PHG
FOUNDATION

making science work for health

Long-read sequencing
Clinical applications and implementation

Long-read sequencing (LRS) is attracting growing attention from the clinical genomics community, because research is increasingly demonstrating that LRS is better able to identify certain categories of genetic variant. This presents unique opportunities to expand the diagnostic potential of sequencing and to enable an 'omics approach due to the epigenomic and RNA sequencing that is possible using LRS (see *Clinical Long-Read Sequencing*). Current LRS technologies have additional technical characteristics of interest for clinical applications, for example, Oxford Nanopore Technologies can be used for real-time sequencing.

The use of long-read sequencing is being explored across a range of cancer and rare disease applications. This briefing will discuss some of the opportunities for LRS in clinical genomics, what this might mean for patients, and considerations relevant to implementation of these technologies.

The essentials

◆ LRS is being considered for a wide range of applications in clinical genomics
◆ LRS could improve diagnostics rates for complex rare diseases
◆ LRS technologies can sequence RNA or DNA modifications, and this can be used to create new tests for cancer to improve treatment decisions
◆ Implementation of LRS presents some challenges, including accuracy and throughput

WYNG
FOUNDATION

The views set out in this briefing note do not necessarily represent those of the WYNG Foundation

Long-read sequencing: clinical applications and implementation

Rare disease

The challenge: diagnosis

Short-read sequencing (SRS) has been instrumental in increasing diagnostic rates for rare disease patients. Globally, larger and more comprehensive sequencing methods are being adopted as a front-line test where rare disease is suspected. Despite these developments, there remains a diagnostic gap which SRS is not able to overcome. Long-read sequencing (LRS) offers a number of opportunities to improve testing for rare disease.

Opportunities using long-read sequencing

Missing diagnoses in rare disease can arise because there is limited evidence available with which to interpret the test findings, resulting in the reporting of variants of uncertain significance (VUS), or no candidate variants being identified. LRS can be used to improve this diagnostic process in a complementary manner by: 1) validating previous findings; 2) further characterising complex variants; and 3) identifying new diagnoses.

Missing diagnoses in rare disease are often caused by what is called the 'n of 1' challenge, where there is insufficient evidence to determine the pathogenicity of a specific rare variant/s found in a single individual. LRS will not be able to overcome this challenge. Improving the diagnostic process will be driven by the identification of new disease genes, data sharing, further study (i.e. functional analysis) and reanalysis with new evidence – some of which may be provided by LRS.

Structural variants

There is strong evidence that identifying structural variants (SV) will result in new diagnoses. A recent diagnostic study found that 13% of diagnoses came from SV, and current estimates suggest that >34% of known disease-causing variants are larger than 1bp [1]. Larger changes in DNA caused by SV are more likely to disrupt or change gene expression leading to disease. LRS improves our ability to identify larger and more complex variants.

The biggest success using LRS in diagnostic discovery studies has been untangling complex variants. These variants can be identified using cytogenetic techniques – laboratory methods to identify chromosomal changes including broken, missing, rearranged, or extra chromosomes. These methods are not as sensitive as sequencing technologies. LRS can be used to identify all affected genes, providing additional information to understand patient phenotypes and guide clinical management. For example, LRS was used to reclassify a very large chromosomal rearrangement in a patient affecting twelve genes, eleven of which were missed in previous genetic testing [2]. Two genes were associated with cardiac diseases and this link resulted in further investigations and follow-up care.

Long-read sequencing: clinical applications and implementation

Interpreting variants: the role of inheritance

Trio sequencing, where a patient and both their parents are sequenced, is used as a tool to determine the inheritance pattern of variants and their role in disease. However, trio sequencing may not always be possible (e.g. for adopted patients) making it harder to interpret the role of a variant in a disease. For example, compound heterozygous inheritance is where the presence of two different variants in the same gene (or allele), one inherited from each parent, result in disease. This is different from autosomal recessive disease, where the same gene variant is present in both alleles. LRS can be used to resolve the inheritance pattern of variants by identifying if they are located on the same or different copies of the gene. This process is known as haplotype phasing, and is a key advantage of LRS over SRS, where these reads are often too short to confirm the relative location of variants.

