Available online at www.jpsscientificpublications.com
Volume – 2; Issue - 6; Year – 2016; Page: 856 – 862
DOI: 10.21276.iajmr.2016.2.6.3

Indo – Asian Journal of Multidisciplinary Research (IAJMR)
ISSN: 2454-1370

IDENTIFICATION OF CLINICALLY RELEVANT POLYMORPHISMS IN GLIOBLASTOMA THROUGH INTEGRATED GENOMIC DATA ANALYSIS

G. Vikraman*, A. C. Gomathi and R. Ananthalakshmi,
Department of Biochemistry, Sacred Heart College (Autonomous), Tirupattur - 635 601, Vellore District, Tamil Nadu, India.

Abstract

In the present study, Genomic abnormalities in glioblastoma multiform (GBM) were analysed using Galaxy Next Generation Sequence (NGS). This research particularly describe a robust gene expression based molecular classification of GBM into Proneural, Neural, Classical and Mesenchymal subtypes and integrate multidimensional genomic data to establish patterns of somatic mutations and DNA copy number. Gene signatures of normal brain cell types show a strong relationship between subtypes and different neural lineages. Additionally, response to aggressive therapy differs by subtype, with the greatest benefit in the Classical subtype and no benefit in the Proneural subtype. Then provide a framework that unifies transcriptomic and genomic dimensions for GBM molecular stratification with important implications for future studies.

Key words: Galaxy NGS, Glioblastoma, Multiform, Mutation and Proneural.

1. Introduction

Cancer is a group of diseases characterized by abnormal cellular growth and malignant tissue invasion. It is among the most lethal diseases, leading to deaths of 7.6 million people yearly worldwide, and this figure is expected to rise in the future. Cancer is also a highly complex disease. Tumours develop gradually over years or decades, and harbour mutations in numerous genes. Further, cancer is heterogeneous at three levels, which increases complexity.

Glioblastoma

Glioblastoma multiform GBM is the most common and most severe brain cancer, defined as grade IV astrocytic glioma. Despite intensive treatment regimens including surgery, radiation and drug therapy, median survival is at 12 months. Histologically, GBM is characterized by highly invasive and diffuse tumour mass that frequently undergoes necrosis, i.e., cell death under abnormal conditions. GBM displays heterogeneity both within a tumour and between patients. GBM tumours are classified into two groups based on their mechanism of formation. Primary GBM, which comprise 90 % of cases, occur de novo with no previous diagnosis of lower grade gliomas. In contrast, secondary GBM with 10 % of cases, have a history of lower grade tumours.

Glioblastoma Multiform

Cancers of brain are the consequence of abnormal growth of brain cells. From its origin, brain cancer can be divided broadly into two categories. First are the primary brain tumours. These mainly involve primary brain cells like gliomas, meningiomas, pituitary adenomas, vestibular schwannomas, primary CNS lymphomas and medulloblastomas. Sometimes the

Corresponding author*: **G. Vikraman
Received: 05.09.2016; *Revised*: 10.10.2016;
Accepted: 09.11.2016.

site of cancers might be exclusive of brain but the tumour made of cancerous cells spread into brain by the bloodstream. These are known as metastatic brain tumour or secondary brain tumour. Glioma is a type of tumour that starts in the brain or spine and 60 % of the primary brain tumour detected is gliomas. It derives its name from its association with the glial cells. These can be classified by cell types, grade and location. Gliomas resemble certain histological features with specific cell types, even though they need not necessarily originate from them. Based on this similarity gliomas are categorised as ependymomas (from ependyman cells), astrocytomas (from astrocytes) and oligodendrogliomas (from oligodendrocytes). Among these, astrocytomas are known to be most prevalent in humans. Based on the pathologic evaluation of the tumour, WHO (World health organisation) classified gliomas into four grade types:

Grade I: This stage is referred when the tissue is benign and in the least advance stage. Best prognosis can be provided during this phase.

Grade II: This stage is referred when the gliomas are well differentiated (not anaplastic). Even though they are not benign, there are chances of better prognosis of patients.

Grade III: This stage consists of anaplastic gliomas, which are highly malignant. Carrying out a prognosis is extremely difficult.

Grade IV: The disease exists in its most advance stage and the prognosis is worse in this stage.

