Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI: 10.26479/2018.0405.43**

META-ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES IN GLIOBLASTOMA STEM CELLS USING MICROARRAY DATASETS

Sweta P Tripathi[1], Vinal Upadhyay[2,3], Himanshu A. Pandya[1], Rakesh M Rawal[2*]

1. Department of Botany, Bioinformatics and Climate Change Impacts Management, University School of Sciences, Gujarat University, Ahmedabad, India.

2. Department of Life Sciences, University School of Sciences, Gujarat University, Ahmedabad, India.

3. Clinical Carcinogenesis Laboratory, Department of Cancer Biology, The Gujarat Cancer Research Institute, Ahmedabad, India.

ABSTRACT: Glioblastoma (GB) type IV, the most vulnerable brain tumor is worsened by the role of glioblastoma stem cells leads to multiple targets in its disease physiology. The objective of the present study is to identify potential key genes, miRNAs and pathways in glioblastoma stem cells (GBSCs) using meta-analysis approach. Using the seven publicly available GEO datasets, 368 differentially expressed genes (DEGs) are actively involved in the GBSCs population in which 172 and 196 DEGs were down-and up-regulated in the GBSCs samples, respectively. Grouping of DEGs revealed the primary involvement of blood coagulation process (P = 9.16E-06) whereas its associated immunological processes including complement and coagulation cascades (P = 6.56E-06) were noted in the KEGG pathway enrichment analysis. Furthermore, the protein-protein interaction networks showed the prominent hub of proteins: Topoisomerase (DNA) IIα (TOP2; degree= 24), Von Willebrand factor (VWF; degree= 23), Fibronectin 1 (FN1; degree= 21) and Secreted phosphoprotein 1 (SPP1; degree= 21). Overall, results suggested the active role of several genes associated with the blood coagulation and complement coagulation cascades in GBSCs and could be regarded as potential biomarkers.

KEYWORDS: Glioblastoma stem cells (GBSCs), differentially-expressed genes, protein-protein interaction network, miRNA-target network.

Corresponding Author: Dr. Rakesh M Rawal* Ph.D.

Department of Life Sciences, University School of Sciences, Gujarat University, Ahmedabad, India. Email Address: rakeshmrawal@gujaratuniversity.ac.in

1. INTRODUCTION

Glioma, especially high-grade tumor Glioblastoma (GB) type IV, is the most common and aggressive type of brain tumor causes death globally [1]. Among human cancers, the malignant primary brain tumor of GB show five years worst survival rate [2]. Several advanced treatments are currently available to treat GB including surgical resection, radiotherapy, and systematic chemotherapy. Standard maximal safe surgical resection followed by temozolomide chemotherapy and radiation forms the safest option of treating GB cases and increasing the survival time to 14.6 months [3, 4]. However, the high inter and intra-tumor heterogeneity of GB brain tissues [5] makes it difficult to distinguish tumor tissues from normal ones and diagnose in the early stage of GB disease progression. To better develop novel therapeutics, researchers around the globe have revived the current understandings of cellular and molecular mechanisms that lead to GB tumor initiation and progression and found glioblastoma stem cells (GBSCs) as the new direction towards therapeutic and development purpose of a biomarker for GB cases [6]. Functionally, they target new groups of cells in the brain and easy to differentiate normal and tumor cells. On the basis of these tumor cells differentiations, several astrocytomas are classified in mitosis, vascular endothelial proliferation or necrosis in which two or three characteristics are used to diagnose GBSCs [7]. Reported glioblastoma stem cells (GBSCs) biomarkers include CD133, nestin, NANOG, SALL4, STAT3, SOX2, c-Myc, Olig2, Bmi1, CD44, L1CAM, and KLF4 [8]. MiR138 is useful as a prognostic biomarker target for MXD1 gene in glioblastoma stem cells [9, 10]. With adcent of new technologies, thousands of genes expression profiles with its experiments are available across various commercially and publically available databases. These data can be further used to discover common or new therapeutic approaches in the area of cancer heterogeneity, drug discovery, biomarker development, key genes identification to understand cancer progression [11]. Especially, microarray technology helps to unveil thousands of genes expressed at a time which can be associated with meta-analysis approach in order to find the best targets [12, 13]. In this study, we aim to analyze the gene expression data and differentially expressed genes (DEGs) in GBSCs using meta-analysis approach. Computational analysis was performed using the microarray datasets on tissues samples of GBSCs: GSE46531, GSE45899, GSE20736, GSE18150, GSE4536, GSE7181 and GSE57978. In addition, functional and pathway enrichment analyses were performed. Moreover, protein-protein interaction (PPI) and miRNA-target regulatory networks were also constructed.

