kDa nanopore sensor consisted of an unfoldase enzyme and the 20S proteasome placed on top of the protein pore[37]. Analyte proteins could be sensed in two modes, thread-and-read with the proteasome inactive, or chop-and-drop with an active proteasome. The molecular machine did not provide full resolution of individual amino acids, yet may be used for fingerprinting proteins. In nanopore fingerprinting, individual peptide molecules with unique amino acid sequence or specific motifs are electronically detected, often coupled with machine learning algorithms for accurate identification[199,200]. For example, small peptides were obtained by trypsin-mediated fragmentation of analyte proteins and subsequently detected upon passing pores such as aerolysin[201] (Figure 6b) or FraC[202].

Fingerprinting is also possible by unfolding and electroosmotic threading entire proteins. The charged denaturant guanidium hydrochloride played a key role by unfolding proteins additionally adhering to the pore wall to induce electroosmosis[203]. Threading through α-haemolysin was initiated by a negatively charged peptide tags fused to a protein end. Notoriously difficult to unfold β-barrel proteins in a mixture were distinguished using machine learning at accuracies of more than 90%. The relatively slow polypeptide translocation of 10 μs per residue may be sufficient for detecting sequence features, once a pore with a shorter reading head is used. Other fingerprinting approaches involve chemical tags on a dipolar peptide[204], fluorescent tagging of proteins[200,205], and unique peptide tags that can be fused to proteins to allow electrical sensing with commercial nanopore DNA sequencing kits[206]. Sequential groups of amino acid in polypeptides give rise to different blockade currents in solid-state nanopore[207]. Furthermore, individual posttranslational modifications within a peptide sequence are detectable via ionic current[208,209] or tunnelling[210]. The above methods unfold and thread polypeptides, but lack single amino acid resolution for sequencing.

Single amino acid resolution

Analogous to DNA[59], early efforts to read peptides focused on detecting individual amino acids. For example, tunnelling electrodes covered by recognition molecules[211] were able to non-covalently bound target amino acids and yield unique and characteristic tunnelling currents. In another study, almost all amino acids were discriminated by ionic current perturbations[38]. The target amino acids were tethered to a short peptide carrier which translocated through the aerolysin pore (Figure 6c). The translocation speed was slow enough to discriminate even structurally similar leucine and isoleucine. This method may be used for sequencing once a scheme is devised to cleave off amino acids from polypeptides and ligate them to carrier peptides for sequential identification. Aerolysin also differentiated peptides carrying a single posttranslational modification in alternating positions[212].

In recent work, amino acid residues in a peptide were resolved, strikingly by using commercial DNA sequencing technology[34]. In the approach, the peptide to be analysed was linked to a DNA strand, and the DNA-helicase Hel308 threaded the two-component strand through the MspA pore (Figure 6e)[60,213]. The helicase moved the peptide component in steps close to the pitch of a β strand so that single amino acid were resolved[34,213,214]. Re-reading the same peptide yielded consensus sequencing accuracies of >99.99%[34]. This approach presents a step-change towards peptide sequencing[215,216] even though it is currently limited to peptides <25 amino acids. Furthermore, highly positively charged peptide may electrophoretically move against the direction of DNA threading. This may be addressed by using a pore variant with strong electroosmosis to overrule electrophoresis.

Non-nanopore-based protein sequencing

Although nanopores have the potential for high-resolution protein sequencing, non-nanopore techniques including Edman degradation and mass spectrometry have been widely used for protein and peptide sequencing for several decades. Recent improvements to these and other methods have increased throughput and sensitivity for use in single-molecule sensing[39,217,218].

Edman degradation involves modifying an N-terminal amino acid and its subsequent cleavage for detection by liquid chromatography, repeated in cycles to read a sequence[219]. Highly parallel sequencing is not feasible as purified peptides are required. To address this, Edman degradation with fluorescence read-out was developed[220] by immobilising millions of protein fragments via their C-terminus onto a glass surface while labelling their N-terminal amino acids with distinct fluorophores. The protein sequences are

established in highly parallel fashion with fluorescence microscopy as N-terminal amino acids are stepwise removed via Edman degradation. In a related approach, fluorescently labelled recogniser molecules temporarily bind to N-terminal amino acids of millions of immobilised peptides. The accompanying fluorescent pulses are recorded, before an aminopeptidases removes the N-terminal amino acids and cycle is repeated to obtain the peptide sequence[221]. Faster run times for the sensing cycles will improve the method's throughput.

Single-molecule mass spectrometry builds on well-established mass spectrometry (MS) techniques. MS measures the mass-to-charge ratio of ions and is used in the probing of charged peptides and proteins. Charge detection MS (CDMS) approaches focus on the analysis of large biomolecular complexes in the range of 1-100 MDa[39]. The development of Orbitrap mass analysers[222–224] has seen increased resolution, directly deriving the charge states of single proteins and their fragment ions[225]. Such techniques have greatly expanded the approach to not only confirm the identity of proteins by matching sequences of peptide fragments to entries of a protein database, but also post-translational modifications and their locations in the protein sequence[225]. Limitations remain, however, as ionization of proteins and peptides is required for MS techniques. Not all peptides are ionized efficiently, which may limit the application of this technique for proteomic research. Other single-molecule approaches to protein sequencing are covered in[39] and Box 3.

Outlook of proteomic sequencing

This Review has highlighted how step-changing scientific innovations, integration of multiple molecular components, and continued improvements have created nanopore DNA sequencing. This winning formula led to an assembly accuracy exceeding 99.8%[23] and read lengths of up to 50 Mb[2] with direct sensing of epigentically modified DNA bases[31,32] using the MinION kit. Similar improvements are likely on the horizon for RNA sequencing to enhance eukaryotic transcriptome analysis and understand the molecular dysregulation leading to diseases including cancer[84,226]. Higher RNA single-read accuracy will likely come from adapted sequencing algorithms or changes in nucleic acid movement schemes. Advances are also expected from nanopore technologies which read out strands by modes other than ionic current, or which rely on exonuclease sequencing.

Nanopore-based peptide sequencing has seen striking improvements in controlled threading and single amino acid resolution. For example, existing DNA sequencing technology has been harnessed to sequence peptide-DNA conjugates[34,213] and assign post-translational phosphorylation to individual residues[227]. The technology currently uses MspA and will likely be implemented with MinION devices. Fingerprinting of proteins is an alternative nanopore technologies using protein fragments with unique signals[37,202] or entire proteins that are electroosmotically threaded[203]. The first approach is well suited for probing a subset of the proteome, such as phosphorylated peptide segments[227] that can optionally be enriched.

While these approaches are successful, challenges remain[41,44,215]. One is the sheer complexity of the proteome in terms of chemical diversity and dynamic range when compared to the genome or transcriptome. As a single eukaryotic cell can contain billions of proteins at highly varying levels, an extremely high-throughput and high-resolution would be required to profile even a fraction on a reasonable time-scale[39]. To cover the proteome as routinely done for the transcriptome, up to 30,000 more peptide reads would be required as performed by RNA sequencing[41]. This is a challenge for current nanopore-based sequencing and highlights the need to improve existing or develop new technologies[190,191] for massively parallel high through-put[228,229]. Increasing throughput is also the aim of next-generation LC-MS/MS-based techniques[228,229] and the SOMAscan assay by SomaLogic which converts the challenging tasks of protein quantification and identification into an easier job of oligonucleotide quantification via established next-generation sequencing[230].

