DOI: 10.55522/jmpas.V11I3.2930 ISSN NO. 2320–7418

International peer reviewed open access journal

Journal of Medical Pharmaceutical and Allied Sciences

Journal homepage: www.jmpas.com CODEN: JMPACO

Research article

RNA-Seq analysis to explore the variants in melanoma cells: molecular diagnosis and therapeutics

Shalini kumari, Ruchi Yadav*

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India

ABSTRACT

High throughput sequencing technology that are also called as second generation or Next Generation (NGS) sequencing techniques has enabled researchers to study genome, transcriptome, metabolome of any organism in high throughput manner. RNA seq (Ribonucleic Acid Sequencing) is a NGS technique that is used to sequence total transcripts of cell and to study gene expression. This technique is widely used to identify differentially expressed genes and to identify variants. RNAseq technique has been used to study several diseases like cancers, neurological diseases, bacterial infections diseases and to understand the key mechanism of genes and its functions. Melanoma is a threatening tumor and one of the most successive metastatic diseases. Melanomas ordinarily happen inside the skin yet inside the mouth, digestion tracts or eye. Melanoma carcinoma is also called as cutaneous melanoma or melanoma of the skin. In current research Pair end RNA-seq sequencing data for melanoma cell was retrieved from ENA (European Nucleotide Archive) database with accession no.:SRP252675. RNAseq analysis pipeline of Galaxy online platform was used for the prediction of single nucleotide variations (SNVs). Total three genes are predicted that are expressed in RNAseq samples and involved in the skin cancer these genes are TNFRSF4 (Tumor Necrosis Factor Receptor Superfamily Member 4), TNFRSF18 and AGRN (Agrin). Protein encoded by TNFRSF4 gene is a member of the TNF-receptor superfamily and AGRN gene is associated with Presynaptic Congenital Myasthenic Syndromes. Pathway enrichment of identified genes shows that TNFRSF4 and TNFRSF18 have function in cytokine-cytokine receptor interaction and AGRN in ECM-receptor interaction. These results highlight the importance of TNFRSF4, TNFRSF18 and AGRN in Melanoma condition and can be further used as potential drug targets.

Keywords: RNAseq, Transcriptome, Melanoma, Differentially Expressed Genes, Galaxy.

Received - 30-01-2022, Accepted- 24-05-2022

***Correspondence:** Dr. Ruchi Yadav ✉ ryadav@lko.amity.edu, **Orchid Id**: https://orcid.org/0000-0002-0971-2667

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India.

INTRODUCTION

RNA seq is high throughput sequencing technique that is used to study gene expression in different conditions and sate of cells [1]. RNAseq technique is used to sequence the total transcript of cells and to identify gene expressions and its variants in different conditions like environmental, disease, development, infections, cultures etc. of cells [2]. NGS technology uses the approach of sequencing short fragments of DNA in high throughput parallel fashion and sequence billions of DNA fragments in single run of NGS machines [3]. NGS technology has made possible to understand, classify, and uncover complexity of genome by discovering new genes, transcripts, variants etc. [4] RNAseq technology has been used to study transcripts of different cells in different condition and most importantly to identify genes that are expressed in diseases [5]. This advantage of RNA seq technology has made it most popular method to identify differentially expressed genes, to study complex pathways and system of diseases [6].

Different technology has been used since decade to study gene expression like microarray technology, SAGE (Serial Expression of Gene Expression) [7]. Microarray uses hybridization of fluorescently labelled probes to identify gene expression and quantification. This technique is based on microarray chip that are spotted with known feature probes and based on hybridization, further data acquisition, gene expression is measured [8]. Microarray result is thus dependent upon the fluorescence dyes, lasers, scanners, and other experimental equipment's [9]. Since gene expression is not directly measured from the sequence data there is lots of possibility of error in the microarray result that can be incorporated due to biasness in the dyes, lasers, chip surface etc. [10]. Also, major drawback of microarray technology is that it can be used to quantify and study only know genes and novel genes cannot be identified. In comparison to traditional approach of studying transcripts and due to drawbacks and limitation of microarray technology, RNAseq has

overcome this traditional approach to study transcriptome [11].

RNAseq is based on sequencing cDNA directly and quantification is done based on read count. This approach of transcriptome analysis is a boon to predict and identify novel genes in different cells [12]. RNAseq is widely applied to study developmental biology, neuro development diseases, cancers, different classes of cancers to identify genes involved and to predict potential drug targets, markers etc. [13]

With the advent of high through put sequencing technology like illumine sequencing, SOLID, pyrosequencing, deep sequencing it is possible to sequence entire transcripts of genome parallelly [14]. NGS sequencing generates high throughput results in the form of read sequences and produces billions of reads fragments. Researchers has used RNAseq technology to study transcripts, various functional regions of transcripts like 5'and 3' UTR region, promoter, exons, introns etc. [15]. It is revolutionary tool study small RNAs like short hairpin RNAs, miRNA, siRNA, shRNA etc. This technique has enabled us to explore all possibility to understand genome complexity at cell, tissue, organs, and organ system level [16].

Hence, there is requirement of exhaustive computational analysis to identify differential genes, predict novel genes, and identify functional pathways and its association with function of genes. With the help of computational software's and tools NGS data can be analyzed and results can be obtained [17].

NGS analysis requires a complete and comprehensive workflow that can be used to identify genes, mutations, SNPs in the data [18]. Downstream analysis is done to identify gene function, pathways, Gene Ontology, etc. of identified genes. Currently number of tools and software's are there that can be used for NGS analysis of these Galaxy server is most widely used because of its availability, user friendly interface, data handling and graphical outputs [19]. Number of workflows has been designed and published for RNAseq analysis using Galaxy server, researcher must carefully use tools in each step of analysis [20]. Here we provide easy and complete protocol for RNAseq analysis using tools of GALAXY server. For this current research Melanoma data was retrieved from ENA database which is freely available online database and can be used to retrieve high throughput experimental data like RNAseq data, ChIpSeq data, Exome seq, etc.

Melanoma is most serious type of skin cancer that is caused due to mutation in melanin genes that result in uncontrolled division in melanocytes [21]. Research in Melanoma has been widely done to identify different classes of variation like mutations, SNPs in skin cells. Melanoma can of be different types and it can spread from outer body parts to internal organs also [22]. There can be different reasons of Melanoma like genetic reasons due to family history, radiations, life style, environmental conditions etc. RNA seq approach has been used to identify variations in melanoma cells and to identify mutations in genes. Genes like ABC transporter gens (ABCB5), FOXA1 have been identified from RNA Seq transcriptomic analysis [23]. In this research we have used melanoma RNAseq data to identify variations in genes and functional analysis has been done to study gene functions

MATERIALS AND METHODS

RNA-seq data retrieval

Pair end RNA-seq sequencing data for melanoma cell was retrieved from ENA database. Two FASTQ format files were downloaded for two different gene sample with Accession no.: - SAMN14372165. The RNAseq sample files in Fastq format were downloaded from ENA database using the SRA accession Id. Here, the link through which we get the data from ENA: www.ebi.ac.uk/ena/browser/view/PRJNA612430. Samples files selected for current research has been shown in table 1.

Table 1: RNAseq data with study accession number PRLNA612430 used for current research

Study Accession	Sample Accession	Experiment Accession	Run Accession	FASTQ FTP
PRLNA612430	SAMN14372165	SRX7906582	SRR11301252	SRR11301252.fastq.gz
PRLNA612430	SAMN14372164	SRX7906583	SRR11301253	SRR11301253.fastq.gz

Quality control (QC)

It is first and preliminary step to study the quality of read files. It stores the quality of all reads per base in Fastq file. Different tools are available that can be used to access the quality of NGS data that is calculated based on quality scores per base [24].QC analysis is performed for the accuracy, completeness, relevancy, validity, and consistency of the data. **MultiQC** tool [25] can be used to compile all the result of FastQC generated from different samples into one single file. MultiQC result was used for comparative analysis.

Binary alignment map (BAM) file [26]

A BAM file (*.bam) is the compressed binary version of an alignment file. BAM file is generated by mapping read sequence to reference genome. BAM file contains all the information of read sequence and alignment. It stores the read sequence information like read sequence, quality score of reads, tags information and alignment data like positions of alignment, alignment scores, alignment quality, start and stop positions of reads. BAM files also details about chromosomal mapping coordinates.

