WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.084
Volume 10, Issue 12, 1326-1339. Research Article ISSN 2277– 7105

RPLP1 GENE COULD BE A POTENTIAL BIOMARKER OF MELANOMA BRAIN METASTASES (MBM): INSIGHTS FROM SCRNASEQ DATA ANALYSIS

Sumedh Choukidar, Prasad Choukidar, Subodh Choukidar, *Dr. Archana Panche and Dr. Sanjay Harke

Dept. of Bioinformatics, MGM's Institute of Biosciences and Technology, Aurangabad.

Article Received on
28 July 2021,

Revised on 16 Aug. 2021,
Accepted on 04 Sept. 2021

DOI: 10.20959/wjpr202112-21677

***Corresponding Author**
Dr. Archana Panche
Dept. of Bioinformatics,
MGM's Institute of
Biosciences and Technology,
Aurangabad.

ABSTRACT

Melanoma brain metastases (MBM) is a secondary brain tumor that has metastasized from skin to the brain. Single cell RNA sequence analysis is a next generation sequencing method that is mainly used to analyze differences in genetic and protein information between cells and to better understand the specific roles of the gene. In this research our objective was to analyze the single cell RNA sequence data and identify the markers for the disease. We analyzed the samples of scRNAseq data of Melanoma brain metastases (MBM) using GALAXY server. After performing the several steps of the scRNAseq pipeline we were able to find the possible biomarker for the secondary brain tumor. 'RPLP1' gene is found to be promoting tumor metastasis and is highly expressed in all the six samples that are taken for the analysis. 'RPLP1' gene could be a potential biomarker of secondary brain tumor metastasis. This study will further lead practitioners and oncologists for the implementation of possible treatments against Melanoma brain metastases (MBM).

KEYWORDS: Scrnaseq, Melanoma brain metastases (MBM), Bioinformatics.

INTRODUCTION

1.1 Melanoma brain metastases (MBM)

Melanoma brain metastases (MBM) is a secondary brain tumor that has metastasized from skin to the brain.[1] Literature demonstrates that approximately 9%-17% of the cancer patients develop brain metastasis.[1] MBM directly causes death in approximately 60-70% of melanoma patients.[2] [Figure 01]. Several gene expression studies demonstrated that

PIK3CG, CCR4, FN14, and GRP94 genes plays role as biomarker for cancer and brain metastasis,[3,5] The need of single cell RNA sequence analysis is become a vital step towards analyzing the molecular targets for early disease diagnosis, disease progression tracking, medication resistance, mechanism identification, and personalized medicine,[6,7] In this study we analyzed the scRNAseq data of Melanoma brain metastases retrieved from NCBI's SRA database. We performed various steps on this data such as quality control, read alignment, quantification, differential expression and functional analysis to identify the possible biomarker for this specific disease. After performing the complete analysis, we report 'RPLP1' gene as possible biomarker for melanoma brain metastasis (MBM).

1.2 Single cell RNA sequence analysis

Single cell RNA sequence analysis is a next generation sequencing method that is mainly used to analyze differences in genetic and protein information between cells to obtain single cell level and to better understand the specific roles of the gene.[8] RNA-seq is usually performed in "bulk", and therefore the data represent a mean of organic phenomenon patterns across thousands to many cells; this might obscure biologically relevant differences between cells.[9] Single-cell RNA-seq (scRNA-seq) represents an approach to beat this problem.[9]

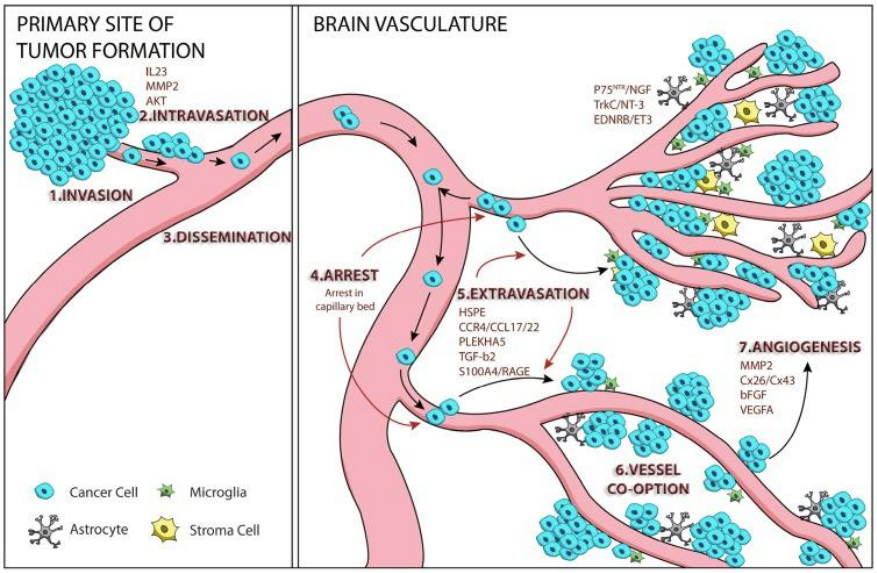


Figure 01: Showing mechanism of Melanoma brain metastases (MBM)

(Source:https://www.ncbi.nlm.nih.gov/core/lw/2.0/html/tileshop_pmc/tileshop_pmc_inline.html?title=Click%20on%20image%20to%20zoom&p=PMC3&id=5037746_ijms-17-01468-g001.jpg).

