Long read sequencing: ready for the clinic?

Emma Johnson
emma.johnson@phgfoundation.org

Sobia Raza
sobia.raza@phgfoundation.org

Long read sequencing (LRS) technologies have been used widely as research tools in human, animal, plant and pathogen genomics. Beyond the research setting, there are examples of these techniques being used to assist in clinical diagnostic assessment. LRS technologies provide several advantages for DNA and RNA sequencing over traditional short-read sequencing (SRS) techniques.

This briefing builds on the accompanying policy briefing, *What is long-read sequencing?* to describe the potential significance of long-read technologies for diagnostic sequencing in a clinical setting and, in this context, the current challenges with the technology.

The essentials

- LRS platforms have the ability to sequence in real time, more easily detect large and complex genomic variants, and to sequence common features of the genome that are difficult to sequence using predominant SRS technologies

- There are several diseases and disorders for which LRS could provide advantages in diagnosis

- Whilst adoption of LRS in the clinical setting is currently limited, these technologies are evolving rapidly and the associated technical challenges are actively being addressed

- Given the above, it will be important to closely monitor how this technology evolves and how it might be incorporated into clinical practice in the future

UNIVERSITY OF CAMBRIDGE

POLICY BRIEFING

What are the potential clinical implications of long-read sequencing?

The advantages of sequencing longer reads are set out in the accompanying note *What is long-read sequencing?* The ability to produce long reads has implications for clinical sequencing and analysis. In the clinical setting, LRS technologies could be especially useful for:

- **Rare disease diagnosis:** Easier detection of large variants and complex structural features which can be implicated in disease. For example, there are around 40 neurological conditions – including the spinocerebellar ataxias, Fragile X syndrome and myotonic dystrophy – that are associated with repeat expansions, and for which LRS might provide a useful diagnostic tool. In 2017, long-read, whole-genome sequencing was used in the clinic to identify a large genomic rearrangement and diagnose a rare disease in a previously undiagnosed patient [1]

- **Cancers:** LRS makes it easier to detect gene fusion events and chromosomal rearrangements that are prevalent in many cancers. These technologies also enable same day genomic and epigenomic analyses which may provide clinically relevant information for diagnosis. Thus, they could facilitate greater speed and personalisation of diagnosis and treatment. In addition, the ability to perform full transcript (RNA) profiling has advantages for understanding the clinical significance of differing RNA forms produced from the same gene (splice variants) in cancers

- **Human leukocyte antigen (HLA) typing:** Improved genotyping of the highly polymorphic human leukocyte antigen (HLA) region for diagnosis of various autoimmune disorders and for examination of this region prior to organ or stem cell transplant

- **Infectious disease genomics:** The comparative speed of LRS, plus flexible depth sequencing in just a few hours, makes it an attractive option for the *in situ* diagnosis of infectious disease agents. This could potentially facilitate a rapid logistical response for the identification and management of disease sources and disease spread

- **Simplifying laboratory workflows:** As LRS can detect a broad range of genomic variation, it could, in principle, replace multiple technologies currently used in the diagnosis of some genetic conditions. In addition, real-time sequencing obviates the need for batch sampling, meaning that samples can be analysed individually and immediately, without the need to collect and prepare a sufficient number of samples to perform a cost-efficient sequencing run. This has particular advantages in scenarios that utilise fresh or fresh-frozen samples, and in time-sensitive settings

Currently, the two primary producers of 'true' LRS platforms are Pacific BioSciences (PacBio) and Oxford Nanopore Technologies (Nanopore).

What are the current challenges with long-read sequencing?

Whilst LRS presents distinct opportunities, it does not necessarily address all the shortcomings of preceding sequencing technologies. In addition, LSR technologies present challenges of their own:

- **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 than the minimum required for SRS. In addition, some LRS approaches require larger DNA input volumes than SRS. These requirements could be problematic for some clinical samples

- **Error rate:** LRS provides comparatively low sequencing accuracy dependent on input sample quality: 85-89% for individual base calls[2] compared to >99% with certain SRS technology [3]. For some technologies error can be reduced by increasing the amount of sequencing performed to produce high consensus accuracy. Multiple sequencing of single, but shorter, molecules enables >99.9% consensus sequence accuracy with PacBio and around 95% with Nanopore[3], however using shorter reads can negate some of the benefit of long-read technology. Hybrid sequencing (using both LRS and SRS) is often currently preferred to optimise accuracy

- **Cost:** The cost of reading a genome to high accuracy is currently higher when compared to SRS. For example, human whole genome characterisation using PacBio systems has been estimated around $3000-6000 compared to <$1000 using SRS [1,4]. This could make it prohibitive for some applications, although costs are falling. The initial system costs are very low for some LRS systems, e.g. $1000 for a Nanopore MinION starter pack, but high for others – $350,000 for PacBio Sequel system at launch. This means that consideration should be given to whether services are better acquired through direct platform purchase or through an external sequencing service provider

