Next Generation Sequencing-Current Status

Suma J[1] and Sivakumar G[2]*

[1]New York Genome Center, New York, USA

[2]Qassim University, College of Applied Medical Sciences, Buraidah, Kingdom of Saudi Arabia

*Corresponding author: Sivakumar Gowder, Qassim University, College of Applied Medical Sciences, Buraidah, Kingdom of Saudi Arabia, Tel: +966566873969; Fax: +96663802268; E-mail: sivakumargowder@yahoo.com

Introduction

Next Generation Sequencing (NGS) is revolutionizing the study of "omics" (e.g., genomics, transcriptomics, methylomics, etc). The sequencing cost has gone down tremendously and thereby allowing most of the methods to deviate from microarrays and traditional Sanger sequencing. Illumina has the highest market share in NGS followed by Thermofisher Scientific, Roche and Pacific Biosciences (http://cen.acs.org/articles/92/i33/Next-Gen-Sequencing-Numbers-Game.html). They currently sell the following instruments: HiSeqX series, HiSeq series, MiSeq, NextSeq and MiniSeq. With the arrival of HiSeqX, Illumina projects that the cost per Gigabase has gone down from $100,000,000 in the year 2000 to $1-10 in year 2014 [1]. Read-length, run time bacterial genome sequencing and maximum throughput on different platforms is reviewed in Van Djik et al., and the number of applications that can be performed using NGS is increasing every day. They can be categorized into three areas as described below [2].

NGS Applications

Genomic sequencing

With the drastic drop in the cost of sequencing, we can sequence the whole genomes since the human genome project. Whole Genome Sequencing (WGS) is used to characterize genetic markers associated with diseases. WGS has revolutionized our understanding of the relationship between genomic variants and its phenotypes. WGS in combination with long read sequencing is used to assemble novel genomes referring to as DeNovo sequencing. With the advance in technologies, WGS can now be done at a single cell level. Few among many single cell genomics applications include reconstructing cell lineage and understanding tumor heterogeneity.

Since many genomic variants that are associated with diseases occur in exome, researchers minimize sequencing cost by targeted sequencing: enriching for exons (Whole exome sequencing) or by multiplex PCR.

RNA Sequencing

Similar to WGS, mRNA sequencing can also be done on cells in bulk as well as at single cell level. WGS when coupled with mRNA sequencing allows us to understand the effect of a variant on gene expression. Total RNA-Seq after depletion of ribosomal RNA will not only shed light on exonic expression levels but also on intronic levels.

Location-Based Techniques

To better understand the transcription and translation processes, location based techniques such as ChIP-Seq (chromatin immunoprecipitation sequencing) and CLIP-Seq (crosslinking immunoprecipitation) have been developed to identify protein-DNA interactions and protein-RNA interactions respectively. Other methods to identify RNA-protein interactions include iCLIP and PAR-CLIP. Chromosome conformation capture (3C)-based systems such as 4C, 5C,6C, Hi-C and ChIA-PET has also been developed to identify DNA-DNA interactions. These methods help in 3D-construction of the genome. Also, methods for epigenetic measurements such as DNaseI-Seq, ATAC-Seq and bisulfite sequencing have been established. While DNaseI-Seq and ATAC-Seq provide information on the openness and structure of chromatin; bisulfate sequencing shed light on DNA methylation. Recently, ATAC-Seq and bisulfate sequencing were developed for single-cell analysis [3,4].

Long Molecule Sequencing

While Illumina is great for sequencing short inserts, long inserts are needed for repetitive regions in genomes which cannot be easily amplified. Key players in long range sequencing are Pacific Biosciences, 10X Genomics, and Oxford Nanopore Technologies. Sample and library preparation cost.

Even though the cost for sequencing is coming down, the challenge remains to bring down the cost of sample and library preparation. Especially for highly multiplexed and low coverage sequencing purposes, the cost per sample of library preparation overweighs the cost per sample of sequencing. To address this challenge, researchers are coming up with methods that can be performed in 96 or 384-well plate formats. Automating sample and library preparation will save a lot of time and thereby bring down the effective cost per preparation. Traditional library preparation includes shearing, end-repair, A-tailing, ligating and amplification. It takes around 1-2 days to perform a library preparation. An exciting alternative to this traditional library preparation method is Transposase based library construction that can be completed in few hours. Illumina commercializes this kit as Nextera. For low coverage sequencing purposes, several multiplexing strategies have very well addressed this challenge. Recent publications demonstrate split-pool barcoding approach for multiplexing thousands of single cells [5,6].

NGS in Clinics

With the advent of many technologies in NGS and a drastic drop in the cost of sequencing, NGS has now made its way into clinics. It is changing the way the diseases are diagnosed. One of the most common

applications is the cancer diagnosis. Until recently, cancers are diagnosed based on morphological criteria, but a huge effort is undertaken to shift the paradigm of diagnosis based on genetic mutations and fusions [7]. Another well-known application is NIPT (Non-invasive prenatal testing). Many companies are now offering tests such as Natera, Verifi, MaterniT21 for NGS-based NIPT. NGS has a role in personalized medicine. Personalized medicine has a control over individualizing medicine both in terms of treatment and diagnosis. NGS also has the potential to accelerate the early detection of disorders and toxicity and to detect pharmacogenetics markers to customize treatments [8-10].

Computational Challenges

With most of the applications tending towards NGS, excess data is being generated. As a result of this, the cost of storing data is considerably increasing than that of the cost of sequencing to produce equivalent amount of data. Also, it is becoming a real challenge to handle the huge amount of data generated in an efficient manner both regarding storage and analysis. One of the possible solutions that are being explored is the usage of cloud-based storage. One such platform that Illumina provides is BaseSpace and can be used for both storages as well as analysis. For now, BaseSpace with Amazon Web Services cloud computing is making an effort to minimize the problem of NGS data storage, processing and interpretation [11].

References

1. An Introduction to Next-Generation Sequencing Technology Table of Contents (2015) 1-16.

2. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genet 30: 1-9.

3. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, et al. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523: 486-490.

4. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, et al. (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11: 817-820.

5. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. (2015) Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell 161: 1187-1201.

6. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161: 1202-1214.

7. Gagan J, Van Allen EM (2015) Next-generation sequencing to guide cancer therapy. Genome Med 7: 80.

8. Gonzalez-Garay ML (2014) The road from next-generation sequencing to personalized medicine. Per Med 11: 523-544.

9. Gowder SJT (2013) An updated review of toxicity of bisphenol A (BPA) with special reference to the kidney. Current Mol Pharm 6: 163-172.

10. Greek R (2013) Animal Models in Drug Development, New Insights into Toxicity and Drug Testing, InTech,Croatia.

11. Wandelt S, Rheinländer A, Bux M, Thalheim L, Haldemann B, et al. (2012) Data Management Challenges in Next Generation Sequencing. Datenbank-Spektrum 12: 161-171.