MATERIALS AND METHODS

2.1 Data Retrieval

Single cell data for Melanoma brain metastasis was retrieved from NCBI's SRA database (GEO ID: GSE174401).[10] We took one pair of samples from each location involved in the disease (Cerebrospinal fluid, Brain metastasis, Cutaneous) [Table no 01]. Each sample (SRR ID) represents a single cell.

Table no 01: Showing samples taken from SRA database with SRA accession and location.

Sr no	SRA accession	Sample	Location
1	SRR14523790	Sample 1	Cerebrospinal fluid
2	SRR14523791	Sample 2	Cerebrospinal fluid
3	SRR14523998	Sample 3	Brain metastasis
4	*SRR14523999*	Sample 4	Brain metastasis
5	*SRR14524074*	Sample 5	Cutaneous
6	*SRR14524075*	Sample 6	Cutaneous

2.2 Quality control and trimming

Quality control is the initial and important step of scRNAseq analysis. This method is useful for filtering out cell barcodes (excess NGS data which do not represent RNA).[11] Quality control is the most common parameters to examine the total number of reads sequenced, GC content and overall base quality score. After quality control we removed redundant/gratuitous sequences from raw data in order to achieve pre-processing of raw data. For this step we used FastQC (Quality control) and Trimmomatic (Trimming unwanted reads). FastQC is a popular FASTQ QC tool because it summarises read quality by position, tells users about adapter content in sequences, reports on tetramer frequencies, and many other features that one would anticipate from raw sequence data.[12]

Trimmomatic detects technical sequences in reads using two different techniques. The first, simple mode, works by trying to match the read to a user-supplied technical sequence.[13] The second mode, called "palindrome mode," is designed to detect the frequent "adapter read-through" scenario, in which a sequenced DNA fragment is shorter than the read length, resulting in adapter contamination at the end of the reads.[13]

2.3 Read alignment

Reads are typically mapped (aligned) to the reference transcriptome using un-gapped aligner. Percentage of mapped reads is an important parameter which indicates overall sequencing

accuracy. To enable fast and sensitive alignment of sequencing reads to a genome and a huge collection of minor variations, HISAT2 uses a novel graph-based data structure and an alignment method.[14] We used the HISAT2 tool for the fast and sensitive alignment.[14]

2.4 Quantification

Counting the number of reads that have aligned to each gene is called quantification. It can be achieved using programs such as String Tie.[15] Because it calculates abundance and exon-intron structure simultaneously, StringTie's network flow approach can re-construct transcripts more correctly than some earlier methods.[15] The method also has significantly better computational run-time efficiency than its predecessor, Cufflinks, allowing it to execute many times faster and consume much less memory.[15]

2.5 Annotation

Annotation is the process in which reads are typically mapped/ aligned to the reference genome to identify the read by the name of gene and chromosome locus from where it belongs to.[16] We performed an annotation step with the Annotate DESEQ2 tool.[17]

2.6 Differential expression analysis

Implementing statistical analysis on normalized gene count data to identify the quantitative changes between two or more samples is called differential expression analysis.[17,18] We performed differential expression analysis using Deseq2 tool, which combines methodological advancements with a number of innovative features to allow for a more quantitative examination of comparative RNA-seq data by employing shrinkage estimators for dispersion and fold change.[18]

2.7 Up-stream analysis

In the up-stream analysis we identify the highly expressed genes within the single cell data. We took Read counts as a parameter to identify the highly expressed genes.

2.8 Function analysis

After performing up-stream analysis of taken data we then analyzed the function of each highly up-regulated gene in the context of tumor metastases.

RESULTS AND DISCUSSION

3.1] Quality control

The analysis of quality control is being done for the testing of the samples and to remove the impurities from the samples such as barcodes, and other sequences that do not represent the transcriptome. Following figures show the quality check of the given samples before and after Trimmomatic [Figure 2-13].

Sample 1 (Cerebrospinal fluid)

