STUDY MATERIALS

On

APPLICATIONS OF NGS ON GENOMIC RESEARCH

Prepared by-

Padmanav Koushik

M.Sc. in Zoology

(Animal Physiology and Biochemistry)

Certified Nutritionist

GATE qualified 2020

Next generation sequencing (NGS), massively parallel or deep sequencing are related terms that describe a DNA sequencing technology which has revolutionised genomic research. Using NGS an entire human genome can be sequenced within a single day. In contrast, the previous Sanger sequencing technology, used to decipher the human genome, required over a decade to deliver the final draft. Although in genome research NGS has mostly superseded conventional Sanger sequencing, it has not yet translated into routine clinical practice. The aim of this article is to review the potential applications of NGS in paediatrics. Several techniques began to appear such as sequencing by hybridization, parallel signature sequencing and pyrosequencing to name a few. These three methods addressed the need for shorter fragments, multiplexing, larger throughput and higher sensitivity, but it was pyrosequencing which was to become the first of the Next Generation of sequencers with the release of the 454 system by 454/Roche in 2005. This opened the floodgates for more next generation sequencers with several platforms entering the market in a relatively short period of time. Each of these machines shared similar performance on throughput, accuracy and cost in comparison to Sanger and so after the success of these initial platforms, new machines and techniques were rapidly developed to keep up with the demands of research laboratories. With rapid improvements in sequencing methods, several of the initial techniques used have already been superseded with simpler, more accurate techniques, but in an ever developing field, no technique remains the same for long.

Over the last 10 years, the field of sequencing has developed at a break neck speed with the ever expanding interest and knowledge of the importance of the underlying genetics of many diseases. One such field that has benefitted greatly from this is that of haematological malignancies. Next generation sequencing has made large advancements in the understanding of the underlying mechanisms of several haematological cancers over the last 10 years which would not have been possible with previous investigative techniques.

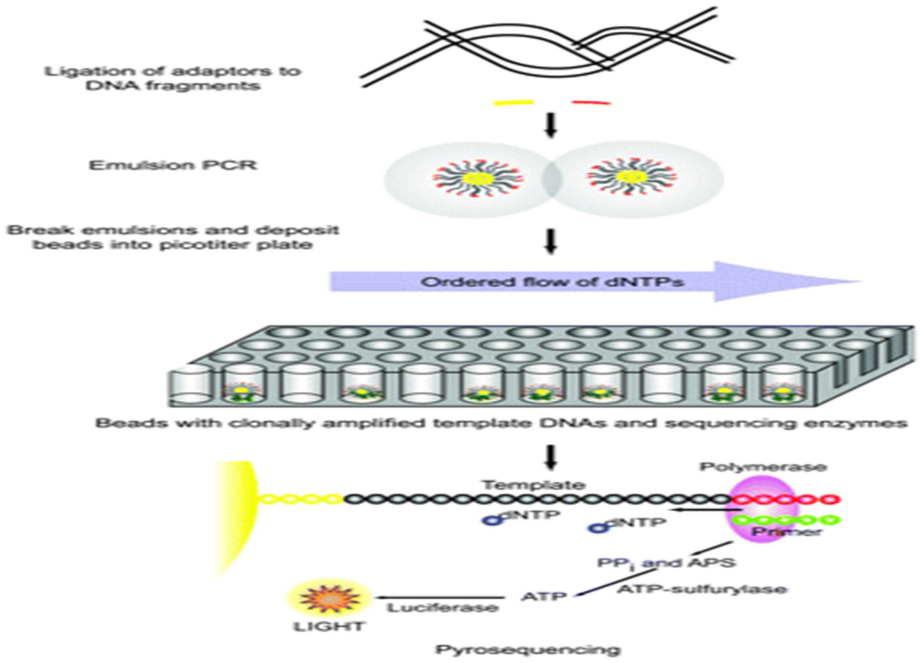
Multiple platforms have been released over the past 8 years which have opened the doors to high throughput, high sensitivity, quick whole genome sequencing. Each platform offers various sequencing approaches, such as whole genome sequencing (WGS), whole exome sequencing (WES), amplicon sequencing to name a few, alongside their own advantages and disadvantages and so a decision on which platform to ultimately go with is down to the needs of the researcher and the data produced, namely transcriptomics, genomics, and epigenomics.

Fundamentals of NGS Platforms

NGS platforms share a common technological feature—massively parallel sequencing of clonally amplified or single DNA molecules that are spatially separated in a flow cell. This design is a paradigm shift from that of Sanger sequencing, which is based on the electrophoretic separation of chain-termination products produced in individual sequencing reactions. In NGS, sequencing is performed by repeated cycles of polymerase-mediated nucleotide extensions or, in one format, by cycles of oligonucleotide ligation. As a massively parallel process, NGS generates hundreds of megabases to gigabases of nucleotide-sequence output in a single instrument run, depending on the platform. These platforms are reviewed next.