Variant call format (VCF) file format [27]

VCF file is obtained from further analysis and variants identification tools like Free Bayes. VCF results includes variations data like SNPs, mutations in read sequences. VCF provides stores variants data in the form of table that includes mutated bases (A, T, G, C), positions of mutation, read sequence, and mapping information. Variants are calculated by alignment of reads sequence

with genomes stored in databases like 1000 genomes projects. To create a VCF file- Free Bayes tool of Galaxy was used BAM file were used as input in Free Bayes tool to identify variants in sample files.

Count file

Count file was generated for all RNAseq data, it summarizes read counts at the different functional region of genomes like exon, UTR, promoter, introns etc. Count file stores number of reads sequences that align at same positions and by counting all reads that share common region of alignment in reference genome. Galaxy provides different tools that can be used for this purpose. Feature count tool was used to count number of reads sequence that maps to same location in genomes and part of same genes, exons, etc. Feature counts tool provide the features of gene, read count and statistical scores. Bam file was used as input for Feature Count tool [28].

Annotation

It is most important part of data analysis that provide details about genes function, pathways, gene ontology etc. Data annotation is the process of adding metadata to the dataset. To get the annotation of the data search- annotateMyIds tool in the galaxy tools panel was used. Using the count file perform the annotateMyIds tool for both the sample file. The result having duplicates. So, to remove the duplicates perform annotation again by selecting no duplicates option. Now the result of annotation was the set of metadata of the data [29].

String database

It is most widely used databases to identify protein interaction and to study network functions. STRING database can be used to predict protein interaction along with functional enrichment of network like pathways, Gene ontology enrichment. Top 50 genes selected from the annotation result was used for STRING database for functional and process enrichment. Clustering was done to identify the cancer-causing genes [30].

RESULTS AND DISCUSSION

Galaxy server was used for complete RNAseq analysis from Quality control to prediction of variants in fastq files. QC analysis of samples files were done using FastQC tool and quality graphs were studied. Preprocessing of raw data was further done before variant analysis for knockout gene sample. FastQC analysis was done for both the sample files used for current study that is SRR11301252 and SRR11301253. For demonstration purpose and explanation of quality graphs only FastQC result of SRR11301252 has been shown. Quality graphs was shown in figure 1(a-d) and figure 2 (e-h).

(a)Adapter content

This plot identifies the contamination of read sequence with the adapter sequence. It plots the reads sequence in the fastq file and shows the location of adapter sequence. In good data this adapter position should lie at the base of the plot as shown in (figure 1 (a)). This analysis is specific to sequencing type and based on NGS machine that is used

(b) Duplication level

This plot is important to identify the reads duplicated in the sample files. Duplication plot shows the high and low abundant transcripts in the Fastq file. There can be chances that duplication of reads sequences arises in the samples due to biasness in PCR amplification or if read sequences is sequenced multiple times. In current sample file this plot shows that there is less level of duplication level (red line) (figure 1 (b)).

(c)Per base N content

This data provides the positions in read sequence where base call by sequencing machine is poor hence instead of any base N is mentioned. This result can be used to identify the quality of read sequence since higher the number of N in read sequence lower the quality. In goo sample data the base call line is below zero and should not raise beyond it as shown in figure 1 (c).

(d)Per base sequence quality

This result shows the quality of each base in read sequence and per base quality control plot was generated between combined quality score statistics at each position along all the reads sequences in the Fastq files. This plot calculates five statistical values that is minimum, lower quartile, median, upper quartile and maximum value and plots it into the form of boxplot. Here, this plot is a good plot as the quality score is good and passes the standard values of per base quality (figure 1 (d)).

(e)Per base sequence content

This plot reports the percent of bases called for each of the four nucleotides at each position across all reads in the file. Here, in this plot, all the four nucleotides % are shown which rises in the start and then became persistent towards X-axis as shown in figure 2 (e). In good NGS data it should be uniform across all reads, and it also depends upon the sequencing type and kit used for sequencing.

(f)Per base sequence GC content

This plot shows the number of reads v/s GC% per reads. It shows the GC percentage across all reads sequences which should be uniform. Here, the result of this plot is right as the GC count per read is nearby to theoretical distribution (figure 2 (f)).

(g)Per sequence quality scores

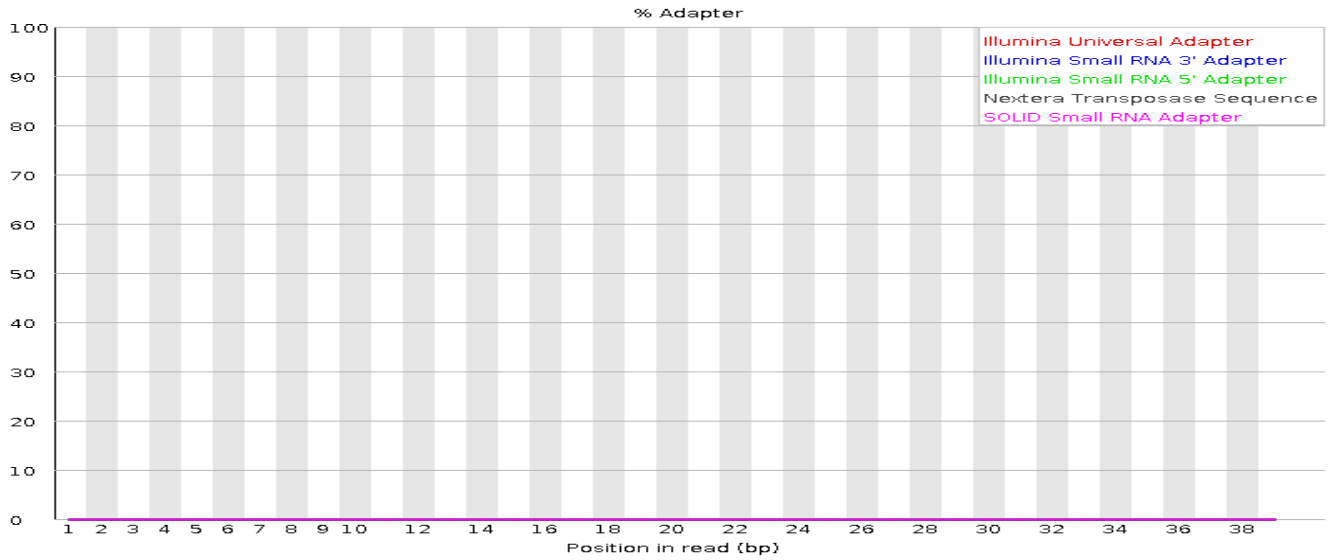
This result shows per read sequence quality and plot the total no. of reads v/s the average quality score over full length of that read. Here, in this plot the average quality per read is constant in the beginning but highly increased at the end (figure 2 (g)).

(h)Sequence length distribution

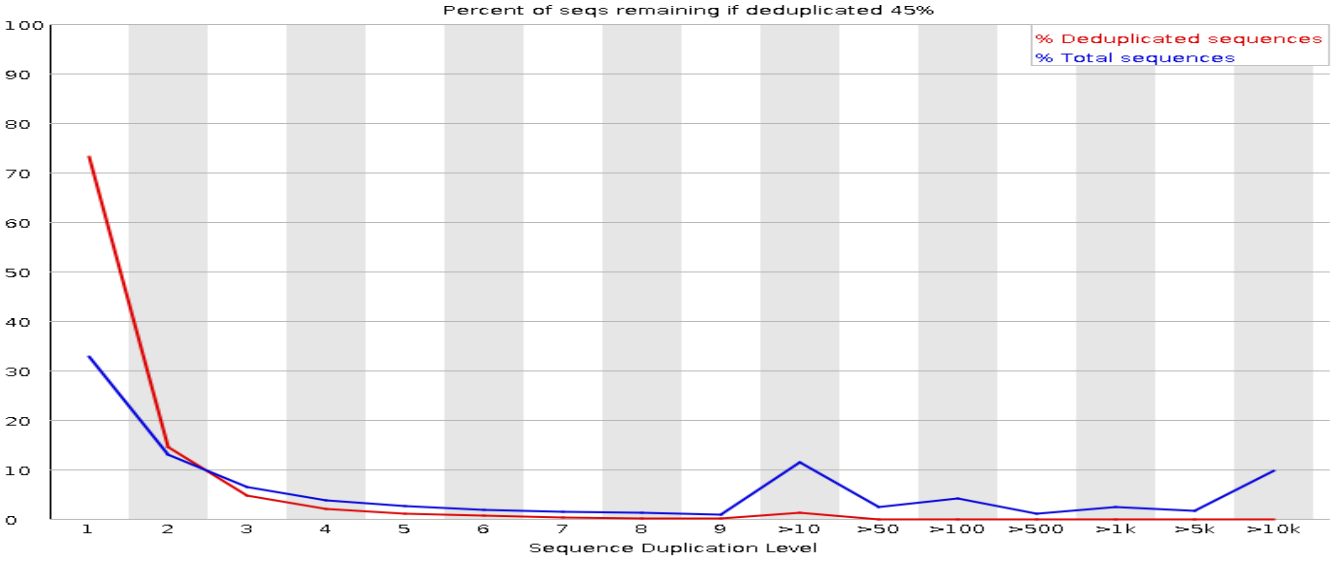
This details the dispersal of read fragment sizes in the sample files across fastq which was further analyzed. Current plot of sequence length distribution it shows that the read sequence length which increases at the end (figure 2 (h)).

