cytiva

Third generation sequencing and the evolution of NGS

Innovations in short read sequencing are increasing the number of next generation sequencing (NGS) applications every year. NGS applications include discovery of Mendelian disease variants, detection of SARS-CoV-2, and novel methods of cancer diagnosis. While faster and easier to perform than previous sequencing techniques, standard short read NGS methods still have limitations. They cannot answer all the difficult questions in cell biology, disease research, and agrigenomics. Advances in third generation (or long-read) sequencing, however, are now providing solutions for these challenges and giving insight into obscure genomic regions unreachable by existing NGS methods.

Next-generation sequencing technology

The introduction of NGS in the mid-2000s marked the beginning of a new age of human discovery, granting us unparalleled access to the genetic code. The fundamental importance of the genome to biological systems and their function makes NGS an extremely versatile tool used widely in research, clinical, and industrial applications.

Fig 1. The introduction of NGS platforms in the mid-2000s marked the beginning of a new age of human discovery.

NGS was developed to address the need for a faster, lower-cost alternative to Sanger sequencing, the dominant method since its inception in 1977 (1). Based mainly on 'sequencing by synthesis' (SBS) technology, prominent manufacturers of NGS systems include US-based companies like Illumina and Thermo Fisher, as well as BGI, which developed a competing approach called DNA nanoball sequencing.

Innovations in NGS are driving scientific breakthroughs and transforming healthcare, enabling and advancing personalized medicine. Several clinical NGS initiatives including the 100 000 genomes project, the Actionable Genome Consortium, and the NCI MATCH trial aim to harness its full power by bringing NGS data into the hands of clinicians and have helped cement NGS as a diagnostic tool. Innovations in sample prep reagents by companies like Illumina are pushing the boundaries of what is possible with short read sequencing, expanding the number of NGS applications every year.

Find out more about key clinical NGS initiatives

NGS is vital for new methods of cancer diagnosis such as liquid biopsies, and is also used for assessment of cancer patients for targeted drug therapies. These types of applications are driving large growth in the sequencing market, which is predicted to reach $24.4 billion by 2023 (2). Other important markets for NGS include synthetic biology and direct-to-consumer testing, both of which are expanding rapidly.

The current trend for labs to automate their NGS process is also driving rapid progress in medicine. With the increased reproducibility and scalability that comes with automating processes like DNA isolation, target enrichment, and library prep, automated NGS protocols might soon be routine in diagnostic applications for diseases such as cancer (3) and HIV-1 (4).

So, what is the future of NGS? We are fortunate to already have a variety of different technologies at our disposal, including those of third generation sequencing, which provides a comprehensive array of approaches for investigating disease, optimizing crop yields, and deepening our knowledge of cell biology. The immense potential of third generation sequencing is only now becoming clear, and the technology is shedding light on complex genomic processes that are not easily studied using other methods.

The emergence of third generation sequencing

What is third generation sequencing?

Third generation sequencing, also known as long-read sequencing, emerged in the late-2000s, and provides a method of sequencing a single DNA molecule without the need for any substantial prior fragmentation or amplification. There are currently two methods in widespread use: single molecule, real-time (SMRT™) sequencing from Pacific Biosciences (PacBio™), and Nanopore™ sequencing from Oxford Nanopore Technologies (ONT).

Base calling with SMRT sequencing involves measuring fluorescence that occurs when labeled nucleotides are added to a growing chain by DNA polymerases affixed to the bottom of a zeptoliter (10^{-21} L) well. Nanopore sequencing uses a motor protein to drive DNA molecules though a nanoscale pore and achieves base calling by analyzing the change in current through the pore. Figure 2 compares these two approaches to long-read sequencing.

While 'second generation' sequencing technologies like the Illumina platform have enabled discoveries and revolutionized medicine over the last decade, there are still some questions that they simply can't answer, especially when the target genomic region has complex rearrangements or repetitive regions. In these cases, third generation sequencing is often a viable alternative. With some vendors even offering portable sequencers, the possibilities for discovery and practical applications are continually expanding.