As another challenge, nanopore-based proteomic analysis requires multiplexing from different samples and potentially single cells[228,229]. Single cell sequencing using droplet-bead technology in combination with Illumina read-out has revolutionised transcriptome analysis[231,232]. To multiplex protein samples for nanopores, the existing technologies for peptide-DNA conjugate sequencing could use the DNA sequence part to encode sample identity. Encoding via DNA sequences is already used for sensing of folded proteins with MinION readout[233]. This approach uses DNA coding strands that carry specific aptamers for proteins such that protein binding results into a unique and easy-to-decipher DNA sequence signal[233]. Multiplexing could also be attained by adding peptide tags to proteins[206] for read-out via peptide sequencing or fingerprinting.

Finally, threading of entire polypeptides could detect the full gamut of amino acid including their post-translational modifications. A challenge is the complexity of associated nanopore signals and stalling of

polypeptide threading due to very large modifications such as glycosylation. To overcome these obstacles, nanopore reading heads could be redesigned by utilising recent advances in de novo pore design[234,235] and protein structure prediction by artificial intelligence. A second head could be added analogous to DNA sequencing, and the pore lumen could be elongated to allow reading of longer peptide sequences. In a similar vein, redesigned motor protein could better unfold and thread peptides through pores in a controlled manner. Such motor proteins could be coupled with nanopore arrays, in which a range of tailor-designed nanopores could simultaneously sequence peptide chains of differing lengths, charges and post-translational modifications[216]. The expected technological advancements will benefit from collaboration of experts across multiple disciplines, which is a particular strength of the nanopore field. Innovation will also benefit from the stimulating effects of competition between multiple commercial participants, as templated by the related field of next-generation DNA sequencing.

The outlook for polypeptide identification and sequencing with nanopores is positive. Nanopores offer a broad spectrum of technologies including direct sequencing, fragmentation-followed-by-detection, and mapping of unique identifiers. Furthermore, different nanopore materials and read-out modalities are available. Unique strengths of nanopore sequencing and fingerprinting are the low entry cost of analysis, high portability while maintaining scalability and direct readout.

Acknowledgements The Howorka Group receives funding from the Human Frontiers Science Program (RGP0047/2020), the Engineering and Physical Sciences Research Council (EP/N009282/1, EP/V02874X/1), the Wellcome Institutional Strategic Support Fund, the Moorfields Biomedical Research Centre, and Oxford Nanopore Technologies Ltd.

Author contributions A.D. and S.H. contributed to discussions and wrote the manuscript.

Competing interests S.H. is named inventor on a patent on the CsgG nanopore which is licensed to Oxford Nanopore Technologies Ltd. S.H. and A. D. are named inventors on patents on DNA nanopores which are licensed to Oxford Nanopore Technologies Ltd.

Table 1: Nanopore-based protein fingerprinting and sequencing techniques

Technique/method	Analyte		Features					Application		Challenges to consider for future development
	Protein	Peptide	Unfolding	Threading	Protease digestion	Label-free	Strand rereading	Protein fingerprinting	Single-aa resolution	
Oligonucleotide tagging of folded protein	✓		✓	✓				✓		Particularly stable protein conformations may be difficult to unfold using this method. Protein charge may impact threading through the nanopore.
ClpX threading	✓		✓	✓	✓			✓		polyGSD-ssrA tag would need to be incorporated into endogenous proteins. Predicting protein ionic current patterns currently not possible.
Proteasome-coupled nanopore	✓	✓	✓	✓	✓	✓		✓		Small peptide fragments translocate too quickly through the pore for detection. Nanopore width may need reducing to improve the resolution of translocating strands and fragments.
Peptide fingerprinting using proteases	✓	✓	✓		✓	✓		✓		Requires access to large databases for accurate fingerprinting. Particularly stable or protease resistant protein conformations may be difficult to digest and fingerprint using this method.
Protein fingerprinting using guanidium chloride	✓	✓	✓	✓	✓			✓		High concentration of guanidium chloride may not be compatible with all biological nanopores. Translocation velocities may vary with each protein of interest. Insufficient resolution of posttranslational modifications.
Chemical tagging of a bipolar peptide		✓		✓				✓		Has the potential for use in protein fingerprinting, but has yet to be used experimentally to fingerprint folded protein analyte. Close proximity of the tags could interfere with detection.
Single amino acid sens. with polycationic carrier		✓		✓					✓	Technique is sensitive enough to distinguish between the twenty naturally occurring amino acids, however only when each amino acid is coupled to a carrier. Exonuclease approach and ligation to carrier peptide .
DNA-linked peptide threading		✓		✓			✓		✓	Currently only possible with peptide strands <20 aa long. Peptide strands with a high concentration of positively charged residues may not thread and sequence efficiently.

Figure 1. Protein nanopores used for portable and scalable DNA sequencing. **a,** CsgG nanopores (PDB: 4UV3, light blue) embedded in array of membrane patches of the hand-held MinION device. **b,** Comparison of the CsgG pore and CsgG-CsgF pore (PDB: 6SI7) inserted in a planar lipid bilayer (grey). The square brackets indicate the single and double reading heads of the CsgG and CsgG-CsgF pores, respectively. A bound helicase (pink) facilitates the nucleotide-by-nucleotide translocation of the single DNA strand (dark blue). The negatively charged (not shown) strand is also electrophoretically directed towards the positively charged membrane side (plus sign). Insets show the pore constrictions in CsgG without (top) and with (bottom) a translocating strand of DNA (dark blue). Without DNA, the ionic current is higher as electrolyte ions K^+ and Cl^- flow unimpeded through the pore (top inset). By contrast, a single DNA strand blocking the pore lumen lowers the flow of ion current depending on the base sequence in the reading head (bottom inset). The structures shown may differ from the pores used in the MinION kit. **c,** Example current readout trace for the sequencing of DNA with a protein nanopore.

Figure 2. The CsgG pore and key regions for improved nucleic acid sequencing. **a,** Side and top view of a CsgG nanopore with indicated molecular regions (red) for improved helicase interaction, polynucleotide capture, sequencing accuracy, and reducing undesired interactions with DNA/RNA. **b,** Helicase Hel308 (purple) bound to a fragment of DNA (blue) (PDB: 2P6R). **c,** Diagram showing the ATP-dependent movement of DNA through Hel308, facilitated by RecA-1, RecA-2, and ratchet domains. The molecular improvements achieved by all noted regions are described in greater detail in Box 2.

Figure 3. Nucleic acid sequencing methods used by sequencing devices from ONT. **a**, 1D sequencing involves sequencing of one DNA strand using membrane tethers and adapters preloaded with the Hel308 motor protein. **b**, RNA sequencing uses tailor-designed adapters and a dual membrane tether. **c**, Methods of nucleic acid threading and translocation. Plus and minus signs at the membrane indicate the polarity of the membrane potential. Arrows indicate the direction of nucleic acid strand translocation. **d,** "Inny" sequencing where a DNA strand binds to the membrane before helicase docking and sequencing. **e,** "Outy" sequencing where a DNA strand is first threaded through the nanopore before sequencing. Arrows indicate the direction of nucleic acid strand translocation. Image modified from ONT.