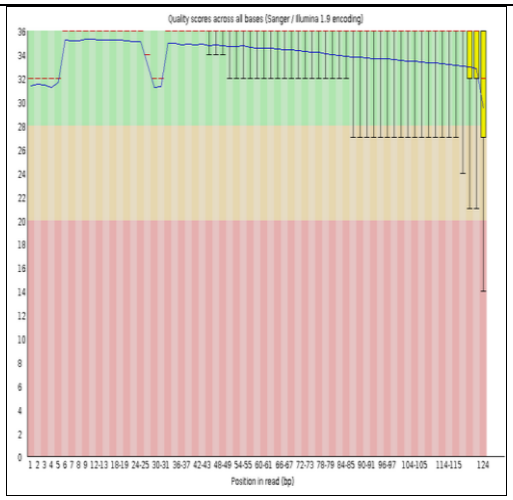


Figure 02: Showing per base quality content before Trimmomatic

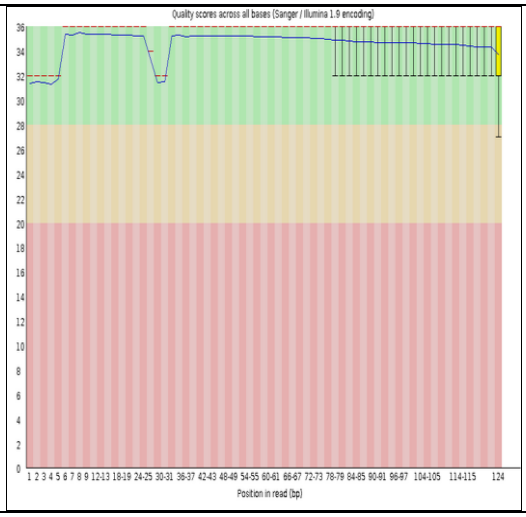


Figure 03: Showing per base quality content after Trimmomatic

Sample 2 (Cerebrospinal fluid)

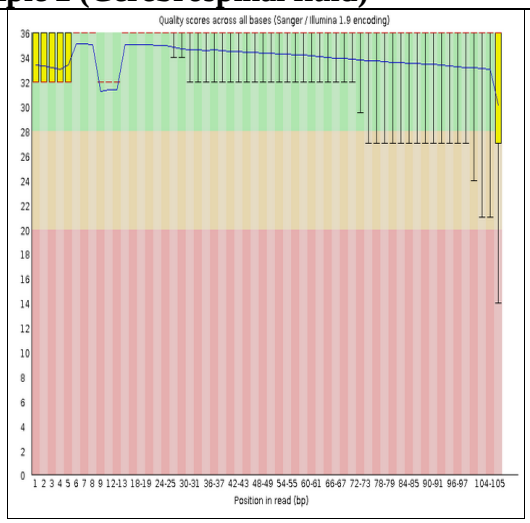


Figure 04: Showing per base quality content before Trimmomatic

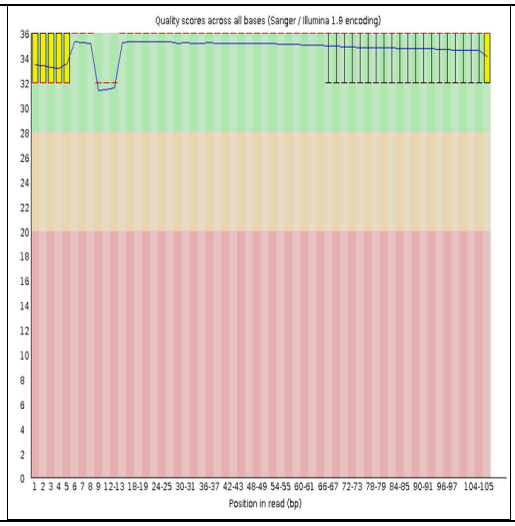


Figure 05: Showing per base quality content after Trimmomatic

Sample 03 (Brain metastasis)

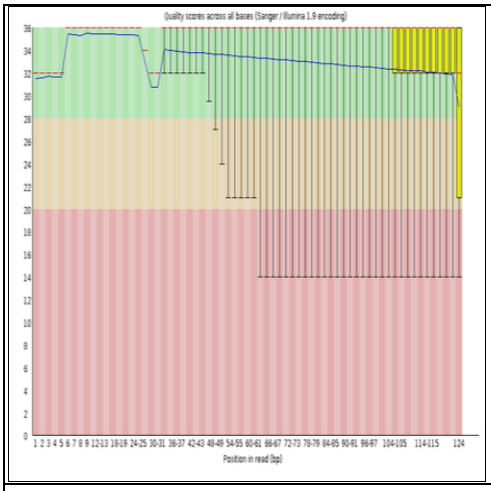


Figure 06: Showing per base quality content before Trimmomatic

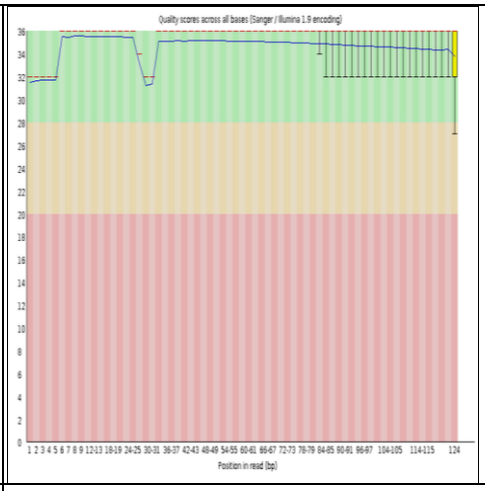


Figure 07: Showing per base quality content after Trimmomatic

Sample 4 (Brain metastasis)