Galaxy Next Generation Sequence (NGS) originally written for biological data analysis, particularly genomics. Galaxy is the tool for analysis the particular sequence and the accuracy of the result. When compare to the other tool the error report less than 3 % and the set of the work flow can set automatically. Here, the availability of the sequences are easily annotated and run the galaxy tool and given the results too fast because the set of data will be run based on the work flow. The set of available tools has been greatly expanded over the years and galaxy is now also used for gene expression, genome assembly, proteomics, epigenomics, transcriptomics and host of other disciplines in the life sciences. Glioblastoma Multiform (GBM) is most common form of brain cancer in adults. Patients with GBM have a uniformly poor prognosis with a median survival on one year (Ohgaki and Kleihues, 2005). Malignant Gliomas are the most prevalent type of primary brain tumor in adults. They constitute approximately 50 % of all Central Nervous Tumors (Barcizewska *et al*., 2014). Direct Cell - Cell Communication synchronizes groups of cells to execute coordinated programs required for growth, differentiation and therapeutic response (Naus and Laird, 2010).

2. Materials and Methods

Gene of Interest

Get Data from EBI-SRA

First we selected the Control sequence from EBI-SRA database SRX098415. The control sample sequence from the study Alu Scan: a method for genome - wide screening of sequence variations in the human genome. This sequence made by Illumina Genome Analyzer II and the 896.7 M bases and also get the tumor sequence from the same database SRX099037. Tumor sample sequence from the study of Aluscan: A method of genome - wide screening of sequence variations in the human genome. The sequence made from Illumina Genome Analyzer II. Here 1.1 G bases presented.

NGS: QC and Manipulation

FASTQ Manipulation

Highly configurable complex manipulations can be performed on selected FASTQ reads by using the Manipulate FASTQ reads on various attributes tool. This tool allows the user to define a set of matching criteria to be used to select the reads in a FASTQ file on which to perform a set of manipulations; any number of match directives can be defined and a read must match each directive to be considered for

manipulation. Matching is currently limited to user specified pattern matching (regular expressions) on sequence identifier/name, sequence content and quality score strings, with defaults set to match all. However, additional matching and manipulation options can be easily implemented as needed. When a read does not match, it will be transferred to the output in an unmodified fashion. Reads that pass all matching criteria are subjected to any number of user-specified manipulations. Manipulations are available that act upon sequence identifier/name, sequence content or quality score strings. Beyond allowing the user to remove matching reads or to perform string translations on any of these attributes, additional manipulations are available for sequence content, including: reverse complementing, reversing (without complementing), complementing (without reversing), trimming, *in silico* transcription of DNA to RNA and vice-versa, as well as changing the adapter base within colour space sequences. Additionally, separate tools exist that can convert FASTQ files to-and-from a tabular format. This allows FASTQ data to be modified using any of the powerful text manipulation tools, which are pre-packaged with Galaxy.

FASTQ Sanger grooming on the data

FASTQ groomer check the uploaded sequence to any gap penalty included and the sequence variants are presents as the origin and the sequence quality to maintain up to the output. Groomed 7246492 sager reads. Based upon quality and sequence, the input data is valid for: sager Input ASCII range: '#'(35) - 'D'(68) Input decimal range: 2 – 35.

FASTQ Groomer

The FASTQ Groomer tool is used to verify and convert between the known FASTQ variants. The data created by this tool was guaranteed to conform to the target variant specified by the user, including the enforcement of quality score minimums and maximums. After grooming, the user is presented with some information about the input such as ASCII character and decimal value ranges and a list of FASTQ variants for which the input data is actually valid. Although, the output created by this tool is now valid, if the user has selected the wrong presumed input variant, it is possible for the resultant score values not to reflect the values intended by the sequencing technology. Users should utilize the provided summary information as a sanity check before continuing with their analysis; for example, if a user provides a Sanger encoded variant (with ASCII values <59), but specifies the input variant as Solexa, this summary information would state that the input was valid only for Sanger.

FASTQ Summary Statistics

FASTQ format is a text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores. Both the sequence letter and quality score are each encoded with a single ASCII character for brevity. It was originally developed at the Welcome Trust Sanger Institute to bundle a FASTA sequence and its quality data, but has recently become the *de facto* standard for storing the output of high-throughput sequencing instruments such as the Illumina Genome Analyzer. The quality of the groomed sequence will be notice and drawn from the galaxy will be establish below. The groomed sequence will be high level of the percentage to match the control sequence.