Figure 1: Fastq graphs as obtained from FastQC tool it shows the quality of reads samples by calculating different scores. For demonstration Fast QC result of SRR11301252 sample has been shown (a) adapter content. (b) Duplication level. (c) Per base N content. (d) Per base quality

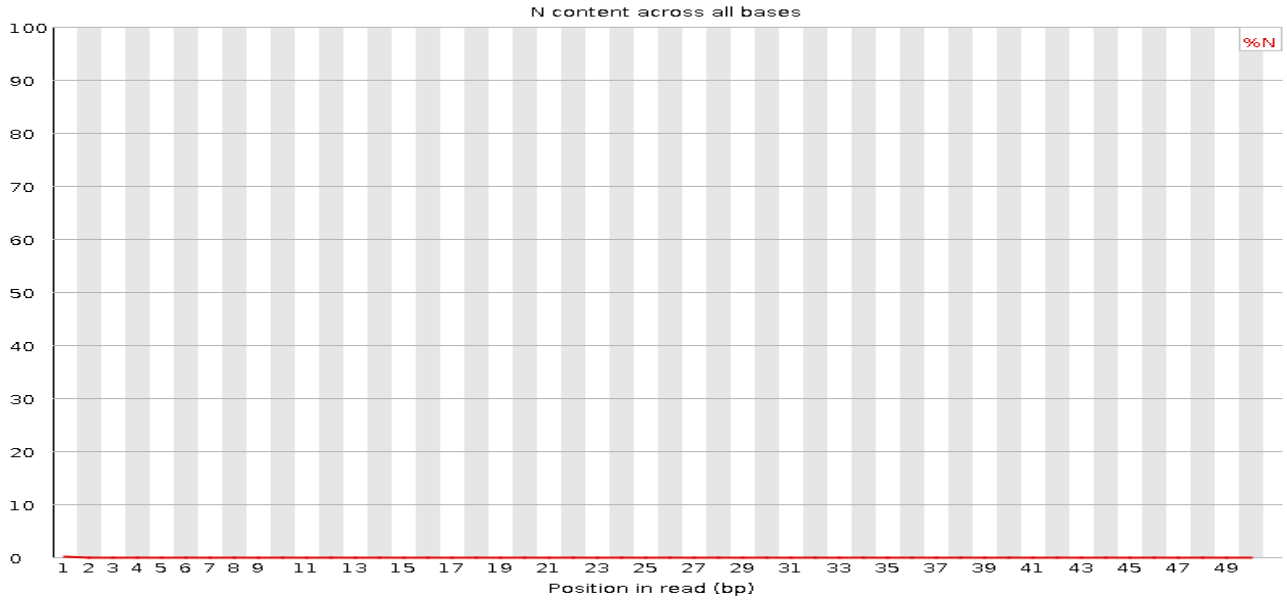
(a)



(b)



(c)



(d)

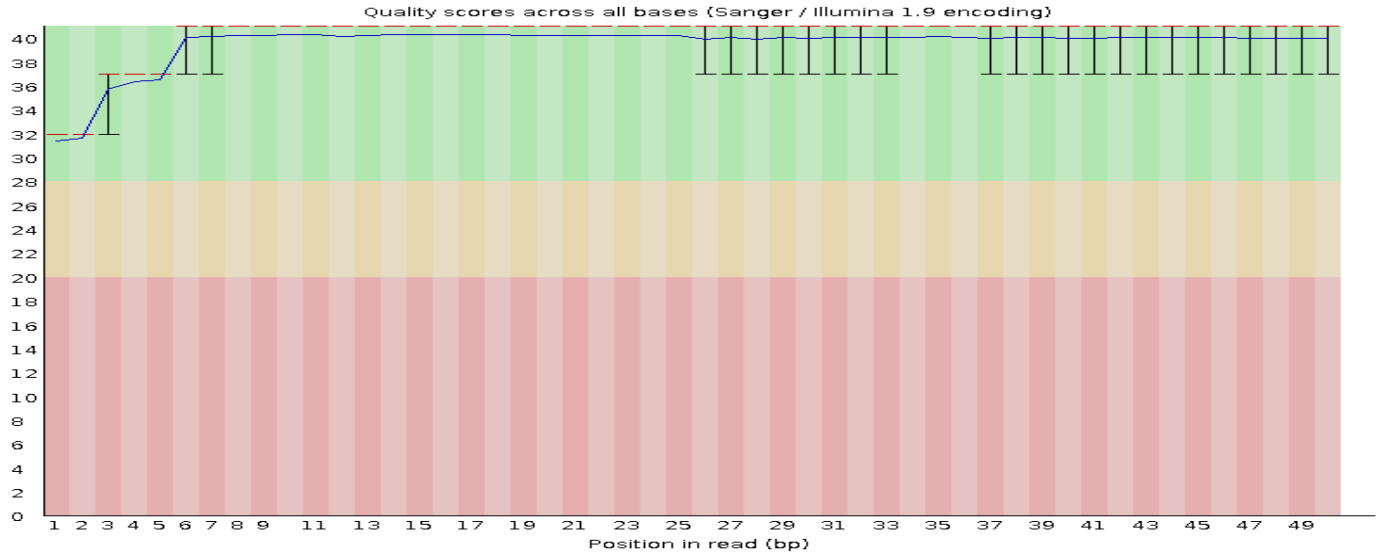
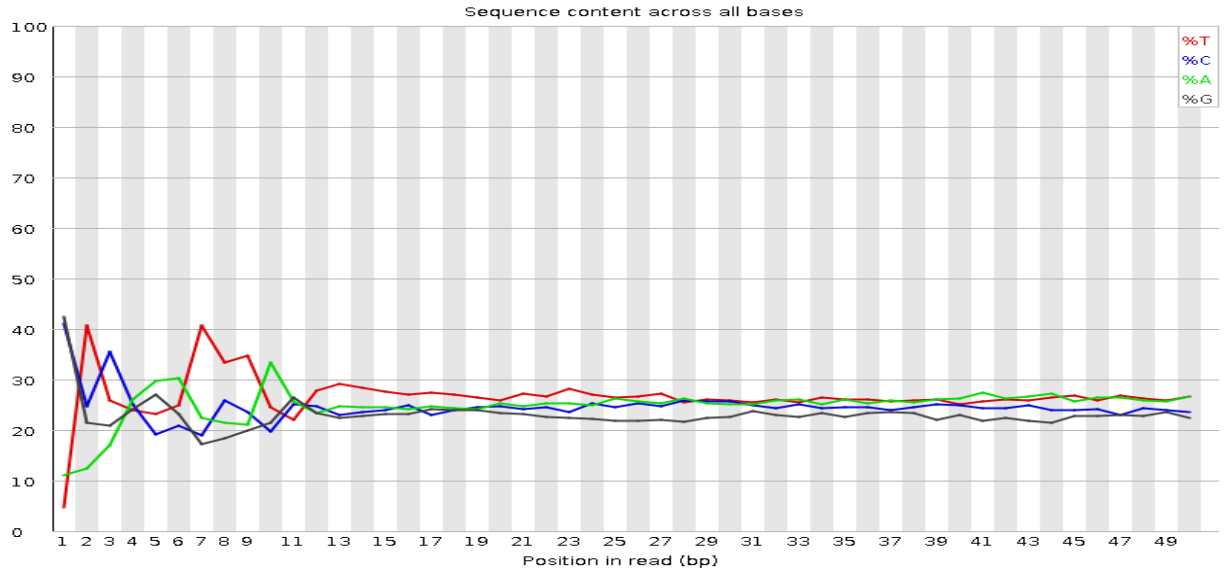
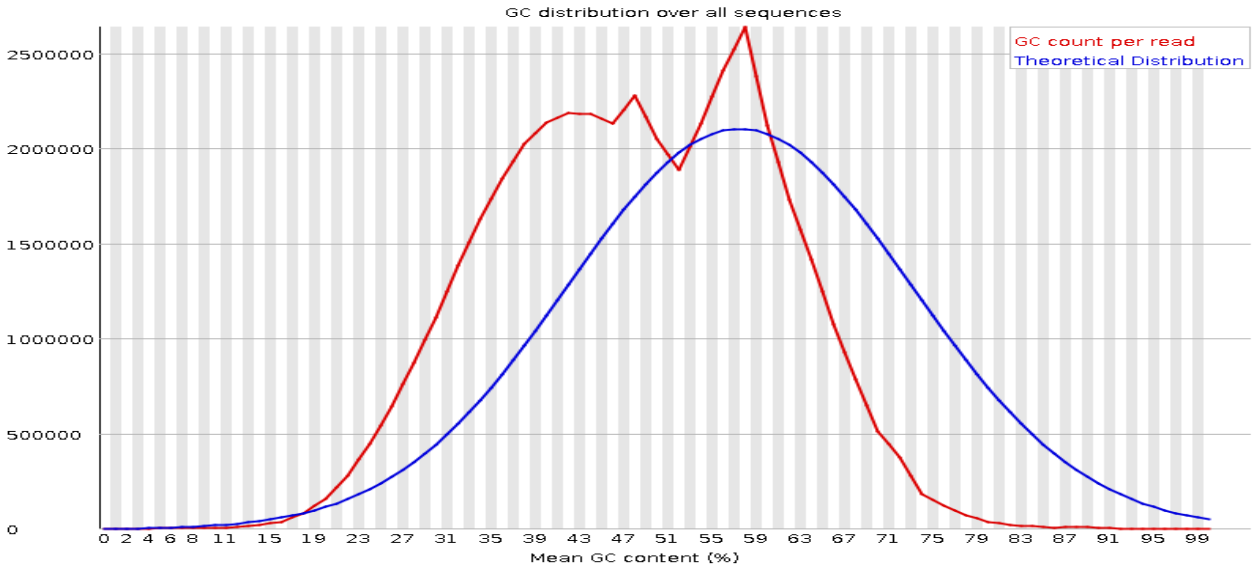