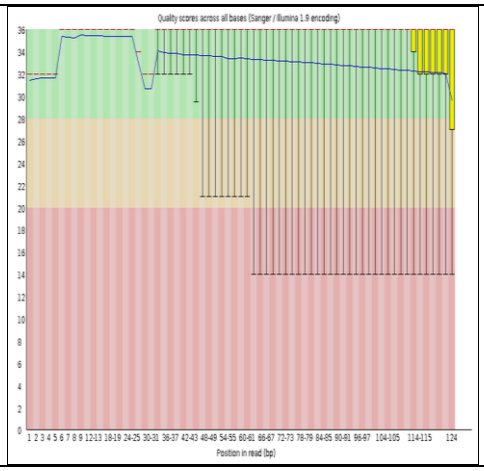


Figure 08: Showing per base quality content before Trimmomatic

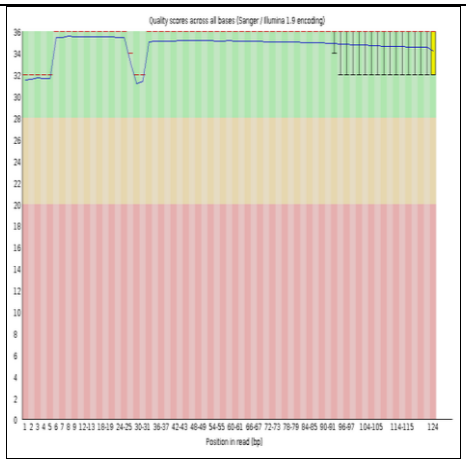


Figure 09: Showing per base quality content after Trimmomatic

Sample 05 (Cutaneous)

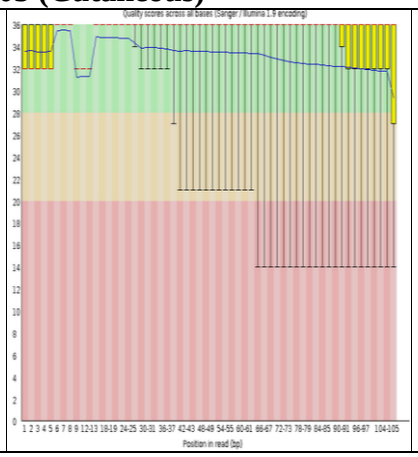


Figure 10: Showing per base quality content before Trimmomatic.

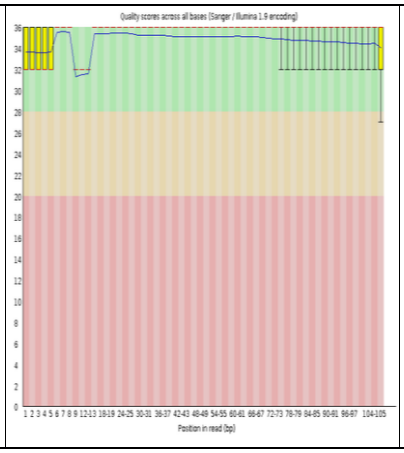


Figure 11: Showing per base quality content after Trimmomatic.

Sample 06 (Cutaneous)

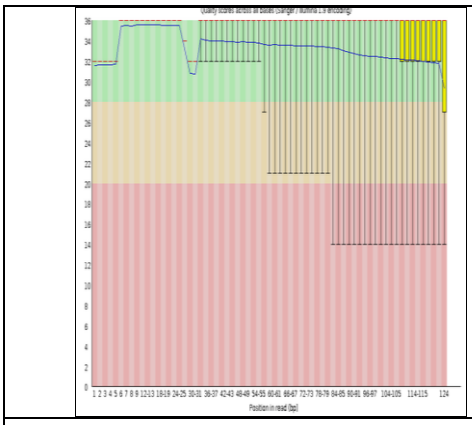


Figure 12: Showing per base quality content before Trimmomatic

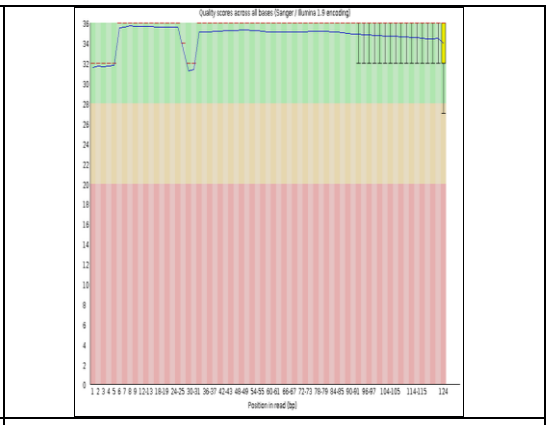


Figure 13: Showing per base quality content after Trimmomatic

3.2 Read alignment

In this step raw reads are aligned to the reference genome version (hg.38) to find homologous sequences for the raw reads. This step enables us to know the possible gene for a read with gene name and its Ensembl ID from which we could move forward to the further analysis of quantification.

3.3 Quantification

In this step we used the StringTie tool for the quantification of the reads. This tool enables us to estimate the actual read counts for the specific read, which is further used for the differential expression analysis and gene expression analysis.