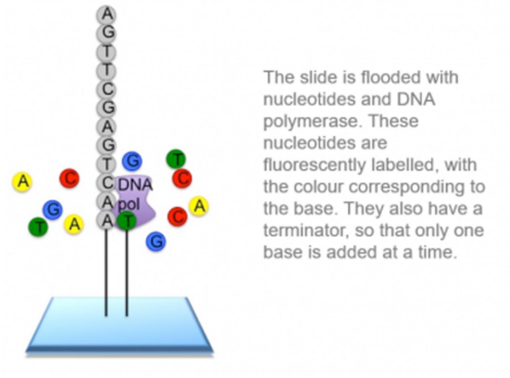
1) Roche/454-

Roche 454 sequencing can sequence much longer reads than Illumina. Like Illumina, it does this by sequencing multiple reads at once by reading optical signals as bases are added. As in Illumina, the DNA or RNA is fragmented into shorter reads, in this case up to 1kb. Generic adaptors are added to the ends and these are annealed to beads, one DNA fragment per bead. The fragments are then amplified by PCR using adaptor-specific primers. Each bead is then placed in a single well of a slide. So each well will contain a single bead, covered in many PCR copies of a single sequence. The wells also contain DNA polymerase and sequencing buffers. Template DNA is fragmented, end-repaired, ligated to adapters, and clonally amplified by emulsion PCR. After amplification, the beads are deposited into picotiter-plate wells with sequencing enzymes. The picotiter plate functions as a flow cell where iterative pyrosequencing is performed. A nucleotide-incorporation event results in pyrophosphate (PPi) release and well-localized luminescence. APS, adenosine 5'-phosphosulfate.

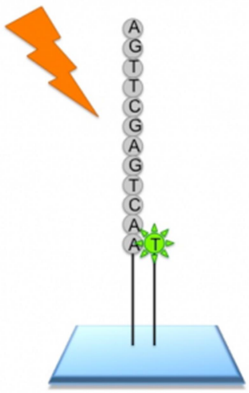


2) Illumina-

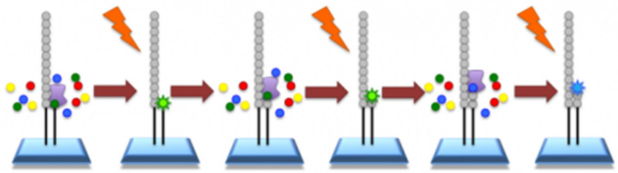
In NGS, vast numbers of short reads are sequenced in a single stroke. To do this, firstly the input sample must be cleaved into short sections. The length of these sections will depend on the particular sequencing machinery used. In Illumina sequencing, 100-150 bp reads are used. Somewhat longer fragments are ligated to generic adaptors and annealed to a slide using the adaptors. PCR is carried out to amplify each read, creating a spot with many copies of the same read. They are then separated into single strands to be sequenced.



The slide is flooded with nucleotides and DNA polymerase. These nucleotides are fluorescently labelled, with the colour corresponding to the base. They also have a terminator, so that only one base is added at a time.

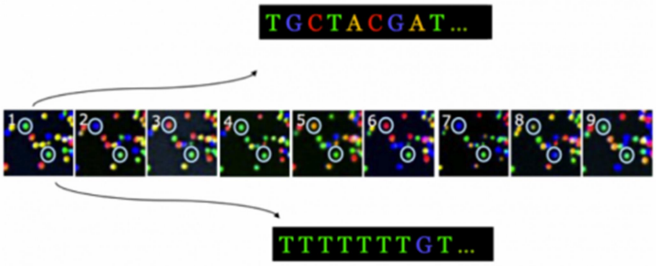


An image is taken of the slide. In each read location, there will be a fluorescent signal indicating the base that has been added.



The slide is then prepared for the next cycle. The terminators are removed, allowing the next base to be added, and the fluorescent signal is removed, preventing the signal from contaminating the next image.

The process is repeated, adding one nucleotide at a time and imaging in between.



Computers are then used to detect the base at each site in each image and these are used to construct a sequence.

All of the sequence reads will be the same length, as the read length depends on the number of cycles carried out.

3) Ion Torrent: Proton / PGM sequencing-

Unlike Illumina and 454, Ion torrent and Ion proton sequencing do not make use of optical signals. Instead, they exploit the fact that addition of a dNTP to a DNA polymer releases an H+ ion. As in other kinds of NGS, the input DNA or RNA is fragmented, this time ~200bp. Adaptors are added and one molecule is placed onto a bead. The molecules are amplified on the bead by emulsion PCR. Each bead is placed into a single well of a slide.