Quality statistics

As quality scores can vary along the length of sequencing reads, determining how to trim and filter read data involves calculating summary statistics on a per column (base position) basis. The FASTQ Summary Statistics by column tool accomplishes this task. The output of this tool contains read counts, minimums, maximums, sums, means, quartiles with ranges, outliers and nucleotide counts for each base position in a FASTQ file. This statistical summary can be graphed by using the Boxplot tool, found under the Graph/Display Data tool section.

Read Trimmer

To prevent otherwise high - quality reads from being rejected assembly processes, it can be beneficial to trim bases from poor - quality ends of reads. The FASTQ Trimmer by column tool allow trim either end of a set of reads by using absolute offsets or by specifying percentage of read length based offsets. Offsets begin at 0 for each end and increase towards the opposing end of the read. For example, to trim the outer 3 bases from each end of a 36 length sequencing read, a user can specify absolute 5 and 3 offsets of 3 or percentage based offsets of 8.33 ($0.0833 \times 36 = 2.9988$, rounded to the nearest integer = 3).

Filter Fast Q

This tool allows the user to filter out reads that have some number of low quality bases and return only those reads of the highest quality. For example, if we wanted to remove from our dataset all reads where the quality of any base was less than 20, we change the Minimum Quality box to 20. The "Maximum number of bases allowed outside of quality range" allows selecting how many bases per read must be below our threshold before we discard that Read. If we leave this at the default setting of 0, all bases in a read must pass the threshold Minimum quality score to be kept. 1. Trim the reads in the GM12878 dataset using the Generic FASTQ manipulation FastQ Trimmer tool. Determine from the boxplot and FastQC figures where the quality of the reads begins to drop off sharply. Calculate how many bases have to be trimmed from the end and use that number as the Offset from 3' end. 2. Using the Generic FASTQ manipulation Filter FastQ tool, filter out all sequences with any bases that have a quality less than 20.

NGS Mapping

There are two mappers available in Galaxy: Bowtie and BWA. The crucial difference between these mappers is that BWA performs gapped alignments, whereas Bowtie does not (although there is a version of Bowtie available which does perform gapped alignments, it is not the one available in Galaxy). This, therefore, gives BWA greater power to detect indels and SNPs. Bowtie tends to be much faster, and have a smaller memory footprint, than BWA. BWA is generally used for DNA projects, whereas Bowtie is used for RNA-Seq projects since the exonic aligner TopHat uses Bowtie to do the initial mapping of reads.

Burrows - Wheeler Alignment

The control and tumour sequences are aligned and merged the following steps and alignment is must to read the base pairs in the edited data. So, we use the BWA tool. The quality controls are checked by using the NGS: QC and manipulation FASTQ Summary Statistics tools.

3. Result and Discussion

The high percentage of supportive TAM in GBM tumor mass makes it possible to be a good tumor mass makes it possible to be a good target for GBM treatment. In this case, the unique features of TAMs in GBMs, including their origin, the tumor - suppressive properties, the secreted cytokines and relevant mechanisms (Wenchar Zhou and Shidney Bard, 2014). Epidermal growth factor receptor (EGFR) is a membrane-bound receptor that has been shown to have a major role in the pathogenesis and progression of different cancers. The frequent amplification of the EGFR gene in glioma was initially reported in 1985 (Qun-Ying Yang *et al*., 2015).

The PTEN gene received its name through its identification on chromosome 10 in a loss of heterozygosity region often mutated in patients with high - grade tumors (Chow and Baker, 2006). Another means of inducing GBM tumors in mice is the suppression of neurofibromatosis type 1 gene (NF1) and Trp 53 gene, located on chromosome 11 in mice (Reilly *et al*., 2004). The Ras pathway, like the EGFR or MET pathways is a signalling pathway used within the cells themselves. When both Ras and the complementary signalling pathway Akt are over expressed in lab mice, glioblastoma tumors develop with similar histological features to