3.4 Annotation

In this step output consists of annotated genes from samples with their Ensembl ID, chromosome locus, and gene name.

Results obtained after performing quantification and annotation are given in Figure 14

Gene ensembl id	Read counts	chr	Start	End	u/d	NA	Gene name
ENSG00000000003.15	24	chrX	100627107	100639991	-	NA	TSPAN6
ENSG00000000005.6	0	chrX	100584935	100599885	+	NA	TNMD
ENSG00000000419.12	263	chr20	50934866	50958555	-	NA	DPM1
ENSG00000000457.14	12	chr1	169849630	169894267	-	NA	SCYL3
ENSG00000000460.17	16	chr1	169662006	169854080	+	NA	C1orf112
ENSG00000000938.13	27	chr1	27612063	27635185	-	NA	FGR
ENSG00000000971.15	23	chr1	196651877	196747504	+	NA	CFH
ENSG00000001036.14	79	chr6	143494811	143511720	-	NA	FUCA2
ENSG00000001084.13	10	chr6	53497340	53616970	-	NA	GCLC
ENSG00000001167.14	13	chr6	41072944	41099976	+	NA	NFYA
ENSG00000001460.18	8	chr1	24356998	24416934	-	NA	STPG1
ENSG00000001461.17	107	chr1	24415801	24472976	+	NA	NIPAL3
ENSG00000001497.16	73	chrX	65512581	65534775	-	NA	LAS1L
ENSG00000001561.7	6	chr6	46129988	46146688	+	NA	ENPP4

Figure 14: Showing results obtained from quantification and annotation with Ensembl ID, Read counts, chromosome number, and Gene name.

3.5 Differential expression analysis

For the differential expression analysis, we used DEseq2 tool which gives quantitative changes in expression levels between experimental groups. This tool gives various plots such as PCA (Principal Component Analysis), Heat-map, Dispersion estimates, mean average, and P value [Figure 15].

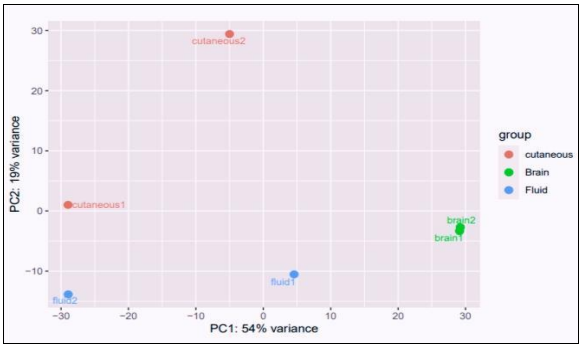
PCA: This figure 15(A) shows association between six samples of three groups. Samples of Brain 1 and Brain 2 are shown more similar to each other suggesting the same set of genes are expressed in both samples and less variance between the data. Samples of fluid and cutaneous are seen to be less similar to each other suggesting the genes expressed in the both samples are different. This phenomenon could happen due to subtypes of the same type of cells.

Heat-map: Heat-map shows the expression of genes between the samples. In figure 15(B) dark blue colour indicates the identical set of genes expressed in the sample which also validates the data as it shows the descending diagonal. The light blue colour indicates less identical genes are expressed in the samples. The white colour shows the dis-similar genes are expressed in the samples.

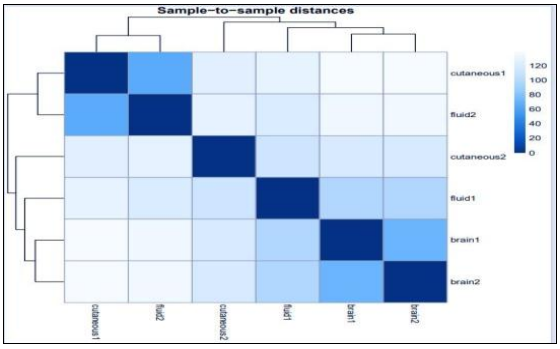
Dispersion estimates: The dispersion estimates for genes with the same mean will differ only based on their variance [Figure 15(C)]. Therefore, the dispersion estimates reflect the variance in gene expression for a given mean value.

MA plot: MA plots are commonly used to represent log fold-change versus mean expression between two treatments. This is visually displayed as a scatter plot with base-2 log fold-change along the y-axis and normalized mean expression along the x-axis [Figure 15(D)].

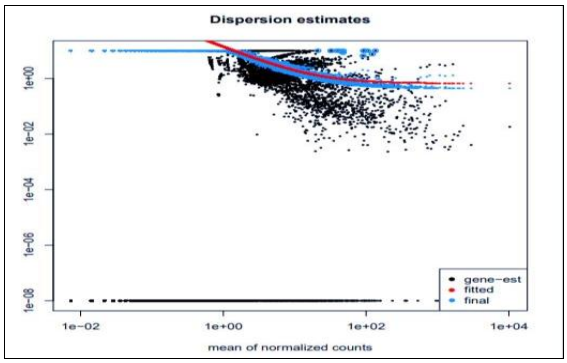
P value: P-values are specifically designed so that they are uniform under the null hypothesis. This graph is conservative type of graph which indicates the data is in discrete form [Figure 15(E)].



