Bioinformatics and computational tools for next generation sequencing analysis in clinical genetics

Abstract

Clinical genetics play an important role at discovering ambiguous and rare disease and also search for the best option of treatment for patients .And because of huge amount of sequencing and reference genome , Next generation sequencing make it easy to analysis hundreds of genes and chromosomes at lower price and perfect time(by using PCR technic. This paper aims to show how next generation sequencing do that operation and attached it to bioinformatics . it will focus on Illumina and Ion Torrent and show how they work by using some algorithms and some data analysis . although NGS has provided too much solves to many problems , further improvements in bioinformatic algorithms are still required to deal with complex and genetically heterogeneous disorders.

Introduction

In medical practice to make it easy to diagnose a disease, researchers use genetics to help them discover it quickly. genetics takes a best way to find a solution or treatment to a disease. It's flexibility comes from the ability to deal with genome at different levels and forms from chromosomal to single base alternation.

Paul Berg, Frederick Sanger and Walter Gilbert made possible several progresses in DNA sequencing field making technology named sanger sequencing . Through sanger technology u can generate a huge amount of sequencing in one machine run in a fast and cost effective way . and they considered it as a Next generation sequencing .

In market sanger sequencing profit has reached to billions dollars by 2025 of course because of it's advantages from doing tasks fast in addition to costs little money compared to other machines but still there are companies and labs use illumine , Ion _Torrent sequencing , Pac Bio and Exford Nano Pore.

Using (PCR) plays an integral role in targeted NGS . it produce multiple of targeted regions simultaneously. Next generation sequencing can produce massive quantities of molecules (PCR) but first generation sequencing produce single sequencing without clonal amplifications.

Related work

Science 1953 to 2017 there are lots of changes and events related to Dna sequencing .
In 1935 watson and crick discovered Dna sequencing and in 1976 the first genome sequencing is published and publication of sanger sequencing method happened in 1977 and in 1983 PCR development and in 1986 the first automated sequencer is launched and sequencing of human genome in 2001 and the NGS is launched in 2005 and the first NG sequencer in 2006 (illumine) and pacBio released first real time sequencer in 2011 and in 2015 the oxford nanopore technology sequencer (the pocket-sized) and finally in 2017 SeqLL released SMS platform (single molecule sequencing).

The disadvantage of the past technologies is the time and the large cost they take to produce the required data.so they head to modern technologies which passed by lots of stages to produce data. And it's advantages is they clean data by trimming and filtering technology to increase the quality control of data and at the same time they are fast , more accurate and less cost.
The raw data is go through the primary and secondary analysis before being a useful data .
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7019349/figure/jcm-09-00132-f003/ in this primary analysis they use sensor , server data and torrent basecaller to produce BAM File. The signal that emitted from nucleotide incorporation is attached to sensor that converts the raw data voltage to Data file passing by server data that consider an input to this Data that converts it to WELLS file that consider an input to torrent basecaller that gives final BAM File.
Then comes the secondary analysis stage.

Methodology

NGS platform perform parallel sequencing and that facilitate huge sequencing at any time by using data analysis and some algorithms like base caller using illumina and they focus on reducing base calling errors by 40% and second algorithm is quality control (Read filtering and trimming) and it help to illuminate error . they also uses sequence alignment and in it they searching for pattern or read in the reference using different mappers and aligners and finally they used variant calling and variant filtering and visualization

Results

Although lots of progress that NGS gives to lots of researches , but we still have long journey to diagnose a perfect answer to all genetic diseases . high performance and improving data are still required to reduce error and improve quality . no doubt that NGS has developed but scientists and doctors need to do more to avoid lack of resources . genetics , scientists and doctor need to coordinate with each other for understanding new data and the development of new tools . AI and machiene learning help NGS to improve in the last years.