Figure 2: FastQC Graphs of SRR11301252 reads file (e) per base sequence content (f) per sequence GC content (g) per sequence quality and (h) sequence length distribution

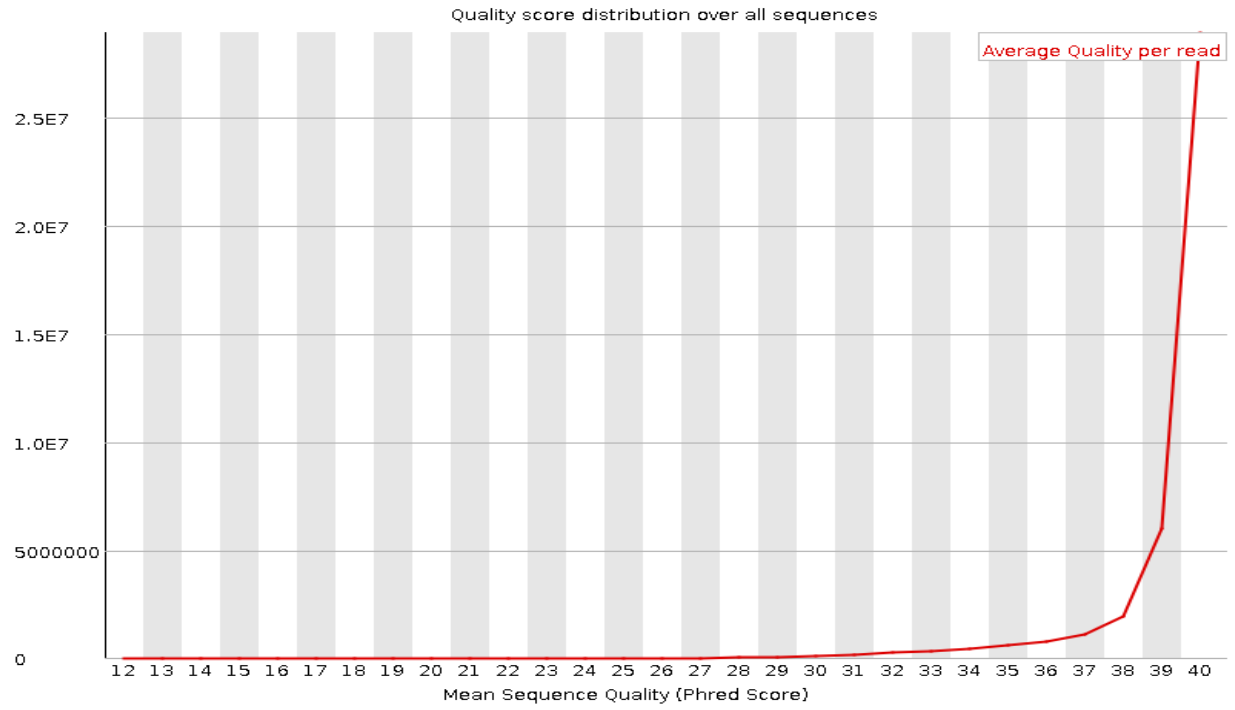
(e)



(f)



(g)



(h)

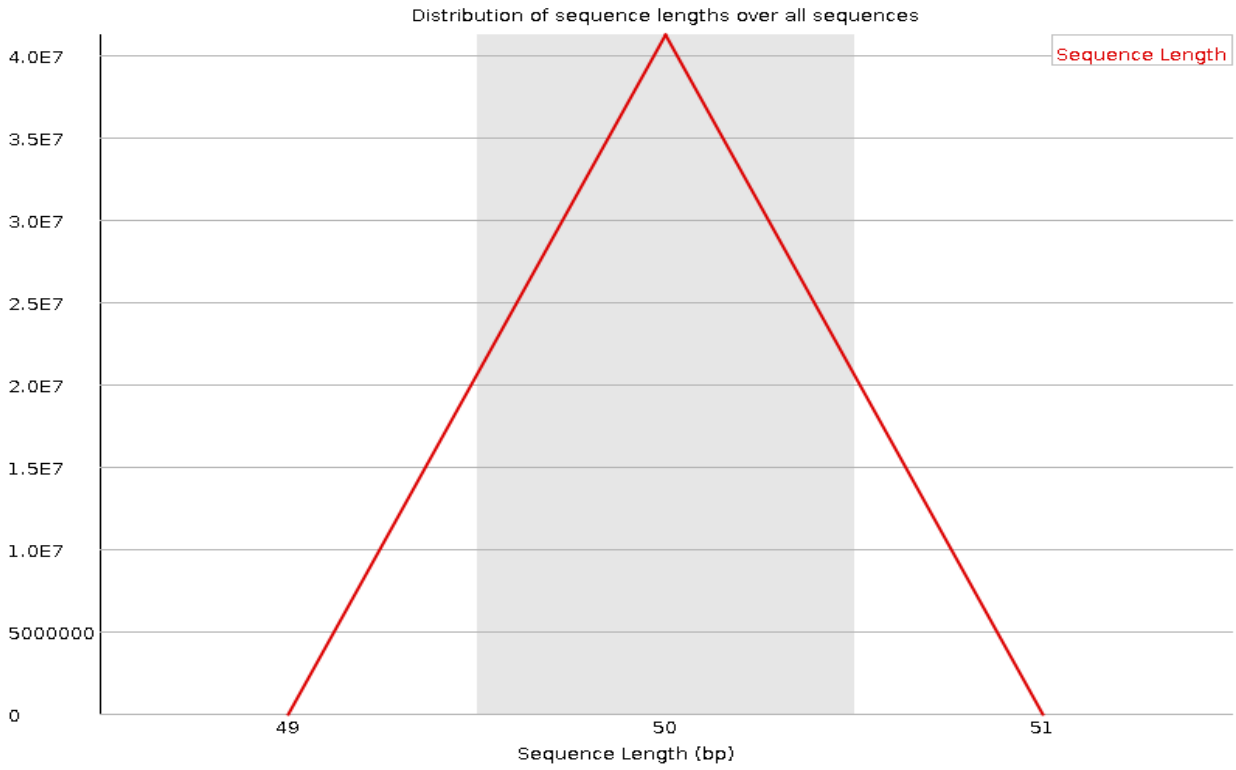


Figure 1 (a-d) & 2 (e-h) give detail about the quality of the sample files. These graphs contain, adapter content, duplication level, sequence content, GC content etc. through which we can analyze whether the data is good for the analysis or not.