Figure 4. Alternative nanopore methods for nucleic acid sequencing. **a,** Quantapore[130] utilises a solid-state

nanopore to capture fluorescently labelled nucleic acid strands on a carrier bead. Upon exonuclease cleavage, individual nucleotides are detected which sequentially pass a fluorescence excitation zone to indicate the DNA sequence. **b,** The Optipore system from NobleGen Biosciences[132] converts each base of a polynucleotide sequence into a binary code within a new custom-designed DNA polymer. The binary code is recognised by complementary fluorophore-encoded beacons to form duplex segments. The beacons dissociate when the DNA polymers threading through a solid-state nanopore and give rise to fluorophore signals that reveal the original DNA sequence. **c,** NanoTag-SBS from Genia Technologies[133] uses a polymerase at a protein nanopore to synthesise a complementary DNA strand with tagged dNTPs. Upon nucleotide incorporation, pyrophosphate PEG nanotags are released which are detected upon passing the nanopore reading head to provide the DNA sequence. **d,** Sequencing by electronic tunnelling from Quantum Biosystems[135] uses nanoelectrodes to detect small differences in the chemical structure of passing bases in polynucleotides. **e,** Nooma Bio's dual-pore system[140–144] uses two nanopores that are independently controlled via electrodes to tightly control the nanoscale movement of DNA between the two pores. **F,** HD-mapping from Nabsys[145] employs a solid-state nanodetector to analyse tagged DNA for genome mapping.

Figure 5. Nanopore-based detection of proteins. **a**, **b**, Comparison of DNA and peptide sequencing highlighting the main differences in the analyte strands. **a**, A DNA strand (dark blue) is threaded through the CsgG pore (light blue) aided by a helicase (pink). The inset shows the reading head with each DNA base in a different colour. **b**, A protein (red) is unthreaded so that the polypeptide strand can thread the CsgG pore (light blue). The inset shows the reading head with each amino acid residue in a different colour. **c**, **d**, **e**, Large-diameter nanopores can be constructed from **c**, solid-state materials, **d**, a glass capillary, and **e**, DNA origami. **f**, Electrophoretic and electroosmotic forces determine the movement of charged proteins through charged nanopores. The image shows a solid-state nanopore (blue) with no, positive and negatives charges at the pore walls. The movement of positively charged proteins (red) is indicated for each of the three cases.

Figure 6. Protein pores used for peptide recognition and sequencing. **a**, A custom designed hybrid nanopore enzymatically unwinds a protein and cleaves the polypeptide chain into peptide fragments[37]. **b,** An aerolysin pore fingerprints peptide fragments generated from proteolytic digest of a target protein. **c**, An aerolysin pore detects target amino acids attached to a polycationic carrier[38]. **d**, An MspA pore used in

combination with Hel308 to sequence small length peptide chains bound to a molecule of DNA[34,35]. For simplicity, Hel308 dissociation and rebinding prior to resequencing is not shown.

Box 1: Commercial next-generation sequencing strategies different from nanopore sequencing

The largest share of the current sequencing market is taken by fluorescence-based sequencing-by-synthesis (SBS) of Illumina[236,237]. As an indirect sequencing approach, SBS copies the strand to be sequenced by stepwise incorporation of nucleotides via template-directed and enzyme-catalysed DNA polymerisation[238]. Each of the four nucleotides is uniquely fluorescently labelled so that the sequence can be inferred from the temporal order of fluorescence signals. The signal strength is enhanced by reading out nanoscale clusters of each identical strand bound to a glass slide. The readout is conducted in a massively parallel fashion for millions of DNA cluster with each different sequence. Illumina uses DNA fragments and the read length is up to 300 base pairs (bp). HiFi sequencing by PacBio[1] uses a related SBS principle, but analyses single molecules of DNA. PacBio accommodates non-fragmented strands for long-read sequences. Short-read sequencing is also possible[239] due to PacBio's acquisition of Omniome. Among newer fluorescence-based sequencing routes[18], MGI applies the massively parallel SBS approach to long single-stranded DNA that contains multiple concatenated copies of the sequence to be deciphered[240]. The DNA strands are formed by rolling circle amplification and condense into 300 nm-wide DNA nanoballs that adhere to glass slides. As each balls contains multiple identical DNA copies, multiple fluorophore-tagged nucleotides are simultaneously incorporated during SBS, and the fluorescence signals are enhanced. Elements' AVITI platform also uses SBS but replaces tagged nucleotides with fluorophore-tagged polymeric structures known as "avidites" whose tentacles terminate in single nucleotides that bind to the clustered strands to be sequenced[19]. The avidites only temporarily bind to the DNA strands for imaging, followed by being washed out, and the polymerisation-catalysed incorporation of a non-tagged nucleotide before a new cycle starts. The Ion Torrent sequencing technology from Thermo Fisher uses SBS and detects nucleotide incorporation not by fluorescence but from pH changes. The pH changes occur upon nucleotide incorporation when the phosphate tail is hydrolysed and are detected with a semiconductor device[20].

Box 2: Molecular regions within CsgG critical for enhanced nanopore sequencing

DNA interaction. The first step to improve the sequencing efficiency is to better capture polynucleotide strands at the pore. The binding of negatively charged DNA can be enhanced[241] by introducing positively charged amino acids at the CsgG channel entrance[69]. The opposite can also be help, by removing positive charges at non-channel positions to avoid misguiding DNA threading across the lumen and reading head[69].

DNA interactions can also be improved at the reading head. Wild type CsgG contains two close-by constrictions, one formed of tyrosine (Y) at position 51 and a second formed of asparagine (N) and phenylalanine (F) at positions 55 and 56 respectively (Figure 2b). Removal of either one of these constrictions reduces electrical signal complexity, whilst increasing the open channel current for increased resolution of nucleic acids[242]. A second reading head can aid sequencing, provided it is sufficiently separate from the first[243]. Undesired nucleic acid interactions are furthermore avoided through deletion of the trans-loop at the distal side of CsgG (Figure 2a), where the DNA strand exits the pore.

Helicase interaction. The nanopore-associated helicase facilitates controlled translocation of DNA strands one nucleobase at a time. To increase helicase interactions and subsequent improve sequencing accuracy, the cis-loop of CsgG[69] can be deleted leading to a reduce hydrophobicity and bulk mass at the pore's cis-end (Figure 2a).

Signal-to-noise. A high signal-to-noise ratio is important in achieving single-nucleotide resolution in sequencing. A lower current noise level was attained by mutating a lysine residue in a vestibule of the CsgG pores[69], thus increasing the resolution of single nucleotide signals.

Monomer biosynthesis. The biosynthesis of CsgG monomers has been genetically eased such as by mutationally reducing the number of arginine residue to result in higher transcription/translation rates[69].

Box 3: DNA-nanotechnology-based protein fingerprinting and sequencing techniques

DNA-PAINT

The DNA-PAINT (point accumulation in nanoscale topography) technique involves transient binding of

fluorescently labelled DNA strands to complementary DNA strands chemically tethered to a molecule of interest, such as a lysine or cysteine residues on a peptide sequence. The transient binding of the fluorescent DNA strands is detected as “blinking” in an intensity versus time trace. Protein characterization can be accomplished through a straightforward method that comprises counting amino acids using quantitative DNA-PAINT (qPAINT) where the blinking rate indicates the number of molecular targets. Highly efficient DNA labelling of specific amino acids, followed by qPAINT analysis, may achieve single-molecule protein fingerprinting for intact proteins[244]. DNA-PAINT techniques can also be combined with ultra-high resolution imaging for single-molecule protein fingerprinting[245]. Such techniques show promise with regards to proteomic analysis of complex protein mixtures and patterns of post-translational modifications, however the low binding rate of DNA means obtaining fluorescence information with high spatial resolutions can be time consuming.