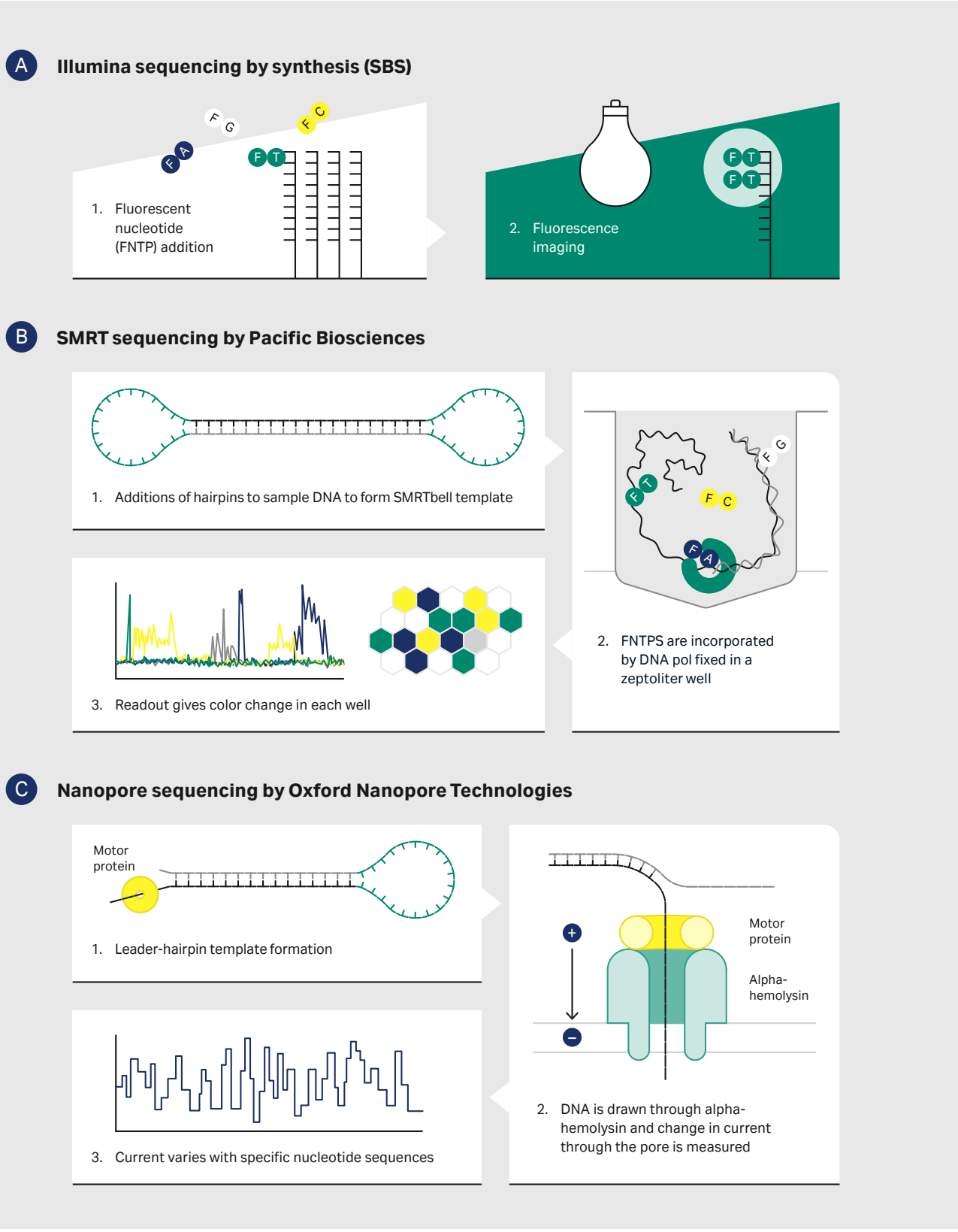

Fig 2. Overview of the main NGS methods in widespread use. A) Illumina sequencing by synthesis (SBS). DNA fragments bind to oligos on a flow cell and undergo steps of clonal amplification. Complementary strands are then synthesized using nucleotides labelled with specific fluorophores and the sequences read upon laser stimulation. B) Third generation SMRT sequencing by Pacific Biosciences. Hairpin sequences are added to double-stranded DNA to form a SMRTbell template, enabling continuous circular sequencing. Then, fluorescently labelled nucleotides are incorporated into a complementary strand by DNA polymerase fixed to the bottom of a zeptoliter volume well, and sequencing performed by laser stimulation and measurement of fluorescence signal from each well. C) Third generation Nanopore sequencing by Oxford Nanopore Technologies. Leader and hairpin sequences with motor proteins are attached to the sample DNA. The motor protein then drives the DNA through a nanoscale pore formed of alpha-hemolysin. Measuring the specific changes in current through the pore provides the sequencing output.