Binary Alignment Map (BAM) of Sample file

After FastQC the alignment of reads sequence with the reference genome was done. This step aligns the reads sequence with complete genome and alignment can be visualized in different genomic browsers like UCSC genome browser (genome.ucsc.edu/). This UCSC graph of BAM file shows the gene of interest that is tumor necrosis factor receptor superfamily member 4 (TNFRSF4) alignment which means it is present in the sample file. Mutation in TNFRSF4 gene is associated with the melanoma. Alignment of TNFRSF4 present in sample file with the reference genome has been shown in figure 3.

Figure 3: TNFRSF4 gene visualized in UCSC browser (genome.ucsc.edu/) using BAM result.

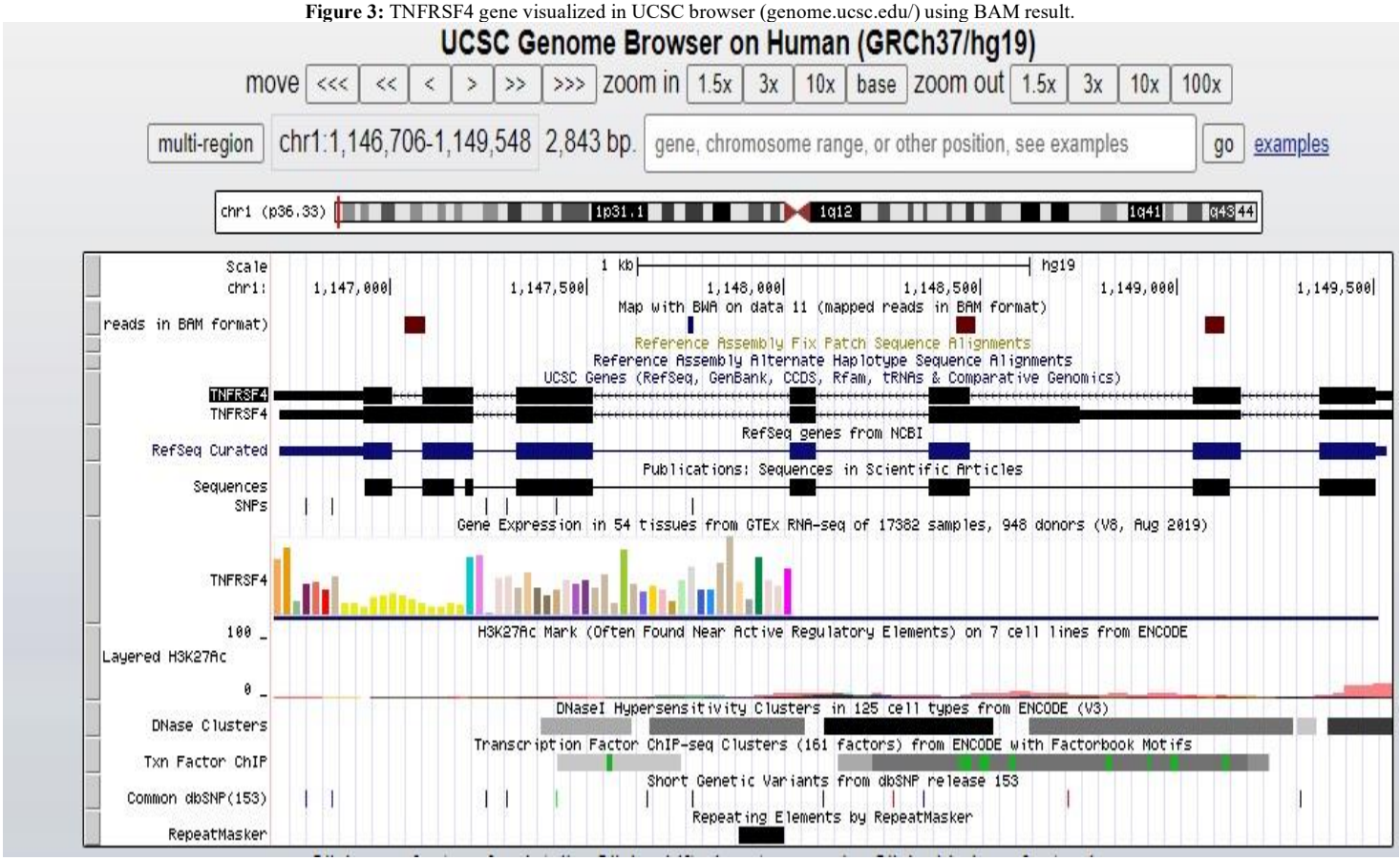


Figure 4: TNFRSF18 gene visualized in UCSC browser (https://genome.ucsc.edu/) using BAM result.