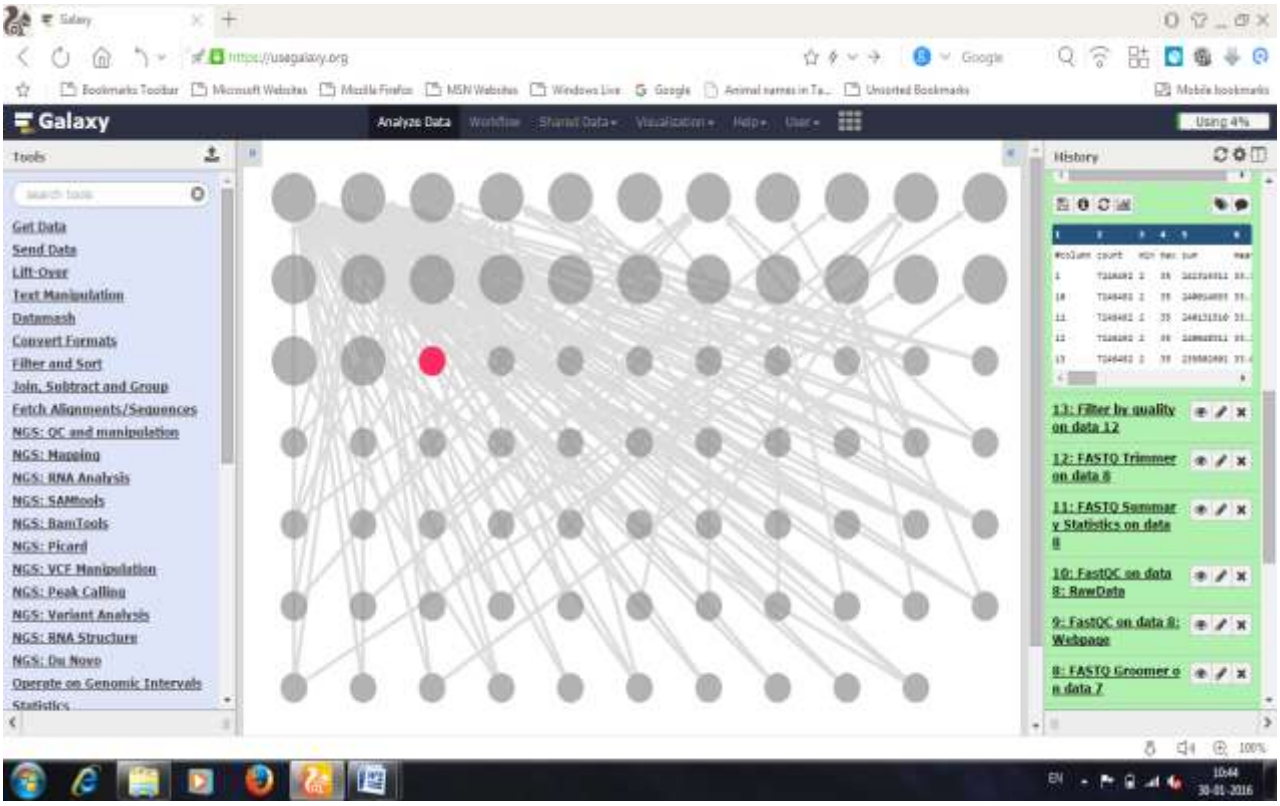

Figure - 1: GALAXY NGS result in overview of expression sequences

The followed graph showed the mutation variations in the tumour sequence, the red high peak line mention the mutation gene. The first mutations found to be in PTEN, CTNND2, MGMT, NF1 & EGFR. In this genes are mutated and the main precursor for the Glioblastoma formation in adult brain CNS and multiform in the mutations.