DNA proximity recording

In DNA proximity recordings, protein identification involves attaching DNA probes to specific amino acids on a peptide chain. DNA amplification between probes in close proximity generates a DNA database of amplicons using a technique called autocycling proximity recording (APR). The amplicons differ in length and abundance in accordance with the associated distances between the amino acid pairs[245]. The database can be analysed to decode the distances between the DNA tags, which in turn can be used as unique identifiers for single-molecule protein identification[245]. DNA proximity recordings make use of next-generation DNA sequencing methods for protein identification and could provide a useful method for identification of proteins in complex mixtures.

DNA-based FRET

DNA-based fluorescence resonance energy transfer (FRET) measurements can be used to determine global pairwise distance measurements[246]. Specific amino acids such as lysine and cysteine are labelled with “docking” DNA strands complementary to DNA strands carrying a donor fluorophore. A fixed position on the protein is labelled with an acceptor fluorophore, such as the N or C terminus. A FRET histogram can be built based on the associated fluorescent intensities, which contains information on the position of each amino acid in relation to the reference point. The smaller the distance between the fluorophore pairs, the more intense the FRET signal will be. This FRET information can be compared to existing FRET information in a database for protein identification.

CITE-seq

CITE-seq (cellular indexing of Transcriptomes and Epitopes by Sequencing) enables simultaneous measurement of both gene expression and cell surface protein markers at the single-cell level. It combines single-cell RNA sequencing (scRNA-seq) with antibody-based protein detection, to providing insight into the molecular and phenotypic characteristics of individual cells within a complex biological sample. In CITE-seq, cells are first labelled with oligonucleotide-conjugated antibodies, specific to cell-surface proteins of interest[247]. These antibodies serve as barcodes that uniquely identify each cell. Next, the cells are processed for scRNA-seq, where the RNA content of each cell is captured, converted to cDNA, and sequenced using NGS techniques. During the sequencing step, the antibody barcode information is also retained, allowing for the correlation of gene expression data with protein expression profiles. This technique is used to quantify both cell-surface proteins and transcriptomic data within a single-cell readout. This technique has broad applications in various fields, including immunology, cancer research, and developmental biology, providing a powerful tool for understanding complex biological systems at the single-cell level, and as such has led to multiple important medical discoveries[248,249].

Single-cell PEA

The Single-cell Proximity Extension Assay (Single-cell PEA) technique can be used to analyse protein expression at the single-cell level. Here, single cells are sorted and lysed in microtiter plates[250]. Each cell is then subjected to PEA, where pairs of oligonucleotide-labelled antibodies are used to recognise and bind to specific protein targets of interest. When the paired antibodies are bound in close proximity, their complementary oligonucleotides are brought together, creating a DNA reporter molecule. The reporters from all cells are subsequently pooled and amplified for analysis. The amplified DNA is then quantified using techniques such as qPCR or NGS. By measuring the amount of amplified DNA corresponding to each protein target, researchers can determine the protein expression levels in individual cells.

PLAYR

The proximity ligation assay for RNA (PLAYR) technique is used to investigate the spatial organisation of RNA molecules within a cell. It combines the principles of proximity ligation and fluorescence in situ hybridization (FISH) to enable the visualisation and analysis of RNA interactions and proximity at the subcellular level. The assays involve fixing, permeabilising and labelling cells with fluorescently labelled antibodies[251]. A pair of RNA probes target proximal regions on target RNA molecules. After hybridisation, the RNA molecules are ligated together using enzymatic reactions, creating circular DNA molecules. These

circular DNA molecules serve as templates for amplification, where they are subjected to rolling circle amplification (RCA). The RCA process generates long, branched DNA structures that are labelled with fluorescent probes, which can be detected by flow cytometry. Protein detection involves binding of fluorescently labelled antibodies, by which protein and RNA detection can occur simultaneously.

References

1. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614 (2020).
2. Nurk, S. *et al.* The complete sequence of a human genome. *Science.* **376**, 44–53 (2022).
3. Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **37**, 783–792 (2019).
4. Deng, X. *et al.* Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* **5**, 443–454 (2020).
5. Zhou, B. *et al.* SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature.* **592**, 122–127 (2021).
6. Cao, X. *et al.* Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nat. Commun.* **12**, 508 (2021).
7. Liu, Y. *et al.* Proteomic profiling of HIV-1 infection of human CD4+ T cells identifies PSGL-1 as an HIV restriction factor. *Nat. Microbiol.* **4**, 813–825 (2019).
8. Leggett, R. M. *et al.* Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat. Microbiol.* **5**, 430–442 (2020).
9. Li, D. *et al.* Genomic profiling informs diagnoses and treatment in vascular anomalies. *Nat. Med.* **29**, 1530–1539 (2023).
10. Macken, W. L., Vandrovcova, J., Hanna, M. G. & Pitceathly, R. D. S. Applying genomic and transcriptomic advances to mitochondrial medicine. *Nat. Rev. Neurol.* **17**, 215–230 (2021).
11. Sero, D. *et al.* Facial recognition from DNA using face-to-DNA classifiers. *Nat. Commun.* **10**, 2557 (2019).
12. Cornelis, S., Gansemans, Y., Deleye, L., Deforce, D. & Van Nieuwerburgh, F. Forensic SNP Genotyping using Nanopore MinION Sequencing. *Sci. Rep.* **7**, 41759 (2017).
13. Liao, J. *et al.* Nationwide genomic atlas of soil-dwelling Listeria reveals effects of selection and population ecology on pangenome evolution. *Nat. Microbiol.* **6**, 1021–1030 (2021).
14. Cuypers, W. L. *et al.* A global genomic analysis of Salmonella Concord reveals lineages with high antimicrobial resistance in Ethiopia. *Nat. Commun.* **14**, 3517 (2023).
15. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science.* **323**, 133–138 (2009).
16. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator

chemistry. *Nature.* **456**, 53–59 (2008).

17. Booth, M. J. DNA and RNA Sequencing. in *Nucleic Acids in Chemistry and Biology* (eds. Blackburn, G. M., Egli, M., Gait, M. J. & Watts, J. K.) (The Royal Society of Chemistry, 2022).
18. Eisenstein, M. Illumina faces short-read rivals. *Nat. Biotechnol.* **41**, 3–5 (2023).
19. Arslan, S. *et al.* Sequencing by avidity enables high accuracy with low reagent consumption. *Nat. Biotechnol.* (2023). doi:https://doi.org/10.1038/s41587-023-01750-7
20. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature.* **475**, 348–352 (2011).
21. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* **122**, 1–15 (2018).
22. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature.* **530**, 228–232 (2016).
23. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).
24. Wang, Y. Y., Zhao, Y., Bollas, A., Wang, Y. Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
25. Marx, V. Method of the year: long-read sequencing. *Nat. Methods.* **20**, 6–11 (2023).
26. Deamer, D., Akeson, M., Branton, D., Kasianowicz, J. J. & Bezrukov, S. M. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
27. Wanunu, M. Nanopores: A journey towards DNA sequencing. *Phys. Life Rev.* **9**, 125–158 (2012).
28. Alper, J. From the bioweapons trenches, new tools for battling microbes. *Science.* **284**, 1754–1755 (1999).
29. Kasianowicz, J. J. & Bezrukov, S. M. On ‘three decades of nanopore sequencing’. *Nat. Biotechnol.* **34**, 481–482 (2016).
30. Bayley, H. Nanopore sequencing: From imagination to reality. *Clin. Chem.* **61**, 25–31 (2015).
31. Liu, Y. *et al.* DNA methylation-calling tools for Oxford Nanopore sequencing : a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 (2021).
32. Yuen, Z. W. *et al.* Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat. Commun.* **12**, 3438 (2021).
33. Lu, H., Giordano, F. & Ning, Z. Oxford nanopore MinION sequencing and genome assembly. *Genom. Proteom. Bioinform.* **14**, 265–279 (2016).
34. 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).