2. MATERIALS AND METHODS

Microarray data retrieval from ArrayExpress database

Seven public microarray datasets related to glioblastoma stem cells (GBSCs) were taken from ArrayExpress database. The selected datasets were based on two platforms: [Affymetrix Human Genome U133 Plus 2.0 HG-U133_Plus_2] and Affymetrix GeneChip Human Gene 1.0 ST Array [HuGene-1_0-st-v1]. The datasets were chosen on basis of the following criteria: every single data

set must contain at least one stage with ≥1 <what is the unit here> of tumor samples and the total number of samples must be more than 10 in any stage. The selected datasets encompassed 149 patient samples which were subjected to meta-analysis. These datasets were downloaded from ArrayExpress ((https://www.ebi.ac.uk/arrayexpress/) [14] as the raw data (.CEL files) (Table 1).

Table 1: Summary of Glioblastoma stem cells (GBSCs) microarray datasets

Datasets	Array type	Total
GSE46531 [15]	HG-U133_Plus_2	12
GSE45899 [16]	HG-U133_Plus_2	12
GSE20736 [17]	HG-U133_Plus_2	6
GSE18150 [18]	HG-U133_Plus_2	6
GSE4536 [19]	HG-U133_Plus_2	101
GSE7181 [20]	HG-U133_Plus_2	6
GSE57978 [21]	HuGene-1_0-st	6

Identification of differentially expressed genes (DEGs) from multiple microarray datasets

R Package version 3.3.1 **(**www.bioconductor.org/) with simpleaffy [22] and limma packages [23] were used for differential gene expression analysis. Retrieved microarray datasets were converted from probe-level into expression values and rectified background noise by quantile normalization and probe summarization. The t-test **[**24**]** was performed to identify significantly expressed DEGs in GBSCs patient samples by adopting the thresholds: $P<0.05$ and $|\log_2 FC \text{ (fold change)}| >1$. A heat map was also generated using Z-score normalization of $\log_2$ expression values to illustrate the relative expression levels of DEGs in GBSCs samples.

Gene Ontology (GO) and pathway enrichment analysis of DEGs

Gene ontology (GO) enrichment analysis was performed as a part of functional annotation study to understand the biological significance of the DEGs through different biological processes, and their functional roles in the GBSCs pathways. Online available software GeneCodis [25] (http://genecodis.cnb.csic.es) was used to perform this analysis. In addition, the pathway enrichment analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database **[**26**]** was conducted.

Protein-Protein (PPI) and Protein-miRNA network construction

StringDB (https://string-db.org/) database was used for the retrieval of known DEGs protein-protein and protein-predicted protein regulatory network. The database contains pre-computed storage of experimental, investigational and analysis based results [27]. The StringDB was used to analyze the interactions between identified DEGs using confidence (combined score: 0.9) and a protein-protein interaction network was constructed with its predicted protein. The cut-off degree was derived from the nodes calculation in the PPI network degree and the nodes with high degree than cut-off were counted to be hub proteins in the PPI network. Based on the constructed PPI, GeneCodis online

software miRNA prediction was utilized to generate PPI and miRNA-target gene results and constructed the regulatory networks of protein-protein, protein-predicted protein and miRNA- target gene in GBSCs, visualized using Cytoscape (v3.4.0) software [28].