Applications of NGS

Next-generation DNA sequencing technology has had a significant impact on many areas of genomic research, by enabling us to design genome-wide and ultra deep sequencing projects that, because of their enormity, would not otherwise be possible. Examples include the following:

1. De Novo Sequencing-

The high yield of data obtained by sequencing run on a next-generation platform can provide sufficient quantity of data to sequence the entire genomes of several bacteria in a single run. The major challenge is piecing the data together to give a single contiguous assembly that represents the entire genome. This is made simpler by the use of long reads, such as those available from the 454 sequencing platform.

2. Whole Genome Resequencing-

Once a reference genome sequence becomes available, as a consequence of de novo sequencing and assembly, resequencing studies can then be performed. Resequencing involves the comparison of the genome sequence of other members of the species, or a related species, to the reference sequence, to delineate differences between them. This helps us to understand genotype/phenotype relationships and may ultimately help us to identify more appropriate treatments for disease and to develop novel diagnostic approaches.

3. Targeted Resequencing-

Even with a second-generation sequencer, sequencing large numbers of complex genomes in their entirety is still too time consuming and costly for most laboratories, and in addition, the analyses and storage of the data present a substantial burden to an establishment's informatics infrastructure. Consequently, 'target enrichment' methods have been developed, which involve the selective capture of desired regions from a DNA sample before sequencing. Resequencing the genomic regions that are retained is necessarily more time and cost effective, and the resulting data are considerably less cumbersome to analyze.

4. Metagenomics-

This is the study of the sequences of large populations of different organisms all growing in a common environment. DNA from all organisms is extracted together, and this mixed sample is sequenced. As with de novo sequencing, the main challenge lies in interpreting the data, which is greatly assisted by long sequence reads.

5. Transcriptomics-

The analysis of complementary DNA by next-generation sequencing (RNA-seq) enables us to build an accurate picture of active transcriptional patterns across the whole transcriptome of an organism.

6. DNA–Protein Interaction (ChIP-seq)-

Chemical cross-linking of DNA to proteins protects the DNA from enzymatic digestion, providing a way of identifying specific DNA–protein interactions. The protein and cross-linked DNA can be precipitated with an antibody, and the DNA can be isolated and sequenced.

7. Epigenetics-

DNA methylation is one of the most widely studied epigenetic modifications. Methylated DNA can be immune precipitated by methyl specific antibodies, and a combination of methylated DNA immune-precipitation with next-generation sequencing allows us to estimate absolute DNA methylation levels, making it a very powerful functional assay that is capable of assaying an entire human epigenome in a single experiment.

Potential uses of NGS in Clinical genetics

There are numerous opportunities to use NGS in clinical practice to improve patient care, including:

NGS captures a broader spectrum of mutations than Sanger sequencing-

The spectrum of DNA variation in a human genome comprises small base changes (substitutions), insertions and deletions of DNA, large genomic deletions of exons or whole genes and rearrangements such as inversions and translocations. Traditional Sanger sequencing is restricted to the discovery of substitutions and small insertions and deletions. For the remaining mutations dedicated assays are frequently performed, such as fluorescence in situ hybridisation (FISH) for conventional karyotyping, or comparative genomic hybridisation (CGH) microarrays to detect sub-microscopic chromosomal copy number changes such as micro-deletions. However, these data can also be derived from NGS sequencing data directly, obviating the need for dedicated assays while harvesting the full spectrum of genomic variation in a single experiment. The only limitations reside in regions which sequence poorly or map erroneously due to extreme guanine/cytosine (GC) content or repeat architecture, for example, the repeat expansions underlying Fragile X syndrome, or Huntington's disease.

Genomes can be interrogated without bias

Capillary sequencing depends on preknowledge of the gene or locus under investigation. However, NGS is completely unselective and used to interrogate full genomes or exomes to discover entirely novel mutations and disease causing genes. In paediatrics, this could be exploited to unravel the genetic basis of unexplained syndromes.

The increased sensitivity of NGS allows detection of mosaic mutations

Mosaic mutations are acquired as a post-fertilisation event and consequently they present at variable frequency within the cells and tissues of an individual. Capillary sequencing may miss these variants as they frequently present with a subtlety which falls below the sensitivity of the technology. NGS sequencing provides a far more sensitive read-out and can therefore be used to identify variants which reside in just a few per cent of the cells, including mosaic variation. In

addition, the sensitivity of NGS sequencing can be increased further, simply by increasing sequencing depth. This has seen NGS employed for very sensitive investigations such as interrogating foetal DNA from maternal blood or tracking the levels of tumour cells from the circulation of cancer patients.

-------0000000-------