35. Yan, S. *et al.* Single molecule ratcheting motion of peptides in a Mycobacterium smegmatis Porin A (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
36. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
37. Zhang, S. *et al.* Bottom-up fabrication of a proteasome–nanopore that unravels and processes single proteins. *Nat. Chem.* **13**, 1192–1199 (2021).
38. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
39. Alfaro, J. A. *et al.* The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods.* **18**, 604–617 (2021).
40. Eisenstein, M. Seven technologies to watch in 2023. *Nature.* **613**, 794–797 (2023).
41. MacCoss, M. J., Alfaro, J., Wanunu, M., Faivre, D. A. & Slavov, N. Sampling the proteome by emerging single-molecule and mass-spectrometry methods. *Nat. Methods.* **20**, 339–346 (2022).
42. Leggett, R. M. & Clark, M. D. A world of opportunities with nanopore sequencing. *J. Exp. Bot.* **68**, 5419–5429 (2017).
43. Brown, C. G. & Clarke, J. Nanopore development at Oxford Nanopore. *Nat. Biotechnol.* **34**, 810–811 (2016).
44. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).
45. Howorka, S. Building membrane nanopores. *Nat. Nanotechnol.* **12**, 619–630 (2017).
46. Feng, J. *et al.* Identification of single nucleotides in MoS_2 nanopores. *Nat. Nanotechnol.* **10**, 1070–1076 (2015).
47. Miles, B. N. *et al.* Single molecule sensing with solid-state nanopores: Novel materials, methods, and applications. *Chem. Soc. Rev.* **42**, 15–28 (2013).
48. Yusko, E. C. *et al.* Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* **12**, 360–367 (2017).
49. Wei, R., Gatterdam, V., Wieneke, R., Tampé, R. & Rant, U. Stochastic sensing of proteins with receptor-modified solid-state nanopores. *Nat. Nanotechnol.* **7**, 257–263 (2012).
50. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2**, 209–215 (2007).
51. Tunuguntla, R. H. *et al.* Enhanced water permeability and tunable ion selectivity in subnanometer carbon nanotube porins. *Science.* **359**, 792–796 (2018).

52. He, Y., Tsutsui, M., Zhou, Y. & Miao, X. S. Solid-state nanopore systems: from materials to applications. *NPG Asia Mater.* **13**, 48 (2021).
53. Xue, L. *et al.* Solid-state nanopore sensors. *Nat. Rev. Mater.* **5**, 931–951 (2020).
54. Moreau, C. J., Dupuis, J. P., Revilloud, J., Arumugam, K. & Vivaudou, M. Coupling ion channels to receptors for biomolecule sensing. *Nat. Nanotechnol.* **3**, 620–625 (2008).
55. Cao, C. *et al.* Discrimination of oligonucleotides of different lengths with a wild-type aerolysin nanopore. *Nat. Nanotechnol.* **11**, 713–718 (2016).
56. Soskine, M. *et al.* An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry. *Nano Lett.* **12**, 4895–4900 (2012).
57. Song, L. *et al.* Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science.* **274**, 1859–1866 (1996).
58. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science.* **303**, 1189–1192 (2004).
59. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
60. 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).
61. Cherf, G. M. *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at five angstrom precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
62. Stoddart, D. *et al.* Nucleobase recognition in ssDNA at the central constriction of the α-hemolysin pore. *Nano Lett.* **10**, 3633–3637 (2010).
63. 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. Chemie Int. Ed.* **49**, 556–559 (2010).
64. Van Gerven, N., Van der Verren, S. E., Reiter, D. M. & Remaut, H. The role of functional amyloids in bacterial virulence. *J. Mol. Biol.* **430**, 3657–3684 (2018).
65. Goyal, P. *et al.* Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature.* **516**, 250–253 (2014).
66. Cao, B. *et al.* Structure of the nonameric bacterial amyloid secretion channel. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5439–E5444 (2014).
67. Boža, V., Brejová, B. & Vinař, T. DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS One.* **12**, 1–13 (2017).
68. 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).

69. Jayasinghe, L. & Wallace, J. E. Mutant pore. US 11186868 B2 (2021).

70. Van der Verren, S. E. *et al.* A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat. Biotechnol.* **38**, 1415–1420 (2020).

71. Chapman, M. R. *et al.* Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science.* **295**, 851–855 (2002).

72. 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).

73. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci.* **93**, 13770–13773 (1996).

74. Meller, A., Nivon, L. & Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys. Rev. Lett.* **86**, 3435–3438 (2001).

75. Laszlo, A. H., Derrington, I. M. & Gundlach, J. H. MspA nanopore as a single-molecule tool: From sequencing to SPRNT. *Methods* **105**, 75–89 (2016).

76. Caldwell, C. C. & Spies, M. Helicase SPRNTing through the nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11809–11811 (2017).

77. Heron, A. *et al.* Modified helicases. US 2021/0123032 A1 (2021).

78. Craig, J. M. *et al.* Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11932–11937 (2017).

79. 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).

80. Derrington, I. M. *et al.* Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.* **33**, 1073–1075 (2015).

81. Laszlo, A. H., Derrrington, I. M. & Gundlach, J. H. Subangstrom measurements of enzyme function using a biological nanopore, SPRNT. *Methods Enzymol.* **582**, 387–414 (2017).

82. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 1–11 (2016).

83. Clarke, J., White, J., Milton, J. & Brown, C. Coupling method. WO 2012/164270 A1 (2012).

84. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).

85. 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).

86. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods.* **15**, 201–206 (2018).

87. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods.* **16**, 1297–1305 (2019).

88. Keller, M. W. *et al.* Direct RNA sequencing of the coding complete Influenza A virus genome. *Sci. Rep.* **8**, 1–8 (2018).

89. 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).

90. Miller, C. *Ion channel reconstitution*. (Springer Science & Business Media, 2013).

91. Cao, C., Liao, D. F., Yu, J., Tian, H. & Long, Y. T. Construction of an aerolysin nanopore in a lipid bilayer for single-oligonucleotide analysis. *Nat. Protoc.* **12**, 1901–1911 (2017).

92. Baba, T. *et al.* Formation and characterization of planar lipid bilayer membranes from synthetic phytanyl-chained glycolipids. *Biochim. Biophys. Acta - Biomembr.* **1421**, 91–102 (1999).

93. Schmidt, J. Membrane platforms for biological nanopore sensing and sequencing. *Curr. Opin. Biotechnol.* **39**, 17–27 (2016).

94. Holden, M. A., Needham, D. & Bayley, H. Functional bionetworks from nanoliter water droplets. *J. Am. Chem. Soc.* **129**, 8650–8655 (2007).

95. Funakoshi, K., Suzuki, H. & Takeuchi, S. Lipid bilayer formation by contacting monolayers in a microfluidic device for membrane protein analysis. *Anal. Chem.* **78**, 8169–8174 (2006).

96. Urban, M. *et al.* Highly parallel transport recordings on a membrane-on-nanopore chip at single molecule resolution. *Nano Lett.* **14**, 1674–1680 (2014).

97. Jeon, T.-J., Malmstadt, N. & Schmidt, J. J. Hydrogel-encapsulated lipid membranes. *J. Am. Chem. Soc.* **128**, 42–43 (2006).