3. RESULTS AND DISCUSSION

Glioblastoma is a malignant brain tumor initiated by glioma-initiating cells (GICs; also called glioma cancer stem cells) is one among the causes of high death rate in the world [29]. It has the high degree of tumor heterogenicity, as a result several oncogenic targets playing an active role in disease progression and the available therapies are effective only to some extent. In this scenario, therapeutic targets sorted based on differential gene expression will be the most effective approach to treat GBSCs.

Identification of differentially expressed genes (DEGs) related to GBSCs

Significantly expressed genes obtained from the above step were classified on the basis of cut-off criteria including log P <0.05 and |log FC| >1.0 for up-regulated genes and |log FC| >-1.0 for down-regulated genes and finally obtained 368 DEGs which comprised 196 up-regulated and 172 down-regulated genes.

Pathway enrichment analysis

The KEGG pathway enrichment analysis recognized significant DEGs in 10 pathways (Table 2). The most significant and highly enriched pathways were complement and coagulation cascades (P=6.56E-08), TGF-beta signaling pathway (P = 5.17E-06) and systematic lupus erythematosus (P = 1.35E-07) (Table 2).

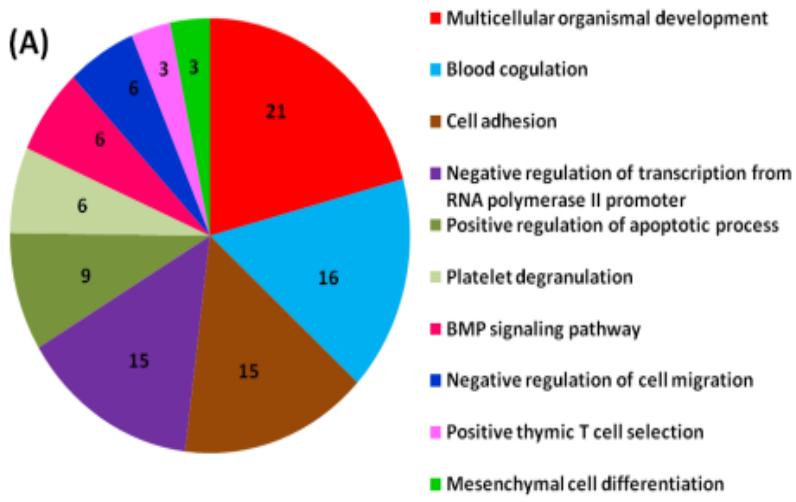
Table 2: The top 10 enriched KEGG pathway of DEGs

KEGG_ID	Pathway	Gene_count	P-value
hsa04610	Complement and coagulation cascades	10	6.56E-08
hsa04350	TGF-beta signaling pathway	9	5.17E-06
hsa05322	Systematic lupus erythematosus	9	1.35E-07
hsa05150	Staphylococcus aureus infection	7	2.15E-07
hsa04512	ECM-receptor interaction	8	1.17E-05
hsa05140	Leishmaniasis	7	2.67E-06
hsa04510	Focal adhesion	11	2.58E-06
hsa05020	Prion disease	5	1.47E-05
hsa05142	Chagas disease (American trypanosomiasis)	7	4.77E-05
hsa05323	Rheumatoid arthritis	6	0.0001

KEGG pathways enrichment analysis resulted in highly enriched pathways such as complement and coagulation cascades, TGF-beta signaling pathway, and systematic lupus erythematosus. Complement and coagulation cascade pathways (genes=10, P=6.56E-08) were strongly activated in GBSCs, PLAUR, C1QA, C3, A2M, SERPINA1, SERPINE1, VWF, F5, C1QC and C1QB. From these genes, VWF and SERPINE1 genes belong to most of the variants. VWF gene is present in the endothelial cells of neurons and it promotes bridge formation of platelets adhesion between sub-endothelial collagen matrix and platelet-surface receptor complex GPIb-IX-V. It also acts as the chaperone for coagulation factor VIII, during injury site. Bibliographic literature listed that its expression in the GBSCs is very heterogenous causing mutational changes and as a result, cytokine is dysregulated which in turn hampers the regulation of apoptosis, cell proliferation, differentiation and tissue homeostasis [30, 31]. Another highly expressed gene is SERPINE1 which is a part of blood-brain barrier, also known as plasminogen activator inhibitor (PA-1). It is also a primary regulator in the plasminogen-plasmin system which inhibits plasminogen activator. We note in the present study that high expressions of this gene may lead to lysosomal degradation of plasminogen activator complexes and its reappearance on the cell surface will disrupt the locomotion and direction of migration of cells and as a result, GBSCs patients survival rate decreases substantially [32].