For example, a recent evaluation of whole genome sequencing for prenatal diagnostics in Hong Kong explored LRS to complement this diagnostic process [3]. In one case, SRS identified a frameshift variant in the gene *PCNT* compatible with clinical presentation, but the mechanism of disease for this gene is known to be autosomal recessive. One further breakpoint was identified in *PCNT*, hypothesised to be caused by a *de novo* inversion disrupting the second allele. However, these variants were 20kb apart. Therefore, LRS was combined with SRS to identify the inversion breakpoints and to confirm these two variants were both located in different copies of the gene. This resulted in a genetic diagnosis of microcephalic osteodysplastic primordial dwarfism type II. Importantly, this inversion was *de novo* and, therefore, the family could be counselled that it was unlikely that further children would be affected.

Cancer

The challenge: a complex landscape

The cancer testing landscape is complex, with different types of tests required at each stage of a patient's care pathway for clinical decision-making. Multiple assay platforms are used to provide these cancer diagnostics. There is more limited evidence for using SRS to identify known drivers (e.g., *BCR-ABL* translocations in CML) identified using cytogenetic or targeted genetic testing. These layers of testing all add to the challenge of getting the right test to the right patient at the right time.

Opportunities of long-read sequencing

LRS has key advantages over SRS because it can: 1) reduce ambiguity when reconstructing complex cancer genomes; 2) expand the use of epigenomic and RNA sequencing; and 3) simplify workflows. LRS can detect a broad range of variation and could, in principle, replace multiple technologies. LRS can also be performed in real-time, reducing the need for batching of samples, which has a number of advantages, such as, utilising fresh or fresh-frozen samples or in time-sensitive scenarios.

Long-read sequencing: clinical applications and implementation

Structural variants

LRS makes it easier to detect gene fusion events that are prevalent in many cancers. For example, chronic myeloid leukaemia is characterised by a specific fusion protein (BCR-ABL) and can be treated using targeted therapies. LRS can be used to test for this translocation and is more sensitive for identifying SV associated with treatment resistance compared to SRS methods [4]. LRS could improve diagnosis and monitoring for resistance caused by structural variants.

'Omics

Cancer diagnostics is increasingly making use of epigenomic and transcriptomic tests to provide clinically significant information. LRS can detect DNA modifications alongside sequence data, improving the sensitivity of cancer diagnostic tests that may be affected by the clonal nature of cancer genomes. LRS can sequence entire RNA sequences, and this can be used to detect changes in gene expression or to identify fusion proteins (e.g. BCR-ABL or ETV6-NTRK3) sensitive to targeted cancer therapies.

Looking to the future: single-molecule sequencing

Cancer is a clonal disease with genetic and epigenetic changes occurring over time in individual cancer cells. Current sequencing methods are not sufficiently sensitive to detect these sub-clonal changes that may lead to treatment resistance. As LRS accuracy improves, sequencing single nucleic acid molecules to generate consensus reads could detect sub-clonal genomic, epigenomic or transcriptomic markers to enable more effective clinical decision-making.

Case study: real-time diagnostics for brain cancer

Brain tumour surgery must balance the benefits of tumour tissue removal against the risks of severe brain damage. Molecular characterisation of tumours before surgery is not possible, but this information can inform prognosis and surgical decision-making. This is further complicated by the need to use a combination of genomic and epigenomic testing for precise molecular tumour classification [5]. LRS could simplify current workflows necessary to distinguish between tumours and enable informed decision-making during surgery.

Nanopore sequencing has been explored to create a rapid sequencing test to return results in time to inform surgical decision-making. In a feasibility study, results were returned after 97 minutes on average, and for 15 of 20 patients this would have been during the surgical window [6]. These results would have changed surgical strategies for 12 out of 20 patients. This shows the potential utility of LRS for these time-critical applications.

Long-read sequencing: clinical applications and implementation

Considerations for implementation

There have been significant advances in LRS and there is growing evidence for the potential value of LRS in clinical genomics. However, LRS does not overcome all shortcomings of preceding technologies and LRS technologies present specific challenges of their own. These factors need to be balanced against expectations regarding the value that these technologies can bring in clinical genomics.

◆ **Sequencing accuracy:** This can be defined based on individual read accuracy or consensus data from multiple reads. There have been significant improvements in LRS accuracy which have narrowed the gap compared to SRS methods. PacBio HiFi sequencing is the highest accuracy LRS method (>99.8% accuracy). Hybrid sequencing can combine LRS and SRS to produce consensus data taking advantage the strengths and limitations of both sequencing approaches.