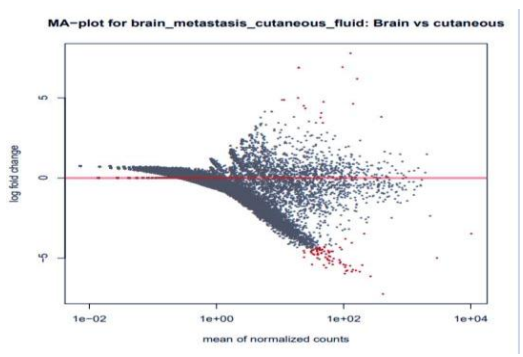
A) Showing PCA (Principal Component Analysis) plot



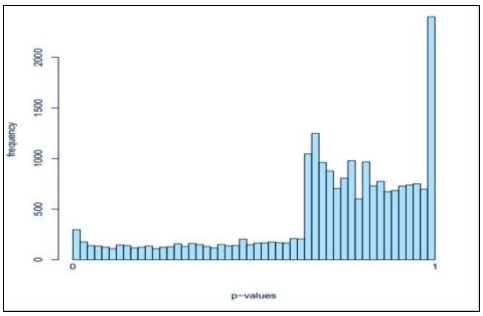
B) Showing Heat-map plot



C) Showing Dispersion estimates plot



D) Showing MA plot (mean average) plot



E) Showing P values plot

Figure 15: Showing various plots generated by Deseq2 tool A) PCA (Principal Component Analysis), B) Heat-map, C) Dispersion estimates, D) Mean average, and E) P value.

3.6 Up-stream analysis

The following bar diagram shows the highly up-regulated genes from all the six samples taken for the study [Figure 16].

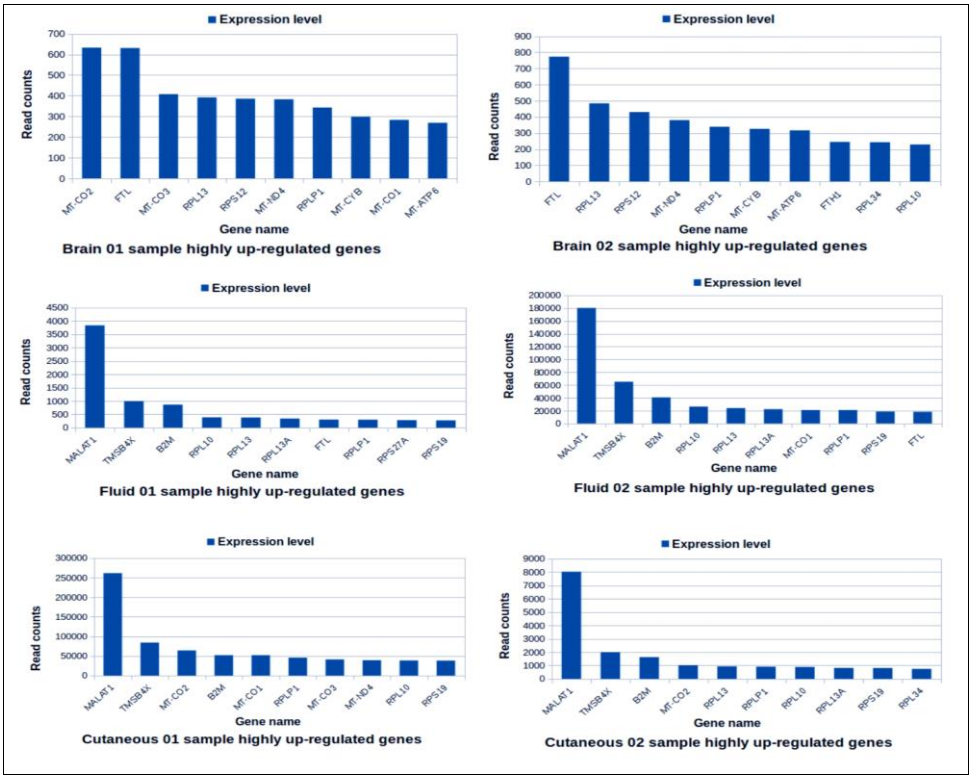


Figure 16: Showing highly up-regulated genes from all the six samples.

Above histograms (Figure 16) shows the 10 highly expressed (up-regulated) genes of all the six samples consisting of 6000 genes/sample taken for the study. Brain 01 sample shows high expression of MT-CO2 and FTL gene with read counts up to 650 where, as Brain 02 sample shows high expression of FTL gene with read counts up to 780. Fluid 01, Fluid 02, Cutaneous 01, and Cutaneous 02 samples, shows high expression of MALAT1 gene with read counts up to 3800, 18000, 26000, and 8000 respectively.

