

The Long and Short of NGS

Editorial

Next-generation sequencing (NGS) has revolutionized the field of genetic diagnostics over the last 10-15 years. NGS technologies have enabled us to look at the DNA at base pair level of resolution at genome wide scale as opposed to the very low-resolution techniques like karyotyping wherein the smallest abnormality which could be detected was around 5 million bases. Sanger sequencing technology enabled us to look at the DNA at base pair level for small sequences of up to 1000 bases. Repetitive Sanger sequencing for large number of fragments was the strategy used to sequence the first human genome in 2003 with a cost of about 3 billion dollars. Recently available NGS technologies have enabled sequencing of the whole human genome for as low as \$100 due to a drastic drop in the cost of sequencing. Most of the contemporary NGS technologies are based on 'short-read sequencing' wherein the DNA is fragmented into small pieces of about 250-300 bases and all such fragments are sequenced in a massively parallel sequencing. These short reads are then aligned to the reference genome using various bioinformatic tools and variants are identified by comparison of sequences with those in the reference genome. Short-read sequencing has been used in research as well as in the clinic for sequencing of groups of genes (panel sequencing), exome sequencing or genome sequencing. Short-read sequencing methods can detect single nucleotide variations and small insertions/deletions with a high degree of accuracy and hence have helped in the diagnosis of a large number of genetic disorders with a diagnostic yield of about 30-70% depending on the indication. Although short-read sequencing technologies have been successfully used for genetic diagnostics over the last few years, there are certain drawbacks of this technology. Structural variations like large deletions/duplications, translocations, insertions, triplet repeat expansions/contractions etc. are not detectable by short-read sequencing with sufficient accuracy.


The disadvantages associated with short-read

sequencing have led to the development of 'long-read sequencing' methodologies wherein large fragments of DNA ranging from few kilobases to megabases can be sequenced as a single strand. This enables the ability to detect structural variations and triplet repeat expansions with greater accuracy compared to short-read sequencing. The two majorly used long-read sequencing technologies are Oxford Nanopore and Pacific Biosciences based techniques. However, the long-read sequencing technologies suffer from lower accuracy for single nucleotide variations and small insertions/deletions. The ultimate ideal tool for DNA sequencing would be the one which can provide highly accurate long-read sequences so that all types of genetic variations can be detected in a single test and that too at a low cost. Newer advances in long-read sequencing technologies are enabling the increase in accuracy for small variants through use of better techniques as well as use of consensus sequences derived from iterative sequencing. The GenExpress in this issue discusses a few articles wherein long-read sequencing technologies have enabled detection of variants in genetic conditions like Duchenne muscular dystrophy (DMD), triplet repeat expansion disorders, etc. resulting from large rearrangements not detectable using short-read sequencing. Of interest is the article published by Dutta et al wherein long-read sequencing helped in detection of the breakpoints, at base-level resolution, of a de-novo reciprocal translocation.

The field of genetics is ever evolving and long read sequences can only allow us to look at DNA fragments of 1-2 Mb length, at the maximum capacity. However, the length of the human chromosomes ranges from few 50 - 250 Mb. The best technology available to look at large chromosomal rearrangements is karyotyping but its resolution is only around 5 Mb. Hence, there was a need for a technology to look at chromosomes at high resolution in order to detect large structural rearrangements and the same

was fulfilled by a recently developed technology called optical genome mapping (OGM). OGM involves labelling of DNA using fluorescent markers and then visualization of large fragments of DNA of hundreds of megabases. OGM is the only technique other than the cumbersome and low-resolution Southern blot technique, which can detect contraction of repeats in facioscapulohumeral muscular dystrophy (FSHD). These aspects of OGM technology have been highlighted in the article by Tallapaka et al in this issue.

It is an exciting time for medical genetics as newer technologies are enabling us to look at the human DNA in much more detail which is likely to help in the identification of genetic variants in all patients with genetic disorders as well as paving the way for high quality research.


(Dr. Ashwin Dalal)
Dr Ashwin Dalal
Assistant Editor
1st July 2023