Gene ontology (GO) enrichment analysis

To understand up and down-regulated differential gene expressions in the selected GBSCs samples, we performed GO enrichment analysis. The GO term for DEGs were significantly enriched in biological process (BP) that included the blood coagulation (GO: 0007596, P = 9.12E-06), negative regulation of transcription from RNA polymerase II promoter (GO: 0000122, 3.88E-05) and BMP signaling pathway (GO: 0030509, P = 1.8E-05), while for molecular functions (MF) the enriched terms was protein binding (GO: 0005515, P = 8.83E-17), and for cellular compartments (CC), the GO enriched terms were cytosol (GO: 0005737, P = 9.67E-07), extracellular region (GO : 0005576, P = 7.79E-07) and extracellular space (GO : 0005576, P = 7.79E-07) (Figure 1).



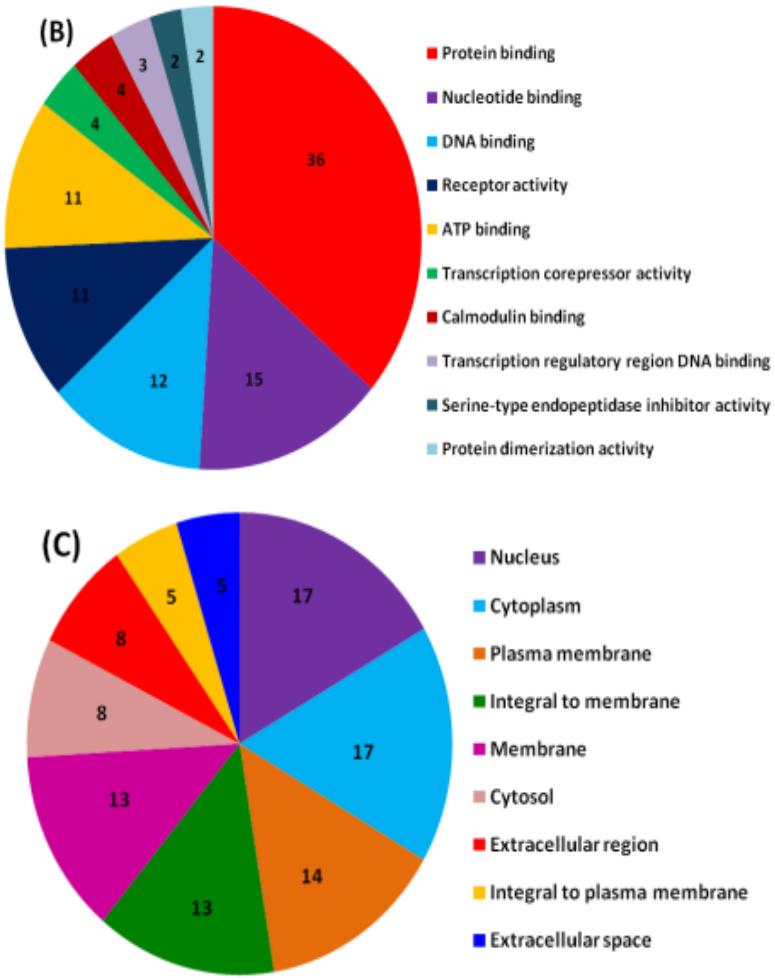


Figure 1: The top list of enriched GO terms for DEGs (P value≤ 1.06E-07, logFc≤1). A. biological process for DEGs; B. molecular functions for DEGs; C. cellular component for DEGs.