NGS facilitates pioneering discoveries

Of the 5,740 single gene Mendelian disorders identified by the Online Mendelian Inheritance in Man™ (OMIM) catalog, only 69.7% have been linked to a causative gene. Rare Mendelian conditions affect 20–30 million people in the US alone, which is why research into their genetic etiology is of such critical importance (5).

Researchers have discovered gene variants responsible for several genetic diseases using NGS, taking advantage of the massively parallel nature of SBS technology that provides a rapid and inexpensive method of variant discovery. Innovations in third generation technology are now also enabling identification of novel disease-causing variants that conventional methods could not sequence.

Discovering gene variants responsible for rare Mendelian diseases

Traditional techniques for identifying disease-causing genes, such as linkage analysis, rely on there being many cases from the same family for success. This requirement means that these techniques are unable to reveal variants responsible for many rare Mendelian diseases, which might be associated with low reproductive fitness or early lethality (6).

In these challenging cases, NGS offers a viable solution, and scientists are now able to identify rare disease-causing variants by sequencing only a small number of unrelated individuals. Often, these studies use whole exome sequencing (WES), which involves enrichment of exonic (protein coding) DNA using microarrays or magnetic beads with attached oligonucleotides.

The first use of WES to discover the genetic basis of a Mendelian disease was in 2010, when variants in the *DHODH* gene were found to be responsible for Miller syndrome (postaxial acrofacial dysostosis) (7). The same year also saw identification of gene variants responsible for rare multi-system disorders including Kabuki syndrome and Sensenbrenner syndrome (8)(9).

By extending the region of the genome that can be analyzed, whole genome sequencing (WGS) can also identify variants in noncoding DNA and shed light on the etiology of diseases that can't be fully explained by other methods. For example, WGS was used recently to identify variants in enhancers that lead to Hirschsprung disease, and discover promoter region variants that lead to Baratela-Scott syndrome (10)(11). Our understanding of both of these rare Mendelian diseases has improved thanks to NGS.

Learn how to streamline your WES or WGS process in our eBook

Using NGS to identify disease-causing variants with complex genetic etiology

An important limitation of short read sequencing, however, is the need to fragment DNA to 150–300 bp, making it difficult to accurately assemble short genomic regions with structural variations (SV) or repetitive sequences, which can be highly individual or population-specific and might not be represented in the reference genome (12). The conventional method of handling these regions on SBS platforms is with paired read sequencing, which involves sequencing from both ends of a fragment, and enables more accurate alignment of genomic regions with SVs or repetitive sequences.

Using a combination of paired read and third generation sequencing, recent studies have identified retrotransposition to be responsible for a certain form of Parkinsonism, and showed that expansion of pentanucleotide repeats causes a subtype of epilepsy (13)(14). These discoveries have important implications for the study of disease and could inform future diagnostics and treatment options for patients.

In some cases, paired read sequencing cannot identify a complex causative gene variant, leading scientists to rely solely on the long read lengths of third generation sequencing for answers. In one instance, SMRT sequencing was used to identify a disease-causing deletion and secure a diagnosis of Carney complex, which is a rare genetic disorder characterized by widespread neoplasia that is difficult to diagnose with short read methods (15).

Discover more clinical applications of NGS in our whitepaper

Clinical uses of NGS

Fig 3. NGS has a variety of clinical uses including testing of individuals with family histories of cancer.

NGS has a variety of clinical uses, such as prenatal testing for diseases of chromosome aneuploidy and BRCA testing of individuals with a family history of ovarian cancer (16)(17).

Novel applications of NGS are providing a critical link between research and medicine, and techniques such as liquid biopsy sequencing are steadily surpassing traditional methods of cancer diagnosis and enabling targeted treatment of specific cancers. NGS-based methods are also well suited for detection of pathogens for rapid diagnosis and better disease control strategies, such as the SARS-CoV-2 virus.

Detection of SARS-CoV-2

As the COVID-19 pandemic seized the world in 2020, implementing efficient mass testing programs became a top priority for disease control agencies, especially since the pathogen can be spread by pre-symptomatic and asymptomatic carriers.

Multiplexed NGS provides the potential for high-throughput detection of SARS-CoV-2. One example is Illumina's dual index COVIDSeq test, which can sequence up to 3000 patient samples in one run with a 24-hour turnaround (8). Other similar studies have described the potential for sequencing tens of thousands of patient samples for SARS-CoV-2 detection in a single run (9).