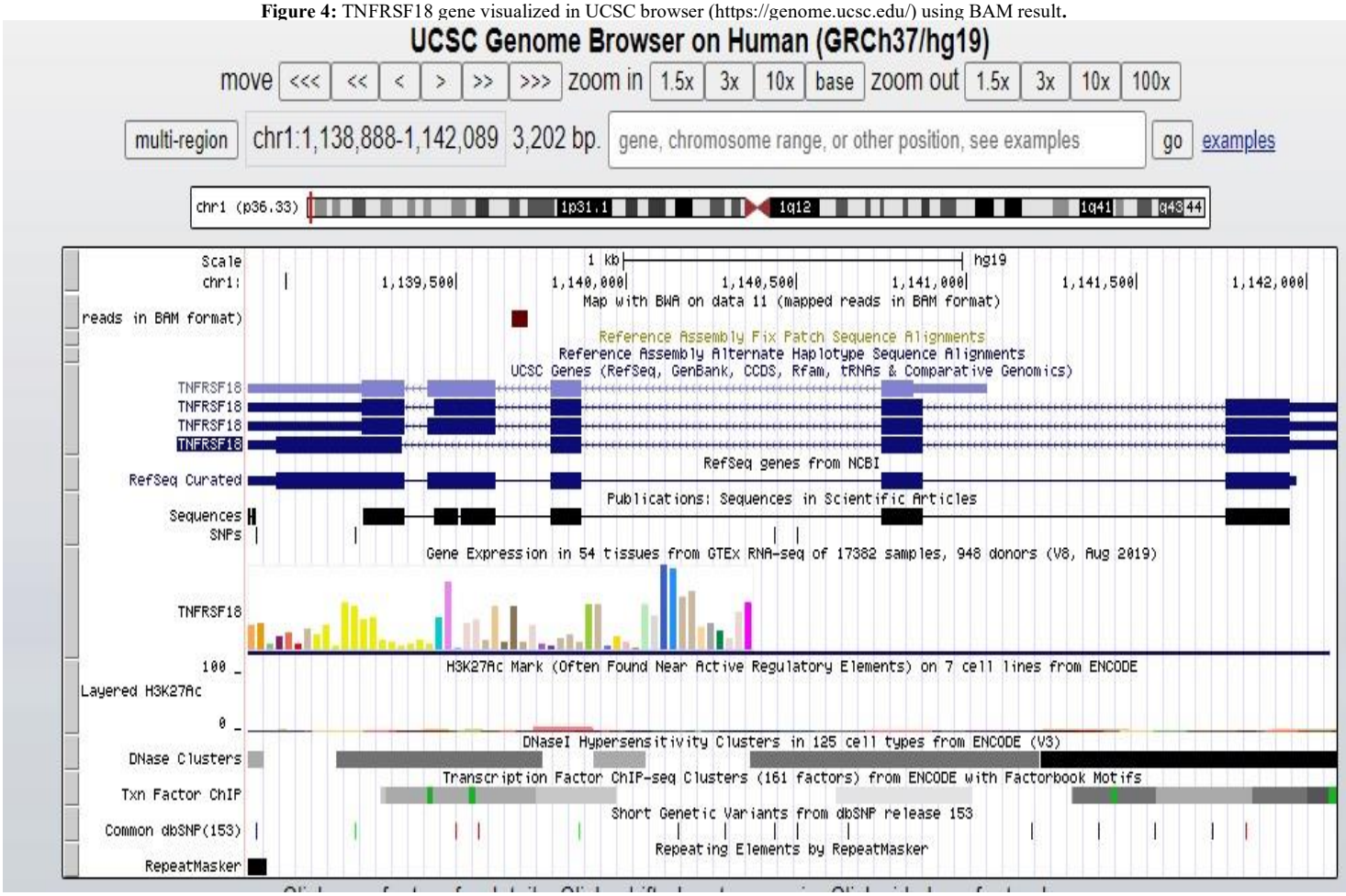
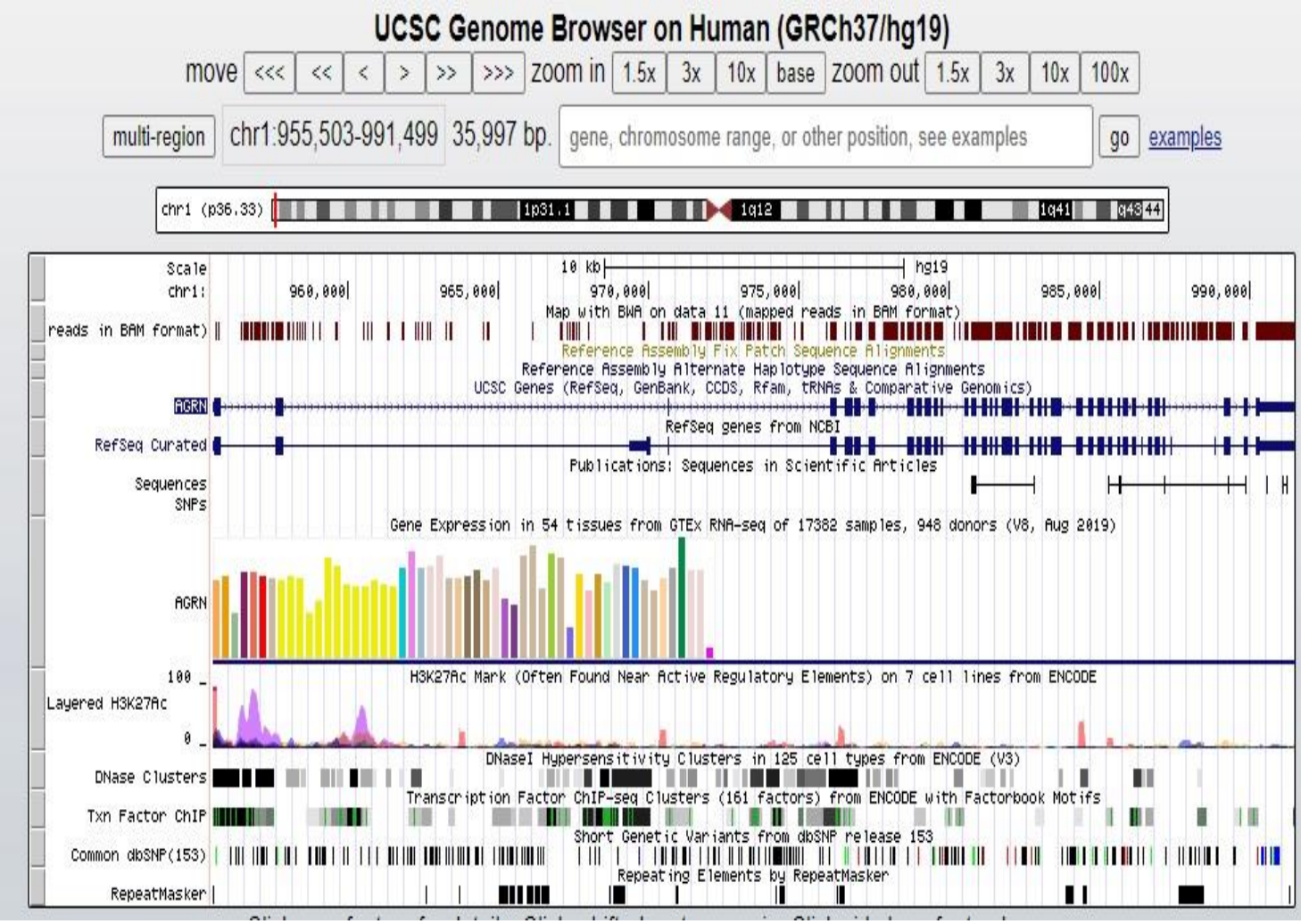


Figure 5: AGRN visualized in UCSC browser (https://genome.ucsc.edu/) using BAM result.



This UCSC graph of bam file shows another gene of interest that is TNF Receptor Superfamily Member 18 (TNFRSF18) alignment which means it is also present in the read file and it is also involved in the melanoma genome. Result of BAM alignment of TNFRSF18 gene has been shown in figure 4.

BAM alignment of another important gene that is Agrin (AGRN) as shown in figure 5. This alignment has also been produced from the alignment of reads file with the reference genome and visualized in UCSC genome browser.

Variant Call File (VCF) of melanoma cell RNAseq

Variants were identified as per methodology already discussed in methodology section. The VCF file was generated that stores the information about any mutation or SNPs in the sample file. VCF file can be visualized in different genomic browsers and variants can be studied in gene of interest. List of variants were visualized in iobio database (iobio.io/) result of VCF is shown in figure 6. It shows the number of alignments between reads and genomic sequence across all chromosomes, variant density in each chromosome, frequency of allele, variant quality, bases change, variant type that is SNP, insertions, deletions, or other types of mutations. This VCF data also describes the variant length, chromosomal location that can be annotated and further studied.

Annotation of Variants in SRR11301252 file

After variant analysis the VCF file was analyzed for further annotation and functional enrichment. There are 25703 Gene Id's in VCF file out of which top 25 Gene Id's were mentioned in the table 2. List of genes identified from annotate My Id and feature count. We perform the feature count of the data for Gene Ids, and number of gene involved in the sample files. Also, we perform the annotate My Id tool for the annotation of the data that annotates the reference ids as mentioned in VCF file and generated annotation table that includes, Gene Id, Gene No, Symbol, Gene Name, Gene Ontology and Pathway for the knockout gene.

Annotation of Variants in SRR11301253 file

The entire methodology was also done for another sample that is SRR11301253. There are 25703 Gene Id's in VCF result file out of which top 25 Gene Id's are mentioned in the table 3. Annotation was done for all 25 genes and annotation table was generated with following information that is Gene Id, Gene No., Symbol, UniProt ID and Pathway for the wild type.

Figure 6: Variants result visualized in vcf.iobio.io database (//iobio.io/).