Protein-Protein interaction (PPI) network

The PPI network is represented by two components: nodes denote proteins while edges indicate interactions between two proteins. The efficiency of a network could be traced by the node shape and the greater the degree of connection the number of proteins they interacted. The PPI network was constructed using 172 down-regulated and 196 up-regulated DEGs and visualized using Cytoscape software that showed 231 nodes and 562 edges. The significant hub proteins identified from this effort included TOP2A (Topoisomerase (DNA) II Alpha, Degree= 24), VWF (Von Willebrand Factor, Degree=23) whereas FN1 (Fibronectin 1), SPP1 (Secreted Phosphoprotein 1) had degree=21. Other hub proteins namely EGR1 (Early Growth Response 1), FOS (Fos Proto-Oncogene, AP-1 Transcription Factor Subunit), PTGS2 (Prostaglandin-Endoperoxide Synthase 2) and SERPINE1 (Serpin Family E Member 1) secured same Degree= 20. The predicted proteins in this network were IFT140 (Intraflagellar Transport 140), TMEM240 (Transmembrane Protein 204), ADC (Arginine Decarboxylase), NAA38 (N(Alpha)-Acetyltransferase 38, NatC Auxiliary Subunit) (Figure 2).

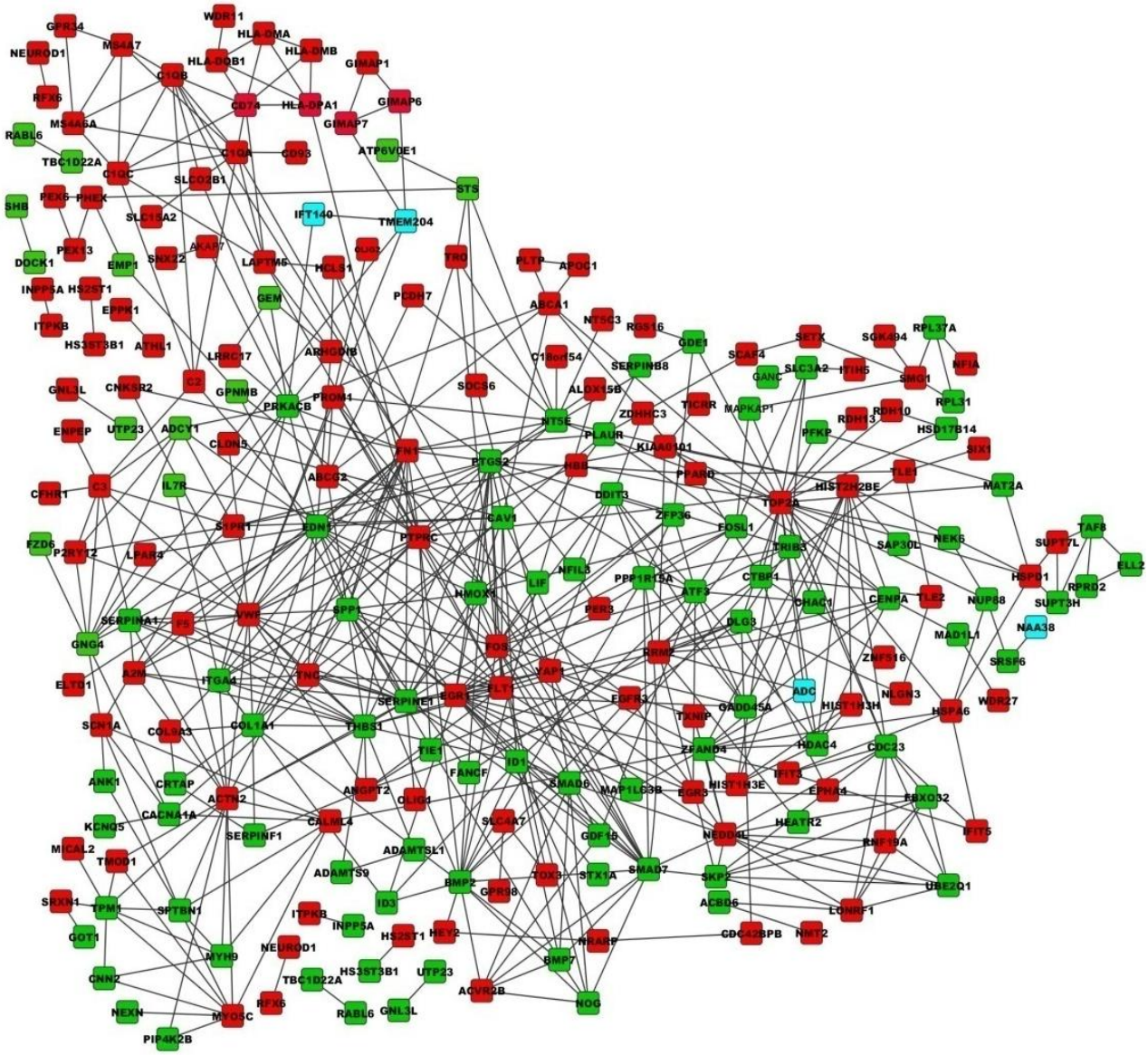