- **Throughput:** For some but not all LRS platforms, the amount of DNA that can be analysed simultaneously is currently lower than is possible with SRS sequencing
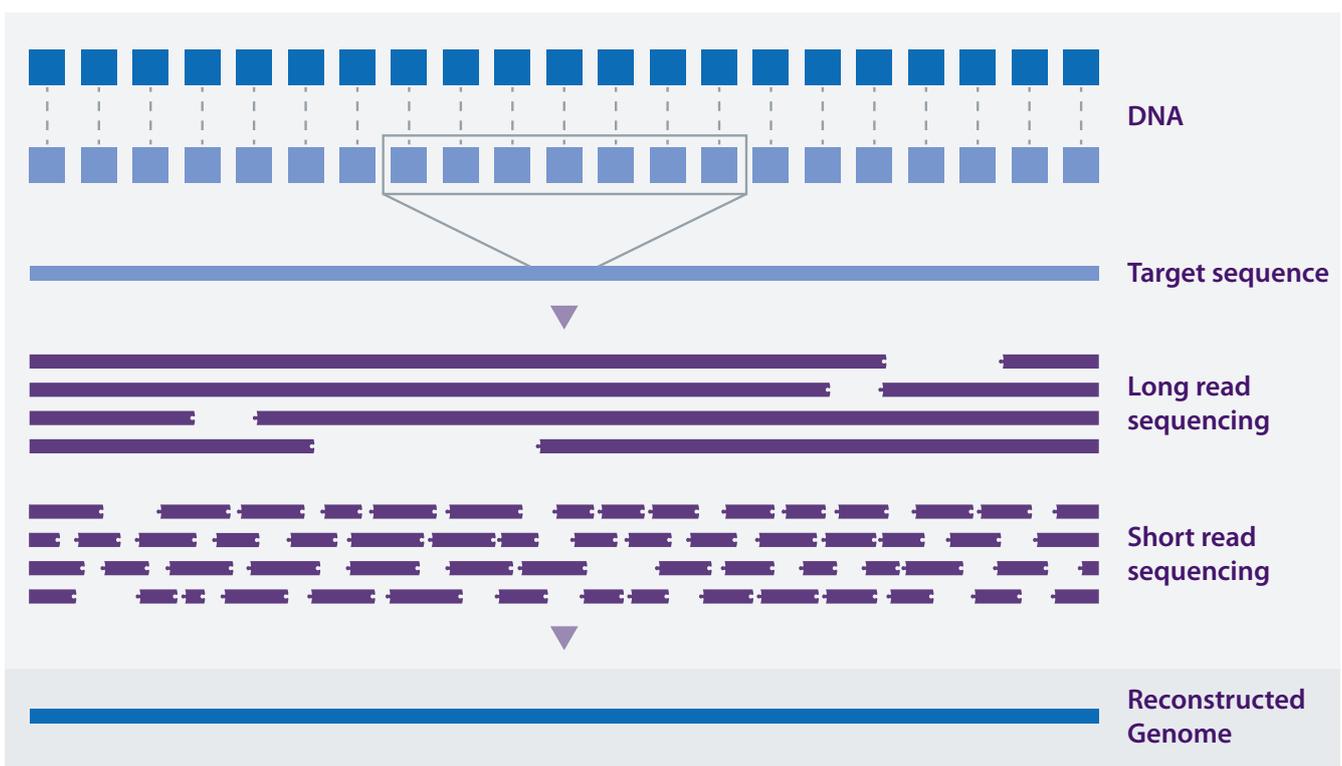
Conclusion

Whilst these issues may be holding back widespread adoption of the technology in the clinical setting, they are actively being addressed. It will be important for health systems to be aware of and critically consider how LRS, and the specific technologies on offer, could be best used in clinical practice. Ultimately, the degree to which these advantages and challenges manifest depends upon the specific combination of techniques, and the application of these flexible technologies.

References

1.   Stanford Medicine News Center. Researchers use long-read genome sequencing for first time in a patient. 2017. **https://med.stanford.edu/news/all-news/2017/06/researchers-use-long-read-genome-sequencing-in-a-patient.html**

2.   Mahmoud M. et al. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. Genomics, 2017. **https://doi.org/10.1016/j.ygeno.2017.12.011**

3.   Understanding Illumina Quality Scores. San Diego, CA: Illumina, Inc, 2014; Pub. No. 770-2012-058

4.   Pacific Biosciences Blog. For Reference-Grade Human Genome Assemblies, SMRT Sequencing Yields Optimal Results. 2018. **https://www.pacb.com/blog/for-reference-grade-human-genome-assemblies-smrt-sequencing-yields-optimal-results/**