◆ **Cost vs. throughput:** LRS is estimated to be approximately 3-4 times more expensive than SRS in terms of cost per gigabase sequenced. This is mostly explained by LRS being lower throughput than SRS and this will increase the cost of high-volume applications (e.g. WGS). However, LRS has advantages such as real time reporting and could be used for applications where a rapid turnaround is required.

◆ **Input material:** In order for long-reads to be sequenced successfully, the DNA sample must be of sufficient quality and have limited breakage. This is higher quality than the minimum required for SRS. In addition, LRS approaches require larger DNA input volumes than SRS. These requirements could be problematic for some clinical samples.

◆ **Computational requirements:** LRS can have high computational requirements to produce sequencing data, and this may put additional demands on clinical laboratories who may have insufficient infrastructure to manage and process this data.

◆ **Genome assembly:** Genome assembly will either compare this sequence data to a reference genome (assembly-based) or reconstruct the original nucleotide sequence without a reference (*de novo* assembly). LRS reduces the complexity of *de novo* methods, and this approach is advantageous for some applications where the genome differs significantly from the reference. However, these methods are still being developed for LRS clinical applications.

In addition to these considerations, LRS technologies have some unique characteristics which may make them more suitable to certain settings.

◆ **Pacific Biosciences:** PacBio HiFi sequencing quality is comparable to, and in some complex genomic regions exceeds, the sequencing quality obtained using SRS [7]. PacBio HiFi sequencers are suited to large, high-resource laboratories for applications requiring high quality sequence data for variant detection in complex genomic regions or complex variants.

Long-read sequencing: clinical applications and implementation

◆ **Oxford Nanopore Technologies:** Nanopore sequencers have specific characteristics uniquely adaptable for use in a point-of-care setting (portability, rapid sample preparation workflows, real-time sequencing). These characteristics are also valuable in low-resource settings with more limited laboratory infrastructure. Additionally, nanopore sequencers can generate ultra-long reads in excess of 100kb and can perform adaptive sampling. Adaptive sampling is a computational method for selective sequencing of target regions of the genome creating more flexibility over traditional targeted methods.

Ultimately, many of the considerations for adoption of LRS will be similar for any sequencing technology. Decisions as to which platform is most appropriate for a particular purpose and setting will be a trade-off between the specific technological capabilities of the sequencing approach and the diagnostic purpose.

Conclusions

Evidence for the value of LRS in clinical genomics is growing. LRS presents wide-ranging opportunities not restricted to any one application. Therefore, the question is not "will LRS be useful?" but rather, "how, where and when will LRS be useful?" Barriers remain to implementation of LRS systems and therefore adoption of these technologies needs to consider appropriate robust validation where LRS may provide the greatest value.

References

1. Eichler EE. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. New England Journal of Medicine, 2019. 381(1): 64-74.
2. Miller DE, Sulovari A, Wang T et al. Targeted long-read sequencing identifies missing disease-causing variation. The American Journal of Human Genetics, 2021. 108(8): 1436-1449.
3. Chung BHY, Kan ASY, Chan KYK et al. Analytical validity and clinical utility of whole-genome sequencing for cytogenetically balanced chromosomal abnormalities in prenatal diagnosis: abridged secondary publication. Hong Kong Med J. 2022. 28 Suppl 1(1): 4-7.
4. Schaal W, Ameur A, Olsson-Strömberg U et al. Migrating to Long-Read Sequencing for Clinical Routine BCR-ABL1 TKI Resistance Mutation Screening. Cancer Informatics, 2022. 21: 11769351221110872.
5. Euskirchen P, Bielle F, Labreche K et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. Acta Neuropathologica, 2017. 134(5): 691-703.
6. Djirackor L, Halldorsson S, Niehusmann P et al. Intraoperative DNA methylation classification of brain tumors impacts neurosurgical strategy. Neuro-Oncology Advances, 2021. 3(1).
7. Foox J, Tighe SW, Nicolet CM et al. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. Nature Biotechnology, 2021. 39(9): 1129-1140.

Author: Heather Turner
Published: November 2022

UNIVERSITY OF CAMBRIDGE