Figure 2: Protein-protein interaction network for DEGs. Red nodes represent protein products of up-regulated DEGs, green nodes represent protein products of down-regulated DEGs, blue nodes represent predicted protein product of up and down DEGs and the lines between two nodes denote the interactions between them.

From the PPI interaction network, we found that TOP2A and SPP1 are most prominent hub protein and involved in immune response and tissue remodeling leads to tumor progression by increasing migration, invasion of cancer stem cells in the tumor [33]. However further validation of DEGs required, and suggest that additional information could give new dimension to the identification of new targets for GBSCs and possibly the development of better cancer chemotherapeutic approaches.

The integrated regulatory network of miRNA-target gene

Using Genecodis online software with its MicroRNA annotation function, the regulatory miRNA-target gene network was built by connecting 144 nodes (134 DEGs and 10 miRNAs families) with 207 edges. In this network, notable genes were found to interact with multiple miRNAs such as

Leucine Rich Repeat Containing 17 (LRCC17; targeted by hsa-miR-556-3p, hsa-miR-568, hsa-miR-299-3p and hsa-miR-19b-2*) and Tropomyosin 1 Alpha (TPM1; targeted by hsa-miR-556-3p, hsa-miR-338-3p, hsa-miR-302c* and hsa-miR-542-3p) (Figure 3).