While SBS tests are an effective tool for diagnosis of COVID-19, it is the third generation platform from ONT that is revolutionizing the way we can monitor different strains of SARS-CoV-2. The ONT MinION™ platform can sequence the entire SARS-CoV-2 genome in less than 24 hours. This capability means the technology can be used by scientists to monitor healthcare-associated COVID-19 cases for effective infection tracking and control in hospital and community settings (8).

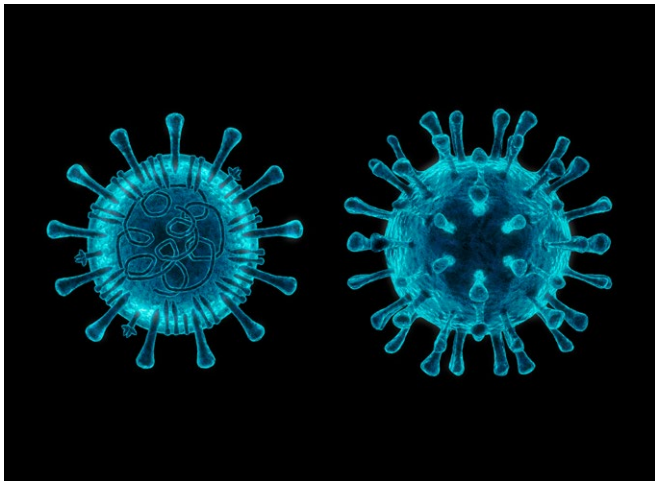

Fig 4. Multiplexed NGS provides the potential for high-throughput detection of SARS-CoV-2.

Liquid biopsies

NGS is revolutionizing cancer diagnosis, and therefore increasing survival rates. While traditional biomarkers are often not specific to a particular cancer and might provide only low diagnostic sensitivity, sequencing of circulating tumor DNA (ctDNA) from liquid biopsies, released by cancer cells undergoing apoptosis or necrosis, provides a versatile method for early detection of a range of cancers that is both rapid and precise. This DNA contains vital clues about the tissue of origin and the molecular signature of the tumor (18).

Fig 5. Sequencing of circulating tumor DNA (ctDNA) from liquid biopsies provides a versatile method for early detection of a range of cancers.

Steadily surpassing tissue biopsies as a diagnostic tool, sequencing of liquid biopsies has been used to identify predisposing mutations up to two years before a cancer was confirmed, making it a promising method for early diagnosis (19). Liquid biopsies are also enabling the diagnosis of tumors from previously unreachable regions including the brain, where repeat sampling and neuroimaging are particularly challenging (20).

Read more about the challenges and opportunities with liquid biopsy in our whitepaper

Targeted treatments

With the emergence of NGS, researchers are now able to put their knowledge of cancer genetics to clinical use. NGS has proved indispensable in cancer diagnostics testing, which involves identifying the molecular profile of a tumor to assess the suitability of a patient for targeted treatment. Examples of FDA-approved targeted therapies include PARP inhibitors for BRCA1/2 mutated breast, ovarian, and prostate cancers and ALK inhibitors for lung cancers with rearrangements in EML4-ALK (11).

Fig 6. NGS has proved indispensable in cancer diagnostics testing to assess the suitability of a patient for targeted treatment.

One of the main challenges to implementing NGS for routine mutation detection in clinical laboratories worldwide has been the large initial investment required for instrumentation setup. As a result, many labs continue relying on the more laborious and time-consuming method of Sanger sequencing for these procedures. However, current research suggests that Nanopore sequencing could offer a promising alternative method, with detection of treatment-informing mutations reported to be more sensitive, manageable, and cost-effective than Sanger sequencing, and run at a fraction of the cost of SBS platforms (12).

Read our 10 top tips to reduce the cost of your NGS assay

Third generation sequencing provides novel genetic insights

As our understanding of the regulation of gene expression has deepened, it has become critical to develop sequencing technology that can effectively study these processes. Read lengths on third generation platforms average around 30 kb for both SMRT and Nanopore sequencing, although DNA molecules up to 2.3 Mb have been sequenced using Nanopore technology (21). Such high read lengths enable these platforms to handle sequencing of whole large genomes and intact mRNA molecules, and provide an efficient and novel means of measuring DNA methylation (22).

