Identification of Multiple Cancer Biomarkers through a Comprehensive Approach: A New Era for Personalized Cancer Therapeutics

[1]Dr.Balaji S, [2]Kishore G R, [3]Chandrika K S [4]Chandana N [5]Swamy Manasvi Rao G M [6]Sandhya P S

[1]CIIRC(Center For Incubation ,Innovation, Reasearch and Consultancy)

[23456]Jyothy Institute of Technology ,Bangalore

Abstract-Cancer contributes to be an important global medical and economic burden to humanity. Cancer genomic studies have supported the theory of genomic aberrations and alterations as the hallmark of cancer cells. Since genomic instability due to accumulated mutations has resulted in tumor heterogenity, metastatic progression and therapeutic resistance, we hypothesized that it is highly relevant to understand the crucial gene mutations that are driving the metastatic transformations and tumor reccurence. Based on the information of crucial gene mutations, relevant therapeutic approaches can be implemented as a part of treatment regimen. To better understand the tumor genome complexities, a global approach of sequencing whole cancer genome was proposed. Though whole genome sequence provides important insights on mutation spectra, lack of relevant tools and pipelines to clinically analyse and interpret the technology for cancer prognosis. As part of this study, we are developing a software tool-Multiple Biomarker Identification for Cancer and Genetic Disorder(MBICGD) for identifying multiple data using computational methods. Galaxy toolbox, which analyses the mutation spectra from cancer genomes, will be utilized for developing a relevent pipeline that encompasses the application tools for genome assembly, alignment, variation, calling and annotation. The developed database will be useful for cancer diagnosis, clinical and translational research and academic studies.

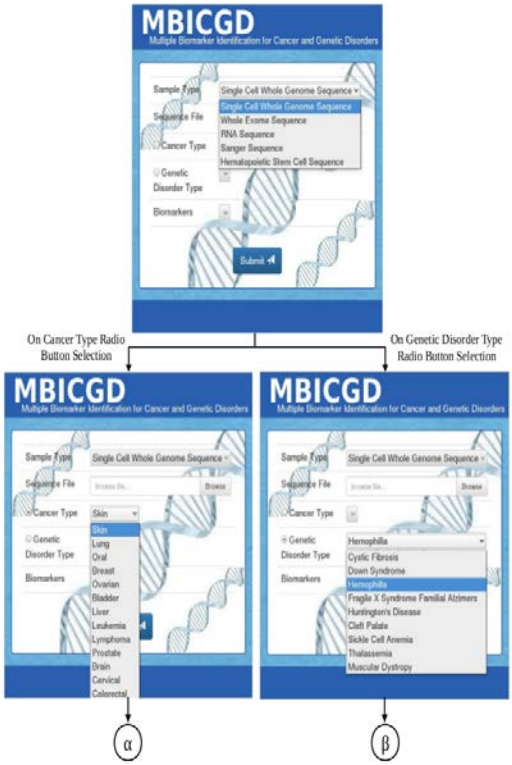
Keywords: MBICGD, cancer genome, genomic instability, gene mutations

I. INTRODUCTION

Aberrant genomic alterations have been the hallmark of cancer cells. Accumulated mutations have resulted in tumor heterogeneity, metastatic progression and drug resistance of specific subpopulation of cancer cells which, confounds with diagnosis and treatment strategies in many tumors. Decoding the whole human genome provides critical insights of the genetic alterations that occur in cancer which can be used as potential biomarkers for detecting the severity of the cancer and form rationale for selective drug treatment. Hence, over the last few years, sequencing and mapping of the whole genome has evolved substantially. Ever since the first human genome was sequenced, the need for the whole genome sequence of number of individuals is increasing with most of the traditional Sanger sequencing being replaced by next generation sequencing(NGS) at reasonable costs. Next generation sequencing technologies have significantly contributed in data generation, deciphering of cancer genome evolution and the dissection of tumor subclones. NGS has made a huge impact in terms of throughput speed, read length and sharp reduction in per base cost sequencing. The European Genome-phenome Archive (EGA) have maintained an archieve of genetic and phenotypic data of different types of cancer and figure 1 shows the graphical representation of meta-data collected from various cancer types using different methodologies. Many cell based technologies have been deployed to understand the contribution of individual biomarkers in cancer progression. However, a more comprehensive approach of using bioinformatics in identifying multiple cancer biomarkers from the meta-data of cancer genome databases is essential for effective cancer treatments. Though progression in DNA sequencing DNA

squencing technolgies have provided the catalogue for sequencing DNA variants among human population, analysis and usuage of massive amount of sequencing data requires computational technologies which is challenging and remains elusive.

