

Long reads become a reality

Despite the high accuracy of short-read sequencing methods, inherent limitations prevent the analysis of some very interesting parts of the genome. In 2009, two fundamentally different sequencing technologies — single-molecule real-time sequencing (SMRT-seq) by Pacific Biosciences (PacBio) and nanopore sequencing by Oxford Nanopore Technologies (ONT) — began to emerge, changing the genomics landscape by introducing multi-kilobase-scale reads.

Instrumental to the development of SMRT-seq were zero-mode waveguides (ZMWs), which modify wave properties of light to allow emitted fluorescence to be observed in a very small volume. This permits single-molecule detection of a high concentration of fluorophore-labelled

pulses provides information on the kinetics of incorporation. Multiple linked copies of a circular template are generated owing to the high processivity and strand-displacement activity of the polymerase — a feature that greatly improved the accuracy of PacBio sequencing in its later iterations.

Work leading to the launch of the other major player in long-read sequencing, the MinION device by ONT, had been ongoing for several decades. The concept of using an electric potential to drive a DNA molecule through a protein pore embedded in a lipid bilayer and inferring the sequence from the subtle changes in detected current as nucleotides temporarily blocked the protein channel was first proposed in 1989.

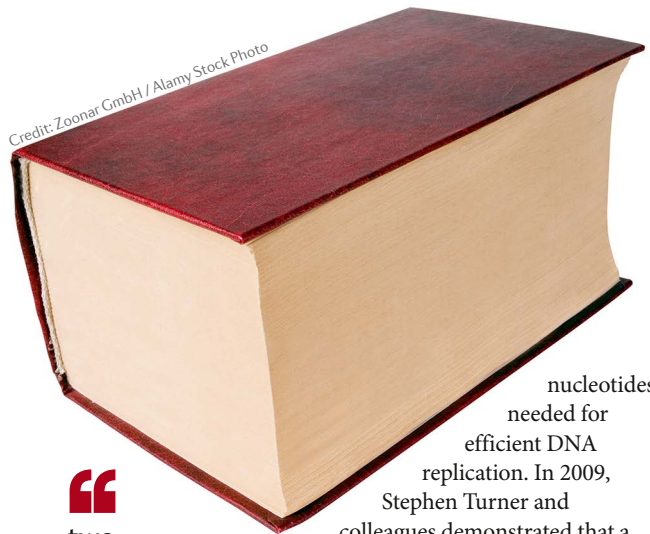
In 2009, Hagan Bayley's group showed that an engineered version of the bacterial channel protein α -haemolysin can be used for ordered continuous detection of cleaved nucleotides that were electrophoresed through the pore. A few years later, the groups of Mark Akeson and Jens Gundlach demonstrated that preloading a polymerase on the DNA allowed it to pass through the pore slowly enough for changes in current to be detected, solving a key challenge for the practical implementation of nanopore sequencing. In 2014, the pocket-sized MinION device was offered to a group of early adopters through the MinION Access Programme (MAP), which soon led to the first publication highlighting the device's performance and computational tools for analysis of the resulting data.

An exciting feature of long-read sequencing is the use of native DNA, which allows modifications to be detected. In SMRT-seq, modified bases affect the kinetics of the polymerase, permitting their identification, but accurate detection relies on high coverage. In nanopore sequencing, modified bases can be detected as electrical current changes;

in 2017, two studies provided the first demonstration of genome-wide DNA methylation profiling by the MinION. The ability to detect modified nucleotides is not limited to DNA, as demonstrated by a study showing that nanopore sequencing can be extended to profile the human transcriptome, providing information on splicing, polyadenylation, haplotypes and RNA modifications.

Sequencing long continuous stretches of DNA or RNA, aided by advances in computational algorithms to decode raw signals and assemble genomes *de novo*, has opened numerous possibilities. Following improvements in library preparation methods and MinION chemistry, in 2018 the first human genome assembled *de novo* from ultra-long (>100 kb) nanopore reads was published. This was an improvement over previous drafts, providing longer contigs and insight into regions such as the major histocompatibility complex (MHC) locus and telomere repeats. This effort has continued; the Telomere-to-telomere (T2T) consortium recently gave us the first gapless assembly of a human chromosome (MILESTONE 17), a feat achieved by a combination of the latest developments in nanopore and PacBio sequencing and assembly algorithms capable of harnessing the power of long reads.

Ivanka Kamenova, *Nature Protocols*



Credit: Zoonar GmbH / Alamy Stock Photo

“
two fundamentally different sequencing technologies ... [changed] the genomics landscape by introducing multi-kilobase-scale reads



nucleotides needed for efficient DNA replication. In 2009, Stephen Turner and colleagues demonstrated that a phage polymerase immobilized on a ZMW could be used to amplify DNA and read its sequence in real time. Using deoxynucleotides linked with distinct fluorophores via the terminal phosphate results in cleavage and diffusion of the fluorescent signal following phosphodiester bond formation, such that a brief pulse in fluorescence is detected only while the nucleotide resides in the active site of the enzyme. The emission spectra of the fluorophores permit different nucleotides to be distinguished, while the interval between successive

ORIGINAL ARTICLES Eid, J. et al. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009) | Clarke, J. et al.

Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009)

FURTHER READING Flusberg, B. A. et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010) | Chertif, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012) | 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) | Jain, M. et al. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015) | Rand, A. C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017) | Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017) | Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018) | Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019)