One early limitation of third generation technology was the potential for base calling errors. However, continued development of SMRT sequencing chemistries over the last few years has increased the raw base calling accuracy from 85% to 99%. The accuracy of Nanopore sequencing has also improved, with the introduction of adaptor reagents pushing base calling accuracy as high as 98% (21).

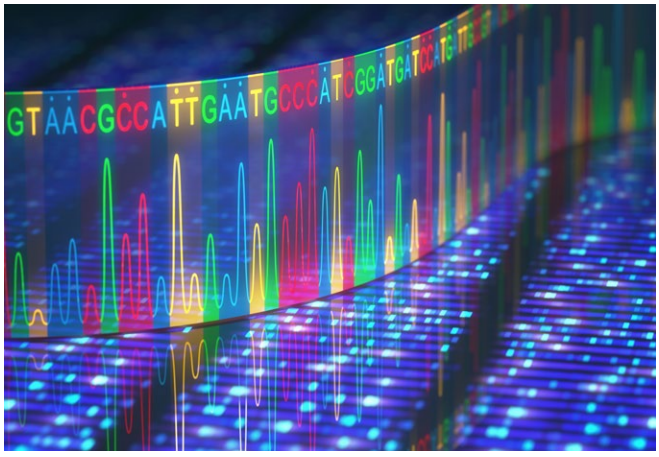

Fig 7. Understanding epigenetic processes is crucial for a better understanding of human biology, as modifications can be acquired by exposure to environmental factors.

Third generation sequencing simplifies study of the epigenome

Epigenetics refers to dynamic and inheritable chemical modifications that control gene expression by altering DNA accessibility and chromatin structure. Common modifications include acetylation of histones and methylation of DNA. Understanding epigenetic processes is crucial for a better understanding of human biology, as modifications can be acquired by exposure to environmental factors, and are important in the development of several diseases including cancer, heart disease, and obesity (23)(24)(25).

Conventional approaches to methylation sequencing on Illumina platforms rely on bisulfite conversion, in which unmethylated cytosines are converted to uracil and then thymine during library prep. While this technique has

enabled many discoveries, it has some disadvantages, such as unwanted DNA fragmentation, as well as the potential for unwanted conversion of 5-methylcytosine to uracil and underestimation of methylation status (26).

By integrating measurement of methylation status into the sequencing workflow without the need for additional steps of chemical conversion, third generation sequencing methods can rapidly and efficiently probe epigenetic changes. As well as 5-methylcytosine, third generation methods can also detect bases with several other non-canonical epigenetic modifications (27)(28).

SMRT sequencing enables measurement of methylation status by recording differences in arrival time and duration of fluorescence pulses that occur when nucleotides are incorporated into the growing chain. This platform has been used recently to investigate how genome methylation affects virulence and can prolong outbreaks of different pathogenic bacteria such as *E. coli* and MRSA (29)(30).

The specific waveforms generated during base calling in Nanopore sequencing enables the identification of methylated and non-methylated DNA. This ability means researchers can now distinguish viral and host DNA during analysis of viral epigenetic status and map differences in chromatin accessibility between cancerous and noncancerous cells (18)(19).

RNA isoform sequencing

It is only recently that we have begun to appreciate the extent to which RNA molecules are processed. Genes can be transcribed from different start sites, pre-mRNAs can be alternatively spliced, and the transcripts can also be polyadenylated at different sites (20). All these events lead to an astonishing variation in RNA transcripts and protein products beyond that apparently encoded in the genome.

While sequencing these different isoforms can reveal valuable biological information, SBS methods require cDNA to be fragmented prior to sequencing: this step risks losing information about the multitude of different RNA variants in a cell. Third generation technology does not have this limitation, so it is extremely useful for studies of RNA isoforms.

Aberrant splicing and intron retention have recently been identified in a gene linked to X-linked Dystonia-Parkinsonism (XDP), a discovery made possible by the long read capability of PacBio's SMRT sequencing platform (13). ONT's Nanopore sequencing is also making an impact in this area, with one group successfully sequencing and quantifying complex isoforms of B-cell receptors at the single cell level. This is a great stride in analytic capability and another step towards truly personalized medicine (31).

NGS and plant genomics

Building reference genomes for crop species using NGS is crucial for enabling the genetic modification of crops for resistance to biotic or abiotic stresses and higher yields. One major challenge to this process, however, is that many plant genomes are polyploid – meaning they contain more than two complete chromosome sets – which makes *de novo* genome assembly on conventional SBS platforms extremely challenging.