98. Shim, J. W. & Gu, L. Q. Stochastic sensing on a modular chip containing a single-ion channel. *Anal. Chem.* **79**, 2207–2213 (2007).

99. Heitz, B. A., Jones, I. W., Hall, H. K., Aspinwall, C. A. & Saavedra, S. S. Fractional polymerization of a suspended planar bilayer creates a fluid, highly stable membrane for ion channel recordings. *J. Am. Chem. Soc.* **132**, 7086–7093 (2010).

100. Daly, S. M., Heffernan, L. A., Barger, W. R. & Shenoy, D. K. Photopolymerization of mixed monolayers and black lipid membranes containing gramicidin A and diacetylenic phospholipids. *Langmuir* **22**,

1215–1222 (2006).

101. Nardin, C., Winterhalter, M. & Meier, W. Giant free-standing ABA triblock copolymer membranes. *Langmuir* **16**, 7708–7712 (2000).

102. Meier, W., Nardin, C. & Winterhalter, M. Reconstitution of channel proteins in (polymerized) ABA triblock copolymer membranes. *Angew. Chemie Int. Ed.* **39**, 4599–4602 (2000).

103. Brown, C. *et al.* Hairpin loop method for double strand polynucleotide sequencing using transmembrane pores. WO 2013/014451 A1 (2013).

104. González-Pérez, A., Stibius, K. B., Vissing, T., Nielsen, C. H. & Mouritsen, O. G. Biomimetic triblock copolymer membrane arrays: A stable template for functional membrane proteins. *Langmuir.* **25**, 10447–10450 (2009).

105. 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).

106. Fukasawa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M.-S. LongQC: A quality control tool for third generation sequencing long read data. *G3: Genes Genomes Genet.* **10**, 1193–1196 (2020).

107. 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).

108. 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).

109. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics.* **30**, 3399–3401 (2014).

110. Payne, A., Holmes, N., Rakyan, V. & Loose, M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics.* **35**, 2193–2198 (2019).

111. Gamaarachchi, H. *et al.* Fast nanopore sequencing data analysis with SLOW5. *Nat. Biotechnol.* **40**, 1026–1029 (2022).

112. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods.* **14**, 407–410 (2017).

113. Liu, Q. *et al.* Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).

114. Ni, P. *et al.* DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics.* **35**, 4586–4595 (2019).

115. Liu, H. *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* **10**, 4079 (2019).

116. Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **39**, 1394–1402 (2021).
117. Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat. Commun.* **12**, 7198 (2021).
118. Miclotte, G. *et al.* Jabba: hybrid error correction for long sequencing reads. *Algorithms Mol. Biol.* **11**, 10 (2016).
119. 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).
120. Madoui, M.-A. *et al.* Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics.* **16**, 327 (2015).
121. Sović, I. *et al.* Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat. Commun.* **7**, 11307 (2016).
122. Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods.* **15**, 461–468 (2018).
123. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* **32**, 2103–2110 (2016).
124. Ewing, A. D. *et al.* Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol. Cell.* **80**, 915-928.e5 (2020).
125. 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–e32 (2017).
126. De Lannoy, C. *et al.* The long reads ahead: de novo genome assembly using the MinION. *F1000Research* **6**, 1083 (2017).
127. Anderson, B. N. Optically based nanopore sequencing. EP 3482196 B1 (2022).
128. Huber, M. & Davidson, S. Efficient optical analysis of polymers using arrays of nanostructures. EP 3209800 B1 (2022).
129. Huber, M. Nanopore-based polymer analysis with mutually-quenching fluorescent labels. WO 2016/057829 A1 (2016).
130. Martin, H., Ossama, A., Terry, C. & Milya, D. Fluorescent polynucleotide sequencing methods and compositions. WO 2021/217146 A1 (2021).
131. Buzby, P. R., Meller, A., McNally, B., Fan, A. & Olejnik-Krzynmanska, E. Sequence preserved DNA conversion for optical nanopore sequencing. US 2012/0316075 A1 (2012).

132. McNally, B. *et al.* Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett.* **10**, 2237–2244 (2010).

133. Kumar, S. *et al.* PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci. Rep.* **2**, 1–8 (2012).

134. Masoud, V. *et al.* Systems and methods for measurment and sequencing of bio-molecules. EP 3449247 B1 (2021).

135. Ohshiro, T. *et al.* Single-molecule electrical random resequencing of DNA and RNA. *Sci. Rep.* **2**, 501 (2012).

136. Tsutsui, M., Taniguchi, M., Yokota, K. & Kawai, T. Identifying single nucleotides by tunnelling current. *Nat. Nanotechnol.* **5**, 286–290 (2010).

137. Lagerqvist, J., Zwolak, M. & Di Ventra, M. Fast DNA sequencing via transverse electronic transport. *Nano Lett.* **6**, 779–782 (2006).

138. Huang, S. *et al.* Identifying single bases in a DNA oligomer with electron tunnelling. *Nat. Nanotechnol.* **5**, 868–873 (2010).

139. Lee, J. W. Nanoelectrode-gated detection of individual molecules with potential for rapid DNA sequencing. *Solid State Phenom.* **121**–**123**, 1379–1386 (2007).

140. Choudhary, A. *et al.* High-fidelity capture, threading, and infinite-depth sequencing of single DNA molecules with a double-nanopore system. *ACS Nano.* **14**, 15566–15576 (2020).

141. Zhang, Y. *et al.* Single molecule DNA resensing using a two-pore device. *Small* **14**, 1801890 (2018).

142. Liu, X., Zhang, Y., Nagel, R., Reisner, W. & Dunbar, W. B. Controlling DNA tug-of-war in a dual nanopore device. *Small* **15**, 1901704 (2019).

143. Liu, X. *et al.* Flossing DNA in a dual nanopore device. *Small* **16**, 1905379 (2020).

144. Rand, A. *et al.* Electronic mapping of a bacterial genome with dual solid-state nanopores and active single-molecule control. *bioRxiv* (2021). doi:10.1101/2021.10.29.466509

145. Oliver, J. S. Devices and methods for determining the length of biopolymers and distances between probes bound thereto. US 8262879 B2 (2012).

146. Oliver, J. S. *et al.* High-definition electronic genome maps from single molecule data. *bioRxiv* (2017). doi:https://doi.org/10.1101/139840

147. Passera, A. *et al.* Characterization of Lysinibacillus fusiformis strain S4C11: In vitro, in planta, and in silico analyses reveal a plant-benefcial microbe. *Microbiol. Res.* **244**, 126665 (2021).

148. Weigand, M. R. *et al.* Screening and genomic characterization of filamentous hemagglutinin-deficient Bordetella pertussis. *Infect. Immun.* **86**, 5–7 (2018).