3.7 Function analysis

In this step we analyzed the functions of the genes which we found to be highly expressed in all the six samples after performing the several steps [Table 2-4]. RPLP1 gene is found to be expressed in all samples taken for the study [Table 5].

Table 02: Showing common up-regulated genes in both brain samples with function analysis.

Sr no	Gene name	Function	Sample
1	FTL	FTL was found to be overexpressed in various malignant tumors, and played a crucial role in regulating malignancy progress of cancers.	Brain 01 and Brain 02
2	RPL13	Role in tumor development is less clear.	Brain 01 and Brain 02
3	RPS12	RPS12 plays a role in cancer metastasis.	Brain 01 and Brain 02

4	MT-ND4	Mitochondrial gene	Brain 01 and Brain 02
5	RPLP1	Promotes tumor metastasis.	Brain 01 and Brain 02
6	MT-CYB	It plays role in melanoma	Brain 01 and Brain 02
7	MT-ATP6	Mitochondrial gene	Brain 01 and Brain 02

Table 03: Showing common up-regulated genes in both fluid samples with function analysis.

Sr no	Gene name	Function	Sample
1	MALAT1	MALAT1 drives tumorigenesis by promoting tumor cell proliferation.	Fluid 01 and Fluid 02
2	TMSB4X	Biomarker in all cancer types.	Fluid 01 and Fluid 02
3	B2M	B2M Is a regulator of cancer cell survival and metastasis.	Fluid 01 and Fluid 02
4	RPL14	Play a role in all types of cancer.	Fluid 01 and Fluid 02
5	RPL13	Role in tumor development is less clear.	Fluid 01 and Fluid 02
6	RPL13A	Role in tumor development is less clear.	Fluid 01 and Fluid 02
7	FTL	FTL was found to be overexpressed in various malignant tumors, and played a crucial role in regulating malignancy progress of cancers.	Fluid 01 and Fluid 02
8	RPLP1	Promotes tumor metastasis.	Fluid 01 and Fluid 02
9	RPS19	Role in tumor development is less clear.	Fluid 01 and Fluid 02

Table 04: Showing common up-regulated genes in both cutaneous samples with function analysis.

Sr no	Gene name	Function	Sample
1	MALAT1	MALAT1 drives tumorigenesis by promoting tumor cell proliferation.	Cutaneous 01 and Cutaneous 02
2	TMSB4X	Biomarker in all cancer types.	Cutaneous 01 and Cutaneous 02
3	MT-CO2	Mitochondrial gene	Cutaneous 01 and Cutaneous 02
4	B2M	B2M Is a regulator of cancer cell survival and metastasis.	Cutaneous 01 and Cutaneous 02
5	RPLP1	Promotes tumor metastasis.	Cutaneous 01 and Cutaneous 02
6	RPL10	Could be an important player in cancer development	Cutaneous 01 and Cutaneous 02
7	RPS19	Role in tumor development is less clear.	Cutaneous 01 and Cutaneous 02

Table 05: Showing common gene in the all samples.

Sr no	Gene name	function	Sample
1	RPLP1	Promotes tumor metastasis.	All samples

CONCLUSION

In this study we did scRNAseq analysis of Melanoma brain metastases (GEO ID: GSE174401) by performing several steps such as quality control, read alignment, quantification, gene expression, differential expression, and functional analysis to identify the highly expressed genes from the samples. After successfully performing the mentioned steps, we identified the common highly expressed gene in all samples to identify the potential biomarkers for improved prognosis and treatment of Melanoma brain metastases. Several other genes are also identified with the high expression level. Up-stream analysis gives out information about the highly expressed genes in which brain samples are having high expression of FTL genes. On the other hand, cutaneous and fluid samples are seen to be highly expressed with MALAT1 gene. Functional analysis reveals their role in the brain tumor metastasis. 'RPLP1' gene is highly expressed in all the six samples that are taken for the analysis and is found to be promoting tumor metastasis. 'RPLP1' gene could be a potential biomarker of Melanoma brain metastases. This analysis could be helpful in preventing tumor metastasis in the primary stage by inhibiting the activity of RPLP1 gene. This study will further lead practitioners and oncologists for the implementation of possible treatments against Melanoma brain metastases.

REFERENCES

1. Wu K, Sharma S, Venkat S, Liu K, Zhou X, Watabe K. Non-coding RNAs in cancer brain metastasis. Front Biosci (Schol Ed), 2016 Jan 1; 8: 187-202. doi: 10.2741/s457. PMID: 26709907; PMCID: PMC5652305.
2. Gutzmer R, Vordermark D, Hassel JC, Krex D, Wendl C, Schadendorf D, Sickmann T, Rieken S, Pukrop T, Höller C, Eigentler TK, Meier F. Melanoma brain metastases - Interdisciplinary management recommendations 2020. Cancer Treat Rev., 2020 Sep; 89: 102083. doi: 10.1016/j.ctrv.2020.102083. Epub 2020 Jul 18. PMID: 32736188.
3. Taylor, E.M., Byrum, S.D., Edmondson, J.L. *et al.* Proteogenomic analysis of melanoma brain metastases from distinct anatomical sites identifies pathways of metastatic progression. *acta neuropathol commun*, 2020; 8**:** 157. https://doi.org/10.1186/s40478-020-01029-x.
4. Klein A, Sagi-Assif O, Meshel T, Telerman A, Izraely S, Ben-Menachem S, Bayry J,

Marzese DM, Ohe S, Hoon DSB, Erez N, Witz IP. CCR4 is a determinant of melanoma brain metastasis. Oncotarget., 2017 May 9; 8(19): 31079-31091. doi: 10.18632/oncotarget.16076. PMID: 28415693; PMCID: PMC5458190.