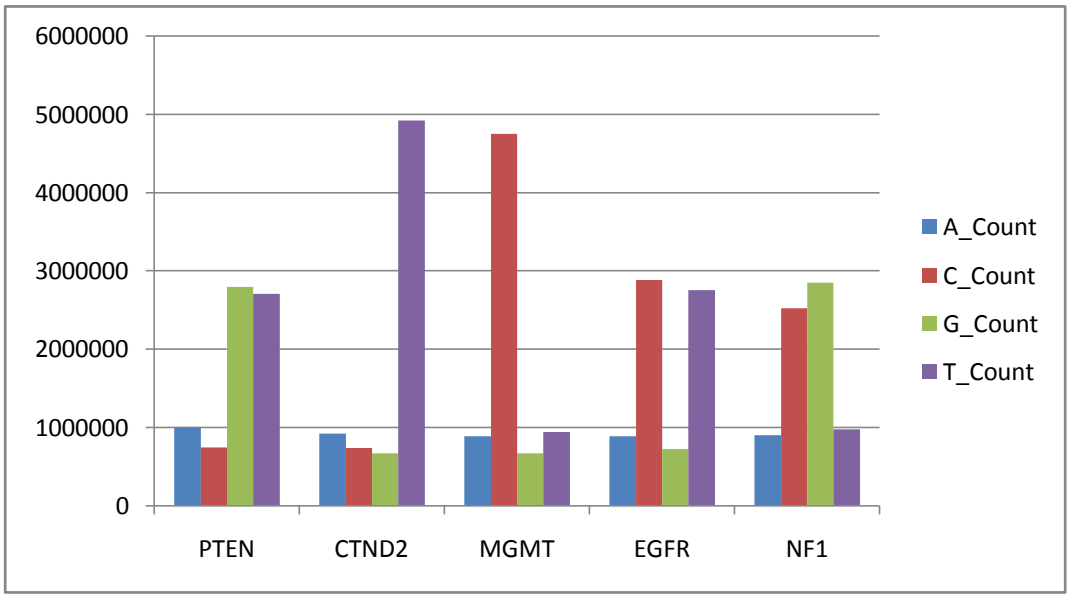

Figure - 2: Expressed genes are shown in the graph

Figure - 3: Groomed tumour sequence of glioblastoma

human GBM (Holland *et al*., 2000). Recent insights into tumor metabolism have further revealed that oncogenic signalling pathway directly promote metabolic reprogramming to up regulate biosynthesis of lipids, carbohydrates, protein, DNA and RNA leading to enhanced growth of human tumors. Therefore, targeting cell metabolism has become a novel direction for drug development in oncology. In malignant gliomas and lipid are significantly reprogrammed. Moreover, molecular mechanisms carrying these metabolic changes are just starting to be unrevealed (Pens *et al*., 2013). The international classification of human tumors published by the world health organization (WHO) was initiated to establish a classification and grading of human tumours that is accepted and used worldwide. Without clearly defined histopathological and clinical diagnosis criteria (David and Ohgaki, 2007). Finally we got the compared sequences of the glioblastoma whole genome sequences display the variations in between these two sequences shown below. The first identified mark shows the first mutation in the whole genome sequence and then the followed mutations has been developed and mentioned.

4. Conclusion

The present study reveals the tumour genome contain the mutated genes of PTEN, CTND2, MGMT, NF1 & EGFR if mutated, these genes can cause single glial cell into multiple cells and hence apoptosis process was not taken place. Brain tumours develop from cancerous glial cells and are called gliomas. Unlike other cancers, glioma tumours grow in the confined space inside the head. In order to grow, most cancers push healthy cells aside, but due to space constrains, glioma tumours must destroy normal brain cells. To kill healthy nerve cells, glioma tumours release large quantities of neurotransmitter glutamate. Excess glutamate is toxic to neurons and causes seizures in up to 80% of people with gliomas. Depending on the tumor's size and location, other symptoms include paralysis, behaviour changes and dizziness. A glioma tumor is particularly damaging because it tends to quickly sprout and spread within the brain.

5. Reference

1) Christian C. Naus and Dale W. Laird. 2010. Implications and challenges of connexin connections to cancer. *Nature*, 25: 55 - 57.
2) Gursel, D. B and K. M. Reilly. 2014. Intratheal gene therapy corrects CNS pathology in a model of Musco polysaccharidosis. *Nature*, 1525 - 0024.
3) Holland, E. C., J. Celestino, C. Dai, L. Schaefer, R. E. Sawaya and G. N. Fuller. 2012. *In vivo* models of primary brain tumours: pitfalls and perspectives. *Journal of Neurobiology*, 5 (2): 156 – 162.
4) Archer, K., V. Mas, K. David, D.Maluf, K. Bornstein and R. Fisher. 2009. Identifying genes for establishing a mutagenic test for hepatocellular carcinoma surveillance in hepatitis C virus positive cirrhotic patients. *Cancer Epidemiology Biomarkers and Prevention*, 18 (11): 2929 - 2932.
5) Kleihues, P and H. Ohgaki. 2006. Primary and secondary glioblastomas: L.M. Chow and S. J. Baker. PTEN function in normal and neoplastic Growth. *Cancer Letters*, 241 (2): 184 - 196.
6) Yanging Zhu, Jessica Begal, Patrica O Donnel, Traces Sikora, Therese ruane, Ping Wang, Mark E. Haskins and James M. Wilson. 2014. Intractal Gene Therapy corrects CNS mucopolysaccharidosis I. *Nature*, 1525 - 1534.