Task: Hamza's research

Abstract:

- To process biological data we need a lot of computational resources and time, both of them are costly for us and we would like to process the data in the least time possible with the least resources available.
- To deal with the enormous and complex biological data, we need efficient ways to analyze that data.
- The target we are aiming for is human genome sequencing within a little unit of time.
- Achieving such a target is challenging but comparing with its results, we will be able to identify strange mutations and the causes of any genetic disease.

Introduction:

- The paper describes the advantages and limitations of current NGS platforms including those using sequencing by synthesis, sequencing by ligation, and real-time sequencing, as well as their significant impact on molecular oncology.
- The main purpose of sequencing the human genome is to obtain valuable information for future care. Genomic sequencing can provide information on genetic changes that can lead to disease or can increase the risk of disease development, even in people without symptoms.
- NGS(the next generation sequencers) enables a simultaneously and massively increased sequencing rate ranging from few gigabases per run to 6000 gigabases and therefore a possible human genome sequencing within 1 week

with only 999 US dollars according to Veritas genomic company (Müllauer, 2017; Goodwin et al, 2016).

- Current NGS is categorized into (1) systems that use sequencing by synthesis chemistry [Illumina® platforms (Illumina®, San Diego, CA, USA), Ion Torrent® platforms (Thermo Fisher Scientific, Waltham, MA, USA), QIAGEN GeneReader® (QIAGEN, Hilden, Germany), Roche® Sequencing platforms (Roche, Pleasanton, CA, USA)] and (2) systems that use sequencing by ligation [SOLiD® (Thermo Fisher, Waltham, MA, USA) and BGISEQ-500® (BGI (MGI) Tech, Shenzhen, China)] allowing short-read sequencing approaches (for review, see Goodwin et al. 2016).
- NGS ranges from the whole-genome sequencing analyzing the totality of the human genome to targeted exome sequencing and finally to focused single genetic alteration assays.
- In this paper, these methods are applied to the human genome noticing the efficiency of each one.

Related work:

- The availability of reference genomes was instrumental in the study of biology. Many competing technologies have been developed to improve the quality and robustness of genome assemblies during the past decade. The 2 widely used long-read sequencing provider Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) have recently updated their platforms: PacBio enables high-throughput HiFi reads with base-level resolution of >99%, and ONT generated reads each one equals 2 Mb. It shows that both PacBio HiFi reads and ONT ultralong reads had their own merits. Further genome reference constructions could leverage both techniques to lessen the impact of assembly errors and subsequent annotation mistakes rooted in each.