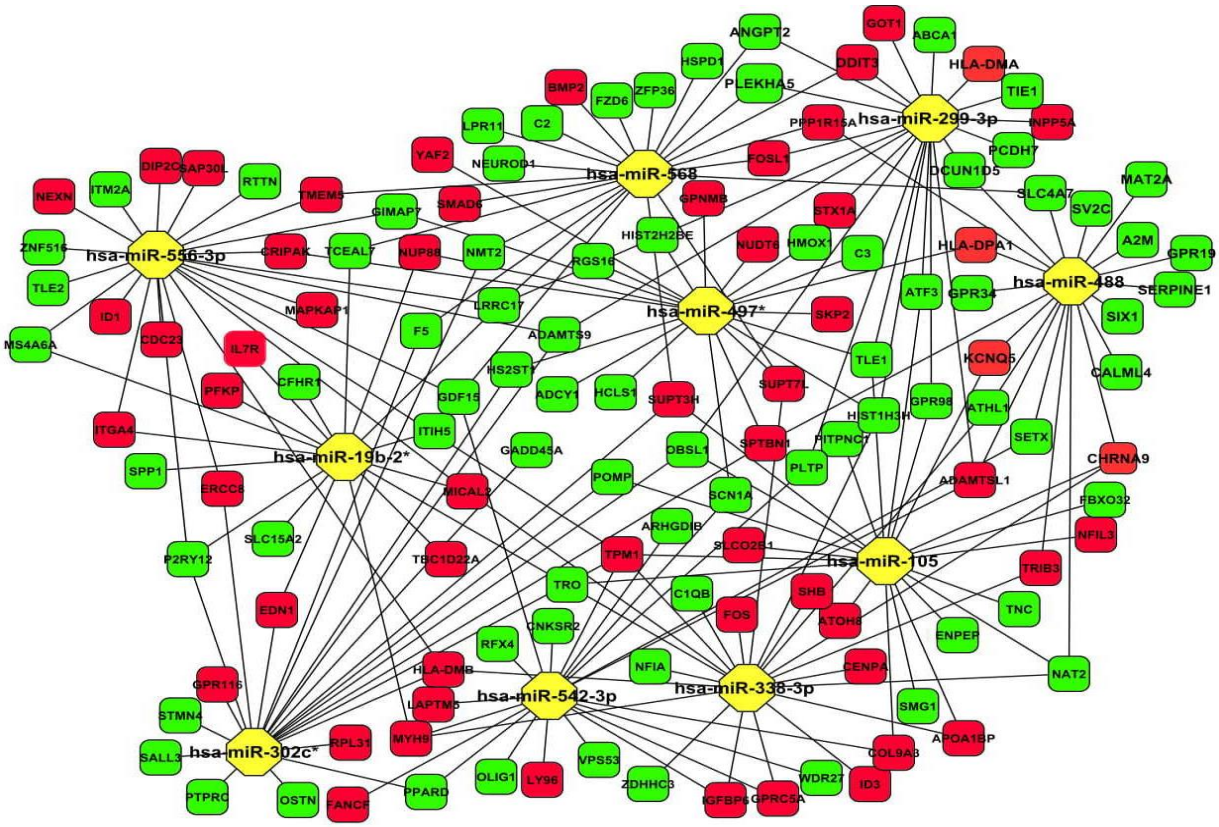


Figure 3: Integrated miRNA-target gene: Red nodes present up-regulated DEGs, and green nodes represent down-regulated DEGs protein products, Hexagonal nodes represents miRNAs and the lines between two nodes show interaction between them.

4. CONCLUSION

We performed a computational analysis on GBSCs and identified 196 up-regulated and 172 down-regulated DEGs in GBSCs samples compared with control samples. SERPINE1 and VWF genes were highly enriched in the blood coagulation biological process to disrupt complement and coagulation pathway which leads to the high risk of GBSCs. In the protein-protein interaction network TOP2A, VWF, FN1, SPP1, EGR1, FOS, PTGS2, and SERPINE1 genes encoded protein were identified which plays crucial roles in GBSCs. PPI network predicted proteins were IFT140, TMEM240, ADC and NAA38. Additionally, LRCC17 targeted by hsa-miR-556-3p, hsa-miR-568, hsa-miR-299-3p and hsa-miR-19b-2* and TPM1; targeted by hsa-miR-556-3p, hsa-miR-338-3p, hsa-miR-302c* and hsa-miR-542-3p. Hence, these identified miRNAs may contribute to the pathogenesis of GBSCs and may be potential therapeutic targets. However, these predictions require further experimental validation.

CONFLICT OF INTEREST

The authors declare that no competing financial interests exist.

