Decrypting the Code of Life DNA Sequencing

Ronak Ashok Borana*

Department of Biotechnology, Bhavan's College, Mumbai-400 058, India

***Corresponding author:** Ronak Ashok Borana, Department of Biotechnology, Bhavan's College, Mumbai-400 058, Tel: +022 2625 6451; E-mail: ronakb128@gmail.com

Received Date: May 14, 2018; **Accepted Date:** July 05, 2018; **Published Date:** July 12, 2018

Abstract

In 1977, Sanger's dideoxy method ushered us into a new realm of DNA sequencing. Since its inception, millions of diverse genomes have been sequenced and published. While Sanger's method fueled the Human Genome Project, a new class of sequencing methods-the Next Generation Sequencing (NGS) has replaced it. Armed with more accurate variant identification, longer read length, decreased cost per megabase pair, and advanced bioinformatics, NGS have spearheaded our understanding of genes and their impact on our phenotype and disease. Several NGS technologies like the Illumina, 454, SOLiD and others are constantly being tweaked and modified to make them more accurate and inexpensive. Oxford Nanopore, a pocket-sized sequencing device is promising to

amplifying them, sequencing them and eventually stitching them together by matching the overlaps in the short sequences [7]. All the short sequences are read simultaneously and hence is also called as Massively parallel sequencing. The primary mode of working for NGS involves 3 broad steps.

1. Template creation and amplification.
2. Sequencing and imaging.
3. Alignment and assembly of genomic data.

Let's look further into the different types of NGS and how these steps fall into their protocol.

Roche 454/Pyrosequencing: Instead of dideoxy dNTPs, this method relies on detection of pyrophosphate that is released on the incorporation of NTPs [8]. Emulsion PCR is used to capture denatured DNA strand on amplification beads. An enzyme called ATP luciferase generates light when pyrophosphate (PPi) is released on induction of dNTP in the denatured DNA strand. This light is different for every dNTP (dATP, dTTP, dCTP, dGTP) and is recorded with each new round of dNTP wash.

Illumina: It was developed by Solexa which was acquired by Illumina Inc in 2007. The prepared sample is used to make small clusters on the flow cell using bridge amplification. The annealed strand is synthesized in double strand using fluorescent dNTPs on flowcells. Each dNTP gives a different color when analyzed by a laser. This color is recorded by a camera and analyzed by computers. Illumina and 454 are together also called as sequencing by synthesis method.

SOLiD: This method uses ligation and two base sequencing. The first step is similar, DNA is denatured, oligos and adapters are added and the strands are amplified using hybridized beads. An 8 base probe (first 2 bases are ligation site, next 3 are cleavage site and the last 3 are the fluorescent site) is used to find the nucleotide on the strand. 5 cycles of ligation, detection, and cleavage is performed for each sequence tag. SOLiD stands for Sequencing by Oligonucleotide Ligation and Detection (Figure 2).

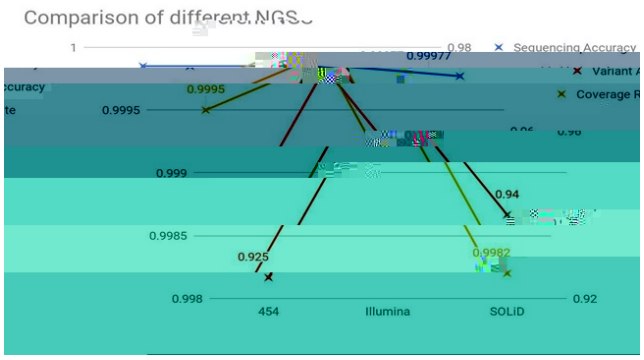

Figure 2: Comparison of different NGS for population target sequencing studies.

NGS outperform Sanger on many fronts. Chemical reaction and signal detection are two individual steps in Sanger but are combined into one in most NGS. Sanger can process one read of a longer length but NGS, on the other hand, can parallelly sequence hundreds of reads. NGS is not just inexpensive, but also easy to use and faster. The domination of NGS has led to the better variant identification and more exhaustive understanding of epigenomes and transcriptomes. The decrease in the cost per base pair and the sequencing of a large

products like MinION and PromethION which have democratized DNA sequencing. The unmatched portability and inexpensive run promises to one day, bring sequencing to every lab on earth.

Future

There is a new wave of TGS or Third generation sequencing tools that have started to flood the market [18]. Nanopore and Pacific Biosciences single polymerase sequencing have been leading this wave. Advancement in cloud computing and better processors is further going to decrease the cost and improve read length and accuracy. Companies like 23 & me and Ancestry.com which commercially genotype their user's DNA sample have seen a massive year on year growth. From Genotyping, the next very likely frontier for these companies will be to offer large-scale Whole Exome and Whole Genome Sequencing. More genomic data will lead to better prediction and of disease and its prognosis and eventually help providers chart a better course for recovery. Google recently released their new Deep Learning algorithm (Artificial Intelligence) called Deep Variant that helps in reconstructing full genomes from raw HTS data and apply Machine Learning to find SNPs.

From microbiologist, taxonomist, archeologist, evolutionary and marine biologist, everyone is using DNA sequencing to understand the flow of life at the most microscopic level. The idea of the most intimate secrets of an organism being sequenced with such ease raises a lot of privacy and surveillance concerns. Will we be moving into a world