Agrigenomics often uses a statistic called the N50 to describe the completeness of a genome assembly, which is a type of median contig length. Only a small number of plant species sequenced using SBS have complete genome assemblies with N50s greater than 5 Mb, the size at which whole chromosomes can be reconstructed (32).

Recent studies using third generation sequencing are producing assemblies with higher contig lengths, with N50 values of 10.53 Mb for Asian rice and 15.78 Mb for maize (33)(34). These long reads can give a deeper insight into genomic architecture that can complement both genomics-based breeding and gene editing technologies for higher crop yields (32).

Conclusions

While optimization of SBS technology is improving read lengths and reliability of conventional SBS technology, the multitude of new approaches offered by third generation sequencing promises to usher in a new age of scientific discovery.

Long-read sequencing is shedding light on the multi-layered processes of gene expression and providing insights into fundamental cell biology and disease processes. This third-generation sequencing technology is also helping us to develop hardier crops with higher yields, which is essential for addressing the food needs of growing populations worldwide. Long-read sequencing of RNA molecules at a single cell level has the potential to revolutionize immunology and could lead to breakthroughs in our understanding of infectious diseases, autoimmunity, and cancer.

The sequencing market is growing, driven in a large part by new direct-to-consumer and clinical applications. With companies such as BGI planning to bring the cost of sequencing whole genomes to less than $100, the potential for more widespread application is huge. These advances in sequencing could even see DNA become a medium of information storage that replaces computer hard drives, and collaborations between industry and academia are currently working to make this a reality (35).

One thing is certain: we now have access to a comprehensive set of sequencing tools and technologies that researchers could barely conceive during the height of Sanger sequencing and the Human Genome Project. The continued evolution of both short- and long-read sequencing will shape the world of medicine and clinical practice for years to come.

Discover how we can support your sequencing process

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-5467. doi:10.1073/pnas.74.12.5463
2. BCC Publishing, Global DNA Sequencing: Research, Applied and Clinical markets. 2019. [Online]. Available: https://www.bccresearch.com/market-research/biotechnology/dna-sequencing-emerging-tech-applications-report.html
3. Koitzsch U, Heydt C, Attig H, *et al.* Use of the GeneReader NGS System in a clinical pathology laboratory: a comparative study. *J Clin Pathol*. 2017;70(8):725-728. doi:10.1136/jclinpath-2017-204342
4. Raymond S, Nicot F, Carcenac R, *et al.* HIV-1 genotypic resistance testing using the Vela automated next-generation sequencing platform. *J Antimicrob Chemother*. 2018;73(5):1152-1157. doi:10.1093/jac/dky003
5. Chong JX, Buckingham KJ, Jhangiani SN, *et al.* The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet*. 2015;97(2):199-215. doi:10.1016/j.ajhg.2015.06.009
6. Duncan E, Brown M, Shore EM. The revolution in human monogenic disease mapping. *Genes (Basel)*. 2014;5(3):792-803. Published 2014 Sep 5. doi:10.3390/genes5030792
7. Ng SB, Buckingham KJ, Lee C, *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010;42(1):30-35. doi:10.1038/ng.499
8. Ng SB, Bigham AW, Buckingham KJ, *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790-793. doi:10.1038/ng.646
9. Gilissen C, Arts HH, Hoischen A, *et al.* Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet*. 2010;87(3):418-423. doi:10.1016/j.ajhg.2010.08.004
10. Fu AX, Lui KN, Tang CS, *et al.* Whole-genome analysis of noncoding genetic variations identifies multiscale regulatory element perturbations associated with Hirschsprung disease. *Genome Res*. 2020;30(11):1618-1632. doi:10.1101/gr.264473.120
11. LaCroix AJ, Stabley D, Sahraoui R, *et al.* GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. *Am J Hum Genet*. 2019;104(1):35-44. doi:10.1016/j.ajhg.2018.11.005
12. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet*. 2019;10:426. Published 2019 May 7. doi:10.3389/fgene.2019.00426
13. Aneichyk T, Hendriks WT, Yadav R, *et al.* Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell*. 2018;172(5):897-909.e21. doi:10.1016/j.cell.2018.02.011
14. Ishiura H, Doi K, Mitsui J, *et al.* Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet*. 2018;50(4):581-590. doi:10.1038/s41588-018-0067-2
15. Merker JD, Wenger AM, Sneddon T, *et al.* Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med*. 2018;20(1):159-163. doi:10.1038/gim.2017.86
16. Manegold-Brauer G, Lapaire O "Clinical Implementation of Next-generation Sequencing in the Field of Prenatal Diagnostics. In Jerzy K Kulski, *Next Generation Sequencing - Advances, Applications and Challenges*. 2016, doi: 10.5772/61799.
17. Rumford M, Lythgoe M, McNeish I, *et al.* Oncologist-led BRCA 'mainstreaming' in the ovarian cancer clinic: A study of 255 patients and its impact on their management. *Sci Rep*. 2020;10(1):3390. Published 2020 Feb 25. doi:10.1038/s41598-020-60149-5
18. Chen M, Zhao H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics*. 2019;13(1):34. Published 2019 Aug 1. doi:10.1186/s40246-019-0220-8
19. Gormally E, Vineis P, Matullo G, *et al.* TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. *Cancer Res*. 2006;66(13):6871-6876. doi:10.1158/0008-5472.CAN-05-4556
20. Le Rhun E, Seoane J, Salzet M, Soffietti R, Weller M. Liquid biopsies for diagnosing and monitoring primary tumors of the central nervous system. *Cancer Lett*. 2020;480:24-28. doi:10.1016/j.canlet.2020.03.021
21. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):30. Published 2020 Feb 7. doi:10.1186/s13059-020-1935-5
22. Liu Y, Cheng J, Siejka-Zielińska P, *et al.* Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biol*. 2020;21(1):54. Published 2020 Mar 3. doi:10.1186/s13059-020-01969-6