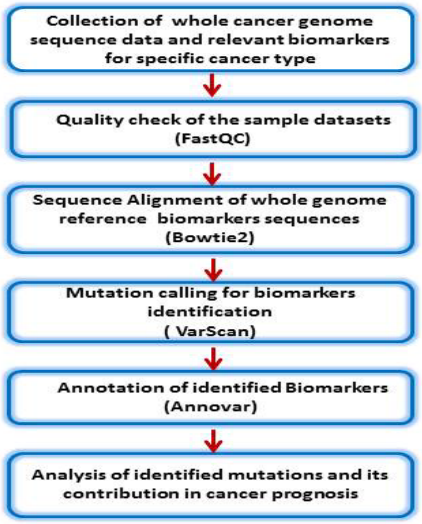
II. SYSTEM MODEL



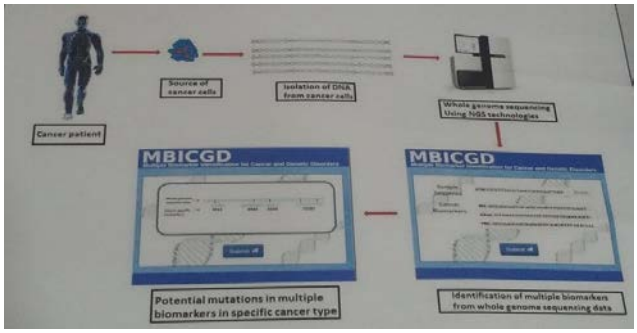
III. PROPOSED METHODOLOGY

- Designing of web portal-Multiple biomarker identification for cancer and genetic disorders(MBICGD)
- Source of whole genome sequence data and the sequences of cancer specific biomarkers
- Development and validation of relevant pipeline for the alignment of multiple biomarkers against whole genome sequence of specific cancer type.

Flowchart of the pipeline developed to identify multiple cancer specific biomarkers.



IV. EXPERIMENTAL RESULTS



V. CONCLUSION

The developed software could potentially trace the mutation patterns of multiple biomarkers and can create a comprehensive map of mutational signatures and their frequencies of occurrences in a specific cancer Apart from identifying cancer specific biomarkers, the software can be customized for identifying other genes for their possible involvement in cancer progression or recurrence.In inherent cancers, familial screening for mutational patterns will provide an insight of high risk group individuals and holds great promise for personalized cancer therapy.

VI. FUTURE SCOPES

One of the important applications of the software MBICGD is in prenatal screening for genetic diseases,which can screen multiple genes associated with any hereditary disease.

With more and more sequencing data available and if we could trace the mutation patterns of multiple biomarkers we could create a comprehensive map of the mutational signatures and their frequency of occurrence in a given type of cancer.

The developed software identifies mutations in all possibleBiomarkers specified in the database Single cell sequencing data also provides informationregarding the clonal evolution of mutations starting from tumor at primary site to distant metastasis.

The developed software will be carefully validated such that it can account for all possible somatic mutations including substitutions, indels, copy number variations,structural rearrangements, and potentially even epigenetic changes (Alexandrov 2014) in the type of cancer being studied.

Apart from detecting cancer specific biomarkers from sequencing data, the software will be customized in future to detect any other gene that may directly or indirectly contribute to tumor growth.

REFERENCES

[1] Meluh pb, koshland d. Evidence that the mif2 gene of saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein cenp-c. Mol biol cell. 1995;6:793–807. [pmc free article] [pubmed]

[2] Mannen h, tseng hm, cho cl, et al. Cloning and expression of human homolog hsmt3 to yeast smt3 suppressor of mif2 mutations in a centromere protein gene. Biochem biophys res commun. 1996;222:178–180. [pubmed]

[3] Johnson es, schwienhorst i, dohmen rj, et al. The ubiquitin-like protein smt3p is activated for conjugation to other proteins by an aos1p/uba2p heterodimer. Embo j. 1997;16:5509–5519. [pmc free article] [pubmed]

[4] Tanaka k, nishide j, okazaki k, et al. Characterization of a fission yeast sumo-1 homologue, pmt3p, required for multiple nuclear events, including the control of telomere length and chromosome segregation. Mol cell biol. 1999;19:8660–8672. [pmc free article] [pubmed]