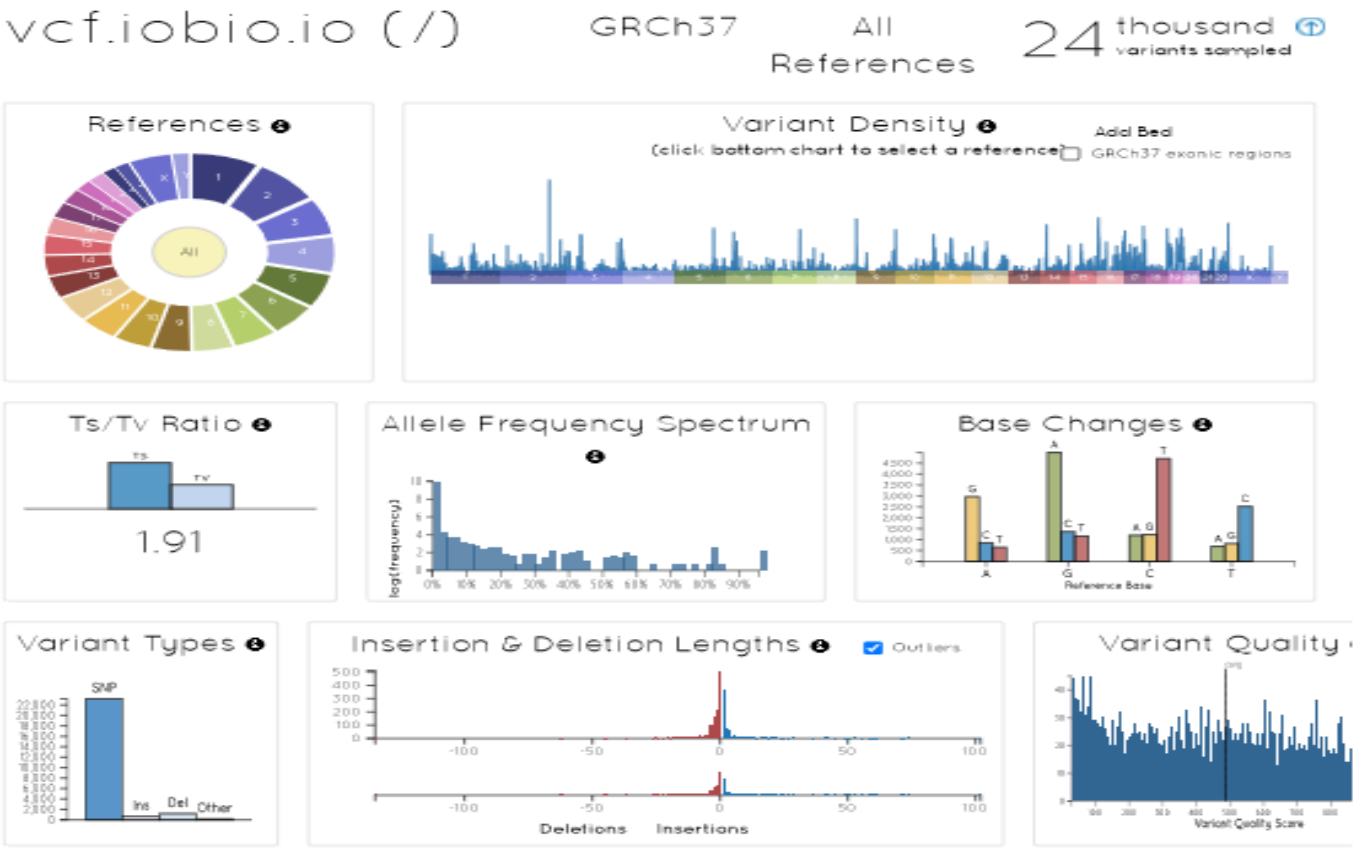


Table 2: List of genes its function and gene ontology (GO) of all the genes expressed in samples files SRR11301252 and SRR11301253. These genes have been predicted by using feature count & annotateMyId tools of Galaxy server.

ENTREZID	GENE SYMBOL	GENENAME	GO
653635	WASH7P	WAS protein family homolog 7, pseudogene	NA
100422834	MIR1302-10	microRNA 1302-10	GO:0035068
645520	FAM138A	family with sequence similarity 138 member A	NA
79501	OR4F5	olfactory receptor family 4 subfamily F member 5	GO:0004930
729737	LOC729737	uncharacterized LOC729737	NA
100507658	NA	NA	NA
100132287	LOC100132287	uncharacterized LOC100132287	NA
100288646	NA	NA	NA
729759	OR4F29	olfactory receptor family 4 subfamily F member 29	GO:0004930
100131754	NA	NA	NA
81399	OR4F16	olfactory receptor family 4 subfamily F member 16	GO:0004930
100287654	NA	NA	NA
100133331	LOC100133331	uncharacterized LOC100133331	NA
100288069	LOC100288069	uncharacterized LOC100288069	NA
100287934	NA	NA	NA
400728	FAM87B	family with sequence similarity 87 member B	NA
79854	LINC00115	long intergenic non-protein coding RNA 115	NA
643837	LINC01128	long intergenic non-protein coding RNA 1128	NA
100506327	NA	NA	NA
284593	FAM41C	family with sequence similarity 41 member C	NA
284600	LOC284600	uncharacterized LOC284600	NA
100130417	LINC02593	long intergenic non-protein coding RNA 2593	NA
148398	SAMD11	sterile alpha motif domain containing 11	GO:0005634

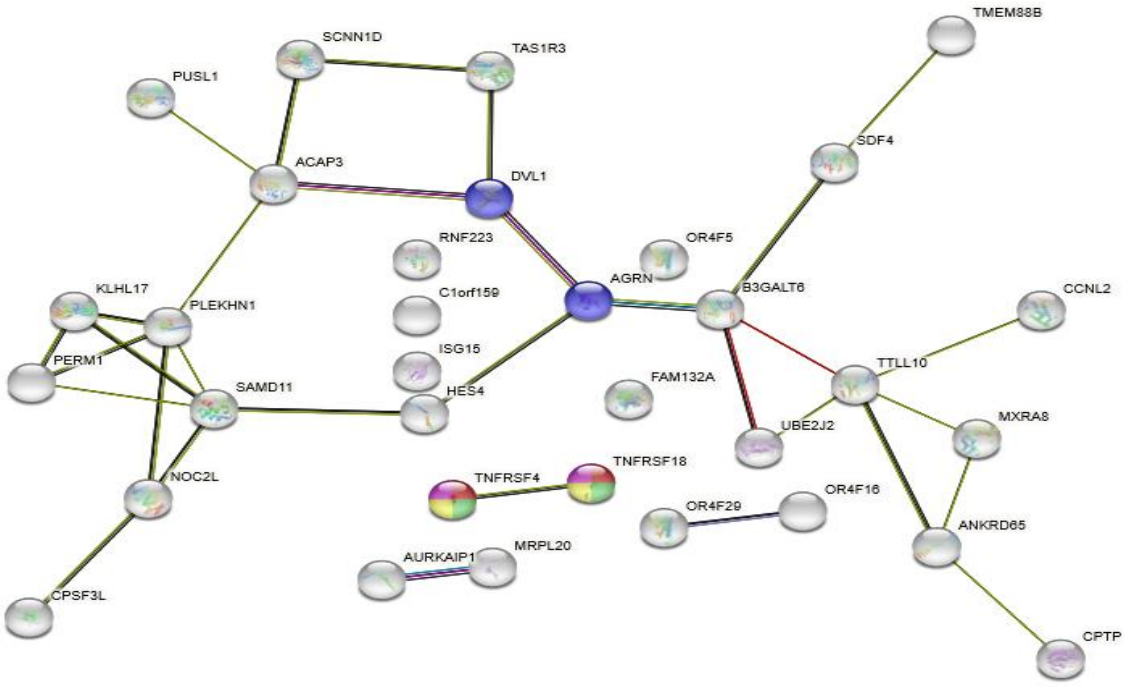
Protein- Protein interaction and Functional Enrichment

Protein- Protein interaction (PPI) of top 50 genes identified after annotation as shown in table 2 were used to study PPI using STRING database. Functional enrichment of proteins was done and Gene Ontology of all the predicted proteins were studied. The result of string was shown figure 7 it shows strong interaction between TNFRSF4 & TNFRSF18 and between AGRN & DVL1 proteins. These proteins are involved in various research related to cancer and causes of cancer [31-33].

Table 3: List of genes its UniProt ID and Pathway of all the genes expressed in samples files in SRR11301252 and SRR11301253. These genes have been predicted by using feature count & annotateMyId tools of Galaxy server

ENTREZID	SYMBOL	UniProtKB ID	Pathway
653635	WASH7P	C4AMC7	NA
100422834	MIR1302-10	NA	NA
645520	FAM138A	Q9NWS6	NA
79501	OR4F5	Q8NG98	Olfactory transduction
729737	LOC729737	NA	NA
100507658	NA	NA	NA
100132287	LOC100132287	NA	NA
100288646	NA	NA	NA
729759	OR4F29	Q6IEY1	Olfactory transduction
100131754	NA	NA	NA
81399	OR4F16	Q6IEY1	Olfactory transduction
100287654	NA	NA	NA
100133331	LOC100133331	NA	NA
100288069	LOC100288069	NA	NA
100287934	NA	NA	NA
400728	FAM87B	P51843	NA
79854	LINC00115	NA	NA
643837	LINC01128	NA	NA
100506327	NA	NA	NA
284593	FAM41C	Q5VWP2	L-proline biosynthesis
284600	LOC284600	NA	NA
100130417	LINC02593	Q86TU6	NA
148398	SAMD11	Q96NU1	NA

Figure 7: Shows the genes interaction network of top 50 genes as mentioned in table 2. Proteins highlighted in colors have function in melanoma.



AGRN genes encodes for the protein that has function in neuromuscular cells and its development. AGRN have important role in neuronal signaling particularly in differentiation of synaptic. This gene has several isoforms that plays key role in formation of neuron cells and its structure in brain regions [34]. Variations of this genes has been found to be associated with neuromuscular diseases like Myasthenic syndrome which shows symptoms such as weakness in muscles, abnormal brain function etc [35].