149. Kaiser, M. D. *et al.* Automated structural variant verification in human genomes using single-molecule electronic DNA mapping. *bioRxiv.* (2017). doi:https://doi.org/10.1101/140699
150. Bošković, F. *et al.* Simultaneous identification of viruses and viral variants with programmable DNA nanobait. *Nat. Nanotechnol.* **18**, 290–298 (2023).
151. Bošković, F. & Keyser, U. F. Nanopore microscope identifies RNA isoforms with structural colours. *Nat. Chem.* **14**, 1258–1264 (2022).
152. Takulapalli, B. Field effect transistor, device including the transistor, and methods of forming and using same. US 2020/0096505 A1 (2020).
153. Traversi, F. *et al.* Detecting the translocation of DNA through a nanopore using graphene nanoribbons. *Nat. Nanotechnol.* **8**, 939–945 (2013).
154. Mojtabavi, M., Vahidmohammadi, A., Liang, W., Beidaghi, M. & Wanunu, M. Single-molecule sensing using nanopores in two-dimensional transition metal carbide (MXene) membranes. *ACS Nano.* **13**, 3042–3053 (2019).
155. Schneider, G. F. *et al.* Tailoring the hydrophobicity of graphene for its use as nanopores for DNA translocation. *Nat. Commun.* **4**, 2619 (2013).
156. Nicoli, F., Verschueren, D., Klein, M., Dekker, C. & Jonsson, M. P. DNA translocations through solid-state plasmonic nanopores. *Nano Lett.* **14**, 6917–6925 (2014).
157. Assad, O. N. *et al.* Light-enhancing plasmonic-nanopore biosensor for superior single-molecule detection. *Adv. Mater.* **29**, 1605442 (2017).
158. Larkin, J., Henley, R. Y., Jadhav, V., Korlach, J. & Wanunu, M. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nat. Nanotechnol.* **12**, 1169–1175 (2017).
159. Zhang, Y., Tang, Y., Tan, C. & Xu, W. Toward Nanopore Electrospray Mass Spectrometry: Nanopore Effects in the Analysis of Bacteria. *ACS Cent. Sci.* **6**, 1001–1008 (2020).
160. Bush, J. *et al.* The nanopore mass spectrometer. *Rev. Sci. Instrum.* **88**, 1–9 (2017).
161. Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nat. Biotechnol.* **18**, 1091–1095 (2001).
162. Thakur, A. K. & Movileanu, L. Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore. *Nat. Biotechnol.* **37**, 96–104 (2019).
163. Lukoyanova, N. *et al.* Conformational changes during pore formation by the perforin-related protein pleurotolysin. *PLoS Biol.* **13**, 1002049 (2015).

164. Wallace, A. J. *et al.* E. coli hemolysin E (Hlye, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell.* **100**, 265–276 (2000).

165. Huang, G. *et al.* Electro-osmotic vortices promote the capture of folded proteins by PlyAB nanopores. *Nano Lett.* **20**, 3819–3827 (2020).

166. Galenkamp, N. S., Soskine, M., Hermans, J., Wloka, C. & Maglia, G. Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores. *Nat. Commun.* **9**, 4085 (2018).

167. Soskine, M., Biesemans, A., De Maeyer, M. & Maglia, G. Tuning the size and properties of ClyA nanopores assisted by directed evolution. *J. Am. Chem. Soc.* **135**, 13456–13463 (2013).

168. Galenkamp, N. S., Biesemans, A. & Maglia, G. Directional conformer exchange in dihydrofolate reductase revealed by single-molecule nanopore recordings. *Nat. Chem.* **12**, 481–488 (2020).

169. Xing, Y., Dorey, A., Jayasinghe, L. & Howorka, S. Highly shape- and size-tunable membrane nanopores made with DNA. *Nat. Nanotechnol.* **17**, 708–713 (2022).

170. Lanphere, C. *et al.* Design, assembly, and characterization of membrane-spanning DNA nanopores. *Nat. Protoc.* **16**, 86–130 (2021).

171. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).

172. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).

173. Jiang, S., Ge, Z., Mou, S., Yan, H. & Fan, C. Designer DNA nanostructures for therapeutics. *Chem* **7**, 1156–1179 (2021).

174. Dey, S. *et al.* A reversibly gated protein-transporting membrane channel made of DNA. *Nat. Commun.* **13**, 2271 (2022).

175. Diederichs, T. *et al.* Synthetic protein-conductive membrane nanopores built with DNA. *Nat. Commun.* **10**, 5018 (2019).

176. Xia, P., Rahman Laskar, M. A. & Wang, C. Wafer-scale fabrication of uniform, micrometer-sized, triangular membranes on sapphire for high-speed protein sensing in a nanopore. *ACS Appl. Mater. Interfaces* **15**, 2656–2664 (2023).

177. Li, W. *et al.* Single protein molecule detection by glass nanopores. *ACS Nano.* **7**, 4129–4134 (2013).

178. Bell, N. A. W., Chen, K., Ghosal, S., Ricci, M. & Keyser, U. F. Asymmetric dynamics of DNA entering and exiting a strongly confining nanopore. *Nat. Commun.* **8**, 380 (2017).

179. Sze, J. Y. Y., Ivanov, A. P., Cass, A. E. G. & Edel, J. B. Single molecule multiplexed nanopore protein screening in human serum using aptamer modified DNA carriers. *Nat. Commun.* **8**, 1552 (2017).

180. Bernèche, S. *et al.* Modeling and simulation of ion channels. *Chem. Rev.* **112**, 6250–6284 (2012).

181. Varongchayakul, N., Song, J., Meller, A. & Grinstaff, M. W. Single-molecule protein sensing in a nanopore: a tutorial. *Chem. Soc. Rev.* **47**, 8512–8524 (2018).

182. Firnkes, M., Pedone, D., Knezevic, J., Döblinger, M. & Rant, U. Electrically facilitated translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* **10**, 2162–2167 (2010).

183. Schmid, S., Stömmer, P., Dietz, H. & Dekker, C. Nanopore electro-osmotic trap for the label-free study of single proteins and their conformations. *Nat. Nanotechnol.* **16**, 1244–1250 (2021).

184. Bandara, Y. M. N. D. Y., Farajpour, N. & Freedman, K. J. Nanopore current enhancements lack protein charge dependence and elucidate maximum unfolding at protein's isoelectric point. *J. Am. Chem. Soc.* **144**, 3063–3073 (2022).

185. John, S., Aksimentiev, A., Joo, C. & States, U. SDS-assisted protein transport through solid-state nanopores. *Nanoscale.* **9**, 11685–11693 (2017).

186. Rivas, F. *et al.* Label-free analysis of physiological hyaluronan size distribution with a solid-state nanopore sensor. *Nat. Commun.* **9**, 1037 (2018).

187. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).

188. Lin, C. Y. *et al.* Ultrafast polymer dynamics through a nanopore. *Nano Lett.* **22**, 8719–8727 (2022).

189. Chau, C. C., Radford, S. E., Hewitt, E. W. & Actis, P. Macromolecular crowding enhances the detection of DNA and proteins by a solid-state nanopore. *Nano Lett.* **20**, 5553–5561 (2020).

190. Keyser, U. F. Enhancing nanopore sensing with DNA nanotechnology. *Nat. Nanotechnol.* **11**, 106–108 (2016).

191. Bell, N. A. W. & Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat. Nanotechnol.* **11**, 645–651 (2016).

192. Tripathi, P. *et al.* Electrical unfolding of cytochrome c during translocation through a nanopore constriction. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016262118 (2021).

193. Movileanu, L., Schmittschmitt, J. P., Scholtz, J. M. & Bayley, H. Interactions of peptides with a protein pore. *Biophys. J.* **89**, 1030–1045 (2005).

194. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **8**, 288–295 (2013).

195. Feng, J. *et al.* Transmembrane protein rotaxanes reveal kinetic traps in the refolding of translocated substrates. *Commun. Biol.* **3**, 159 (2020).

196. Rosen, C. B., Bayley, H. & Rodriguez-Larrea, D. Free-energy landscapes of membrane co-

translocational protein unfolding. *Commun. Biol.* **3**, 160 (2020).