23. Wajed SA, Laird PW, DeMeester TR. DNA methylation: an alternative pathway to cancer. *Ann Surg*. 2001;234(1):10-20. doi:10.1097/00000658-200107000-00003

24. Westerman K, Sebastiani P, Jacques P, Liu S, DeMeo D, Ordovás JM. DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure. *bioRxiv*, pp. 1–14, 2018, doi: 10.1101/471722.

25. Samani N. Epigenetics Studies Uncover Obesity-Driven Methylation Signatures. *iCommunity Newsletter*. October, 2014.

26. Kint S, De Spiegelaere W, De Kesel J, Vandekerckhove L, Van Criekinge W. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLoS One*. 2018;13(6):e0199091. Published 2018 Jun 14. doi:10.1371/journal.pone.0199091

27. Liu Q, Fang L, Yu G, Wang D, Xiao CL, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat Commun*. 2019;10(1):2449. Published 2019 Jun 4. doi:10.1038/s41467-019-10168-2

28. Flusberg BA, Webster DR, Lee JH, *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7(6):461-465. doi:10.1038/nmeth.1459

29. Forde BM, McAllister LJ, Paton JC, Paton AW, Beatson SA. SMRT sequencing reveals differential patterns of methylation in two O111:H- STEC isolates from a hemolytic uremic syndrome outbreak in Australia. *Sci Rep*. 2019;9(1):9436. Published 2019 Jul 1. doi:10.1038/s41598-019-45760-5

30. Webster E, Pak TR, Deikus G. Complete genome screening of clinical MRSA isolates identifies lineage diversity and provides full resolution of transmission and outbreak events. *bioRxiv*, pp. 1–42, 2019.

31. Singh M, Al-Eryani G, Carswell S, *et al.* High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat Commun*. 2019;10(1):3120. Published 2019 Jul 16. doi:10.1038/s41467-019-11049-4

32. Bohra A, Chand Jha U, Godwin ID, Kumar Varshney R. Genomic interventions for sustainable agriculture. *Plant Biotechnol J*. 2020;18(12):2388-2405. doi:10.1111/pbi.13472

33. Choi JY, Lye ZN, Groen SC, *et al.* Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. *Genome Biol*. 2020;21(1):21. Published 2020 Feb 5. doi:10.1186/s13059-020-1938-2

34. Yang N, Liu J, Gao Q, *et al.* Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet*. 2019;51(6):1052-1059. doi:10.1038/s41588-019-0427-6

35. Heckel R, Mikutis G, Grass RN. A Characterization of the DNA Data Storage Channel. *Sci Rep*. 2019;9(1):9663. Published 2019 Jul 4. doi:10.1038/s41598-019-45832-6

cytiva

cytiva.com/sequencing

CY19115-11Mar21-WP