DVL1 gene is also associated with neuronal function and have role in development and functioning of neuronal cells. DVL1 genes shows function in cell development, differentiation, segmentation etc. Pathway associated with gene is Wnt signaling pathway that is important for cell development, cell interactions, proper functioning of brain and have key role in developmental stages also. Mutation in this gene can result in syndrome like Schwartz-Jampel syndrome and Robinow Syndrome [36].

CONCLUSION

RNAseq data of Melanoma has been used to identify genes that have function in melanoma cancer. Three genes which were identified that have function in melanoma cells are: -TNFRSF4,

TNFRSF18 and AGRN. Protein interactions result shows that there is strong interaction between TNFRSF4 & TNFRSF18, AGRN with the DVL1 gene. TNFRSF4 & TNFRSF18 genes belong to family of Tumor Necrosis Factor (TNF) and have role in immunity. These gene show increased expression upon activation of T-cells and play a key role in self-defense mechanism through apoptosis mediated signaling pathway. TNF family receptors like TNFRSF4 has been widely used as drug targets and explored in medicine industry but studies are still in clinical trial phase [38]. Result shows that TNFRSF4 & TNFRSF18, AGRN and DVL1 gene are important genes that can be considered as potential drug targets and can be considered for further wet lab verification.

ACKNOWLEDGMENT

We would like to acknowledge Amity Institute of biotechnology, Amity University Uttar Pradesh, Lucknow campus for providing us facilities to conducting this study. This research project is not funded by any specific grant from funding agencies in the public, commercial, or non-profit sectors.

DECLARATIONS

Funding: No funds were provided for this research

Conflict of Interest: Authors do not have any conflict of interests.

Ethical approval: Not applicable to this research as it is purely computational based study

REFERENCES

1. Wang J, Dean D C, Hornicek F J, et al, 2019. RNA sequencing (RNA-Seq) and its application in ovarian cancer. Gynecologic oncology, 152, 194-201.
2. Simoneau J, Dumontier S, Gosselin R, et al, 2021. Current RNA-seq methodology reporting limits reproducibility. Briefings in Bioinformatics, 22, 140-145.
3. McDermaid A, Monier B, Zhao J, et al 2019. Interpretation of differential gene expression results of RNA-seq data: review and integration. Briefings in bioinformatics, 20, 2044-2054.
4. Birnbaum K D, 2018. Power in numbers: single-cell RNA-seq strategies to dissect complex tissues. Annual review of genetics, 52, 203-221.
5. Sudhagar A, Kumar G, El-Matbouli M, 2018. Transcriptome analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: a comprehensive review. International journal of molecular sciences, 19, 245.
6. Zheng J, Wang K, 2019. Emerging deep learning methods for single-cell RNA-seq data analysis. Quantitative Biology, 7, 247-254.
7. Moncada R, Barkley D, Wagner F, et al, 2020. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nature biotechnology, 38, 333-342.
8. Rai M F, Tycksen E D, Sandell L J, et al, 2018. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. Journal of Orthopaedic Research®, 36, 484-497.
9. Wang J, Dean D C, Hornicek, F J, et al, 2019. RNA sequencing (RNA-Seq) and its application in ovarian cancer. Gynecologic oncology, 152, 194-201.
10. Tarca A L, Romero R, Xu Z, Gomez-Lopez N, et al, 2019. Targeted expression profiling by RNA-Seq improves detection of cellular dynamics during pregnancy and identifies a role for T cells in term parturition. Scientific reports, 9, 1-13.
11. Romero J P, Ortiz-Estévez M, Muniategui A, et al, 2018. Comparison of RNA-seq and microarray platforms for splice event detection using a cross-platform algorithm. BMC genomics, 19 ,1-14.
12. Stark R, Grzelak M, Hadfield J, 2019. RNA sequencing: the teenage years. Nature Reviews Genetics, 20, 631-656.
13. Van den Berge K, Hembach K M, Soneson C, et al, 2019. RNA sequencing data: Hitchhiker's guide to expression analysis. Annual Review of Biomedical Data Science, 2, 139-173.
14. Luecken M D, Theis F J, 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology, 15, e8746.
15. Li W V, Li J J, 2018. Modeling and analysis of RNA-seq data: a review from a statistical perspective. Quantitative Biology, 6, 195-209.
16. Ji Q, Zheng Y, Zhang G, et al, 2019. Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. Annals of the rheumatic diseases, 78, 100-110.
17. Fan X, Dong J, Zhong S, et al, 2018. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell research, 28, 730-745.
18. Moreno P, Huang N, Manning J R, et al, 2021. User-friendly, scalable tools and workflows for single-cell RNA-seq analysis. Nature methods, 18, 327-328.
19. Fattore L, Ruggiero C F, Liguoro D, et al, 2019. Single cell analysis to dissect molecular heterogeneity and disease evolution in metastatic melanoma. Cell death & disease, 10, 1-12.
20. Cursons J, Souza-Fonseca-Guimaraes F, Foroutan M, et al,2019. A gene signature predicting natural killer cell infiltration and improved survival in melanoma patients. Cancer immunology research, 7, 1162-1174.
21. Hogan S A, Levesque M P, Cheng P F, 2018. Melanoma immunotherapy: next-generation biomarkers. Frontiers in oncology, 8, 178.
22. Wee S K, Yap E P H, 2021. GALAXY Workflow for Bacterial Next-Generation Sequencing De Novo Assembly and Annotation. Current Protocols, 1, e242.
23. Wolff J, Rabbani L, Gilsbach R, et al, 2020. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Research, 48, W 177-184.
24. Lee C T, Maragkakis M, 2021. SamQL: a structured query language and filtering tool for the SAM/BAM file format. BMC bioinformatics, 22, 1-8.
25. Manisha S, Anil S, Vikas N, 2021. Supramolecular drug carriers in cancer therapeutics. J. Med. P'ceutical Allied Sci., V 10 - I 2, , 2704-2711
26. Lachmann A, Torre D, Keenan A B, et al, 2018. Massive mining of publicly available RNA-seq data from human and mouse. Nature communications, 9, 1-10.
27. Lei X, Zhao J, Fujita H, et al, 2018. Predicting essential proteins based on RNA-Seq, subcellular localization and GO annotation datasets. Knowledge-Based Systems, 151, 136-148.
28. Szklarczyk D, Gable A L, Lyon D, et al, 2019. protein–protein

association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research, 47, D 607 - 613.
29. Hinterbrandner M, Rubino V, Stoll C, et al, 2021. Tnfrsf4-expressing regulatory T cells promote immune escape of chronic myeloid leukemia stem cells. JCI insight, 6.
30. Gu S, Zi J, Han Q, et al, 2020. Elevated TNFRSF4 gene expression is a predictor of poor prognosis in non-M3 acute myeloid leukemia. Cancer cell international, 20, 1-13.
31. Zheng Y H, Cai X B, Xia L Q, et al, 2021. Mutational screening of AGRN, SLC39A5, SCO2, P4HA2, BSG, ZNF644, and CPSF1 in a Chinese cohort of 103 patients with nonsyndromic high myopia. Molecular vision, 27, 706.
32. Previtali S C, Zhao E, Lazarevic D, et al, 2019. Expanding the spectrum of genes responsible for hereditary motor neuropathies. Journal of Neurology, Neurosurgery & Psychiatry, 90, 1171-1179.
33. Servais, L., Baudoin, H., Zehrouni, K., et al, 2013. Pregnancy in congenital myasthenic syndrome. Journal of neurology, 260, 815-819.
34. Mizutani K, Miyamoto S, Nagahata T, et al 2005. Upregulation and overexpression of DVL1, the human counterpart of the Drosophila dishevelled gene, in prostate cancer. Tumori Journal, 91, 546-551.
35. Sharma M, Castro-Piedras I, Simmons Jr G E, et al, 2018. Dishevelled: A masterful conductor of complex Wnt signals. Cellular signalling, 47, 52-64.
36. So T, Ishii N, 2019. The TNF–TNFR family of co-signal molecules. Co-signal Molecules in T Cell Activation, 53-84.

How to cite this article

Shalini kumari and Ruchi Yadav, 2022. RNA-Seq analysis to explore the variants in melanoma cells: molecular diagnosis and therapeutics. J. Med. P'ceutical Allied Sci. V 11 - I 3, Pages - 4869 - 4880 doi: 10.55522/jmpas.V11I3.2930.