197. Rodriguez-Larrea, D. Single-amino acid discrimination in proteins with homogeneous nanopore sensors and neural networks. *Biosens. Bioelectron.* **180**, 113108 (2021).

198. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano.* **8**, 12365–12375 (2014).

199. Yao, Y., Docter, M., Van Ginkel, J., De Ridder, D. & Joo, C. Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys. Biol.* **12**, 055003 (2015).

200. Ohayon, S., Girsault, A., Nasser, M., Shen-Orr, S. & Meller, A. Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. *PLOS Comput. Biol.* **15**, e1007067 (2019).

201. Afshar Bakshloo, M. *et al.* Nanopore-based protein identification. *J. Am. Chem. Soc.* **144**, 2716–2725 (2022).

202. Lucas, F. L. R., Versloot, R. C. A., Yakovlieva, L., Walvoort, M. T. C. & Maglia, G. Protein identification by nanopore peptide profiling. *Nat. Commun.* **12**, 5795 (2021).

203. Yu, L. *et al.* Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* (2023).

204. Restrepo-Pérez, L. *et al.* Resolving chemical modifications to a single amino acid within a peptide using a biological nanopore. *ACS Nano.* **13**, 13668–13676 (2019).

205. Wang, R. *et al.* Single-molecule discrimination of labeled DNAs and polypeptides using photoluminescent-free TiO_2 nanopores. *ACS Nano.* **12**, 11648–11656 (2018).

206. Cardozo, N. *et al.* Multiplexed direct detection of barcoded protein reporters on a nanopore array. *Nat. Biotechnol.* **40**, 42–46 (2022).

207. Kennedy, E., Dong, Z., Tennant, C. & Timp, G. Reading the primary structure of a protein with 0.07 nm 3 resolution using a subnanometre-diameter pore. *Nat. Nanotechnol.* **11**, 968–976 (2016).

208. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* **32**, 179–181 (2014).

209. Restrepo-Pérez, L., Wong, C. H., Maglia, G., Dekker, C. & Joo, C. Label-free detection of post-translational modifications with a nanopore. *Nano Lett.* **19**, 7957–7964 (2019).

210. Ohshiro, T. *et al.* Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotechnol.* **9**, 835–840 (2014).

211. Zhao, Y. *et al.* Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **9**, 466–473 (2014).

212. Ensslen, T., Sarthak, K., Aksimentiev, A. & Behrends, J. C. Resolving isomeric posttranslational modifications using a biological nanopore as a sensor of molecular shape. *J. Am. Chem. Soc.* **144**, 16060–16068 (2022).
213. Brinkerhoff, D. H. & Dekker, C. Protein and peptide fingerprinting and sequencing by nanopore translocation of peptide-oligonucleotide complexes. WO 2021/133168 A1 (2021).
214. Heron, J. A., Edward, G. J. & Slawa, S. M. Method of characterising a target polypeptide using a nanopore. WO 2021/111125 A1 (2021).
215. Filip, B. & Keyser, U. Toward single-molecule proteomics. *Science.* **374**, 1443–1444 (2021).
216. Wanunu, M. Back and forth with nanopore peptide sequencing. *Nat. Biotechnol.* **40**, 172–173 (2022).
217. Aggarwal, V. & Ha, T. Single-molecule fluorescence microscopy of native macromolecular complexes. *Curr. Opin. Struct. Biol.* **41**, 225–232 (2016).
218. Cohen, L. & Walt, D. R. Single-molecule arrays for protein and nucleic acid analysis. *Annu. Rev. Anal. Chem.* **10**, 345–363 (2017).
219. Edman, P. A method for the determination of the amino acid sequence in peptides. *Arch. Biochem.* **22**, 475–476 (1949).
220. Swaminathan, J. *et al.* Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
221. Reed, B. D. *et al.* Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science.* **378**, 186–192 (2022).
222. Kafader, J. O. *et al.* Measurement of individual ions sharply increases the resolution of Orbitrap mass spectra of proteins. *Anal. Chem.* **91**, 2776–2783 (2019).
223. Makarov, A. & Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1486–1495 (2009).
224. Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. R. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods.* **9**, 1084–1086 (2012).
225. Kafader, J. O. *et al.* Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nat. Methods.* **17**, 391–394 (2020).
226. Mock, A., Braun, M., Scholl, C., Fröhling, S. & Erkut, C. Transcriptome profiling for precision cancer medicine using shallow nanopore cDNA sequencing. *Sci. Rep.* **13**, 2378 (2023).
227. Nova, I. C. *et al.* Mapping phosphorylation post-translational modifications along single peptides with nanopores. *bioRxiv* (2022). doi:10.1101/2022.11.11.516163

228. Bennett, H. M., Stephenson, W., Rose, C. M. & Darmanis, S. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nat. Methods.* **20**, 363–374 (2023).

229. Motone, K. & Nivala, J. Not if but when nanopore protein sequencing meets single-cell proteomics. *Nat. Methods.* **20**, 336–338 (2023).

230. Candia, J. *et al.* Assessment of variability in the SOMAscan assay. *Sci. Rep.* **7**, 14248 (2017).

231. Niu, M. *et al.* Droplet-based transcriptome profiling of individual synapses. *Nat. Biotechnol.* (2023). doi:10.1038/s41587-022-01635-1

232. Mamanova, L. & Turner, D. J. Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq). *Nat. Protoc.* **6**, 1736–1747 (2011).

233. Koch, C. *et al.* Highly multiplexed detection of microRNAs, proteins and small molecules using barcoded molecular probes and nanopore sequencing. *bioRxiv* (2022). doi:10.1101/2022.12.13.520243

234. Mahendran, K. R. *et al.* A monodisperse transmembrane α-helical peptide barrel. *Nat. Chem.* **9**, 411–419 (2017).

235. Xu, C. *et al.* Computational design of transmembrane pores. *Nature.* **585**, 129–134 (2020).

236. Mardis, E. R. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* **6**, 287–303 (2013).

237. 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).

238. Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nat. Biotechnol.* **27**, 1013–1023 (2009).

239. Kandaswamy, V., Eugene, T. & Bernard, M. A. Nucleic acid sequencing methods and systems. WO 2017/014762 A1 (2017).

240. Meslier, V. *et al.* Benchmarking second and third-generation sequencing platforms for microbial metagenomics. *Sci. Data* **9**, 694 (2022).

241. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature.* **461**, 1248–1253 (2009).

242. Remaut, H., Jayasinghe, L. & Howorka, S. Mutant CsgG Pores. WO 2016/034591 A3 (2016).

243. Van der Verren, S. E. *et al.* A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat. Biotechnol.* **38**, 1415–1420 (2020).

244. Iinuma, R. *et al.* Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT. *Science.* **344**, 65–69 (2014).

245. Dai, M., Jungmann, R. & Yin, P. Optical imaging of individual biomolecules in densely packed clusters. *Nat. Nanotechnol.* **11**, 798–807 (2016).

246. Filius, M., Kim, S. H., Severins, I. & Joo, C. High-resolution single-molecule FRET via DNA eXchange (FRET X). *Nano Lett.* **21**, 3295–3301 (2021).

247. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods.* **14**, 865–868 (2017).

248. Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell.* **184**, 2988-3005.e16 (2021).

249. Trzupek, D. *et al.* Discovery of CD80 and CD86 as recent activation markers on regulatory T cells by protein-RNA single-cell analysis. *Genome Med.* **12**, 55 (2020).

250. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102–e102 (2011).

251. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods.* **13**, 269–275 (2016).