Abstract

The recent trend of BigData in Healthcare is overpowering and necessity increasing rapidly because of its data type diversity in addition to its volume, managing speed and leads to improving care even at the lowest cost. Cancer prevails as a challenging issue because of its different mutations. Identification of the each tumor's root for mutations and mapping of their evolution of genetics that leads to growth in the conflict against the cancer disease, "GenomeAnalysis "plays an important role. In order to accumulate and categorize the enormous revenue of information from genome analysis, research field coalesced with a data Platform ApacheHadoop supporting parallelization, composability for extremely huge upsurge in activity of sequencing data. By aggregating all aids of BigData Analytics Tools and EHR, this proposal presents a study about how to incorporate the Hadoop Tool integrated with GATK(Genome Analysis Tool Kit) through MapReduce to map cancer genomic data problems with the conscious of financially low cost and high speed of accessing data.

References
1. Aaron McKenna1, Matthew Hanna1, Eric Banks1, Andrey Sivachenko1 Kristian Cibulskis1, Andrew Kernytsky1, Kiran Garimella1, David Altshuler1,2, Stacey Gabriel1, Mark Daly1,2 and Mark A. DePristo1,3. Genome Res. 2010. 20:1297-1303, (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.


3. “Cloudera Chief Scientist Jeff Hammerbacher Teams with Mount Sinai School of Medicine to Solve Medical Challenges Using Big Data.”