5. Martínez-Aranda A, Hernández V, Guney E, Muixí L, Foj R, Baixeras N, Cuadras D, Moreno V, Urruticoechea A, Gil M, Oliva B, Moreno F, González-Suarez E, Vidal N, Andreu X, Seguí MA, Ballester R, Castella E, Sierra A. FN14 and GRP94 expression are prognostic/predictive biomarkers of brain metastasis outcome that open up new therapeutic strategies. Oncotarget., 2015 Dec 29; 6(42): 44254-73. doi: 10.18632/oncotarget.5471. PMID: 26497551; PMCID: PMC4792555.
6. Zhang X, Liang Z, Wang S, Lu S, Song Y, Cheng Y, Ying J, Liu W, Hou Y, Li Y, Liu Y, Hou J, Liu X, Shao J, Tai Y, Wang Z, Fu L, Li H, Zhou X, Bai H, Wang M, Lu Y, Yang J, Zhong W, Zhou Q, Yang X, Wang J, Huang C, Liu X, Zhou X, Zhang S, Tian H, Chen Y, Ren R, Liao N, Wu C, Zhu Z, Pan H, Gu Y, Wang L, Liu Y, Zhang S, Liu T, Chen G, Shao Z, Xu B, Zhang Q, Xu R, Shen L, Wu Y, Tumor Biomarker Committee OBOCSOCOC. Application of next-generation sequencing technology to precision medicine in cancer: joint consensus of the Tumor Biomarker Committee of the Chinese Society of Clinical Oncology. Cancer Biol Med., 2019 Feb; 16(1): 189-204. doi: 10.20892/j.issn.2095-3941.2018.0142. PMID: 31119060; PMCID: PMC6528448.
7. de Unamuno Bustos B, Murria Estal R, Pérez Simó G, de Juan Jimenez I, Escutia Muñoz B, Rodríguez Serna M, Alegre de Miquel V, Llavador Ros M, Ballester Sánchez R, Nagore Enguídanos E, Palanca Suela S, Botella Estrada R. Towards personalized medicine in melanoma: implementation of a clinical next-generation sequencing panel. Sci Rep., 2017; 7(1): 495. doi: 10.1038/s41598-017-00606-w, PMID 28356599.
8. Hwang, B., Lee, J.H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med.*, 2018; 50: 1–14. https://doi.org/10.1038/s12276-018-0071-8
9. Olsen TK, Baryawno N. Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol., 2018 Apr; 122(1): e57. doi: 10.1002/cpmb.57. PMID: 29851283.
10. Smalley I, Chen Z, Phadke M, Li J, Yu X, Wyatt C, Evernden B, Messina JL, Sarnaik A, Sondak VK, Zhang C, Law V, Tran N, Etame A, Macaulay RJB, Eroglu Z, Forsyth PA, Rodriguez PC, Chen YA, Smalley KSM. Single-Cell Characterization of the Immune Microenvironment of Melanoma Brain and Leptomeningeal Metastases. Clin Cancer Res., 2021 May 25. doi: 10.1158/1078-0432.CCR-21-1694. Epub ahead of print. PMID: 34035069.

11. Jiang P. Quality Control of Single-Cell RNA-seq. Methods Mol Biol., 2019; 1935: 1-9. doi: 10.1007/978-1-4939-9057-3_1. PMID: 30758816.
12. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics, 2017 Oct 1; 33(19): 3137-3139. doi: 10.1093/bioinformatics/btx373. Epub 2017 Jun 9. PMID: 28605449; PMCID: PMC5870778.
13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014 Aug 1; 30(15): 2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
14. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol., 2019 Aug; 37(8): 907-915. doi: 10.1038/s41587-019-0201-4. Epub 2019 Aug 2. PMID: 31375807; PMCID: PMC7605509.
15. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc., 2016 Sep; 11(9): 1650-67. doi: 10.1038/nprot.2016.095. Epub 2016 Aug 11. PMID: 27560171; PMCID: PMC5032908.
16. Burgoon LD, Zacharewski TR. Bioinformatics: databasing and gene annotation. Methods Mol Biol., 2008; 460: 145-57. doi: 10.1007/978-1-60327-048-9_7. PMID: 18449486.
17. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol., 2010; 11(10): R106. doi: 10.1186/gb-2010-11-10-r106. Epub 2010 Oct 27. PMID: 20979621; PMCID: PMC3218662.
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol., 2014; 15(12): 550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.