REFERENCES

1. Davis M. Glioblastoma: Overview of Disease and Treatment. Clinical Journal of Oncology Nursing. 2016Jan;20(5).
2. Smoll NR, Schaller K, Gautschi OP. Long-term survival of patients with glioblastoma multiforme (GBM). Journal of Clinical Neuroscience. 2013;20(5):670–5.
3. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO Classification of Tumours of the Central Nervous System. Acta Neuropathologica. 2007Jun;114(2):97–109.
4. Fuller GN, Scheithauer BW. The 2007 Revised World Health Organization (WHO) Classification of Tumours of the Central Nervous System: Newly Codified Entities. Brain Pathology. 2007;17(3):304–7.
5. Denysenko T, Gennero L, Roos MA, Melcarne A, Juenemann C, Faccani G, et al. Glioblastoma cancer stem cells: heterogeneity, microenvironment and related therapeutic strategies. Cell Biochemistry and Function. 2010Oct;28(5):343–51.
6. Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proceedings of the National Academy of Sciences. 2013;110(10):4009–14.
7. Nduom EK-E, Hadjipanayis CG, Meir EGV. Glioblastoma Cancer Stem-Like Cells. The Cancer Journal. 2012;18(1):100–6.
8. Brown DV, Filiz G, Daniel PM, Hollande F, Dworkin S, Amiridis S, et al. Expression of CD133 and CD44 in glioblastoma stem cells correlates with cell proliferation, phenotype stability and intra-tumor heterogeneity. Plos One. 2017;12(2).
9. Pointer KB, Clark PA, Zorniak M, Alrfaei BM, Kuo JS. Glioblastoma cancer stem cells: Biomarker and therapeutic advances. Neurochemistry International. 2014;71:1–7.
10. Chan XHD, Nama S, Gopal F, Rizk P, Ramasamy S, Sundaram G, et al. Targeting Glioma Stem Cells by Functional Inhibition of a Prosurvival OncomiR-138 in Malignant Gliomas. Cell Reports. 2012;2(3):591–602.
11. Sweta P Tripathi, Vinal Upadhyay, Himanshu A. Pandya, Rakesh M Rawal. Metaanalysis of differentially expressed genes in glioblastoma based on microarray data, Int J of sci res. July (2017);6(7).
12. Liang M, Cowley AW, Greene AS. High throughput gene expression profiling: a molecular approach to integrative physiology. The Journal of Physiology. 2003Oct;554(1):22–30.
13. Liu B. High-throughput Gene Expression Profiling: Its Application in Studying Transfusion-Transmitted Infections. Journal of Bioanalysis & Biomedicine. 2013;05(01).

14. Trevino V, Falciani F, Barrera-Saldana HA. DNA Microarrays: a Powerful Genomic Tool for Biomedical and Clinical Research. Mol Med. 2007 Sep-Oct;13(9-10):527-41.
15. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, et al. ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Research. 2007Mar;35(Database).
16. Ye F, Zhang Y, Liu Y, Yamada K, Tso JL, Menjivar JC, et al. Protective Properties of Radio-Chemoresistant Glioblastoma Stem Cell Clones Are Associated with Metabolic Adaptation to Reduced Glucose Dependence. PLoS ONE. 2013;8(11).
17. Rheinbay E, Suvà ML, Gillespie SM, Wakimoto H, Patel AP, Shahid M, et al. An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. Cell Reports. 2013;3(5):1567–79.
18. Nogueira L, Ruiz-Ontañon P, Vazquez-Barquero A, Lafarga M, Berciano MT, Aldaz B, et al. Blockade of the NFκB pathway drives differentiating glioblastoma-initiating cells into senescence both in vitro and in vivo. Oncogene. 2011;30(32):3537–48.
19. Suva M-L, Riggi N, Janiszewska M, Radovanovic I, Provero P, Stehle J-C, et al. EZH2 Is Essential for Glioblastoma Cancer Stem Cell Maintenance. Cancer Research. 2009;69(24):9211–8.
20. Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell. 2006;9(5):391–403.
21. Beier D, Hau P, Proescholdt M, Lohmeier A, Wischhusen J, Oefner PJ, et al. CD133 and CD133-Glioblastoma-Derived Cancer Stem Cells Show Differential Growth Characteristics and Molecular Profiles. Cancer Research. 2007;67(9):4010–5.
22. Singer E, Judkins J, Salomonis N, Matlaf L, Soteropoulos P, Mcallister S, et al. Reactive oxygen species-mediated therapeutic response and resistance in glioblastoma. Cell Death & Disease. 2015;6(1).
23. Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics. 2005Feb;21(18):3683–5.
24. Smyth GK. limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor Statistics for Biology and Health. :397–420.
25. Jeanmougin M, Reynies AD, Marisa L, Paccard C, Nuel G, Guedj M. Should We Abandon the t-Test in the Analysis of Gene Expression Microarray Data: A Comparison of Variance Modeling Strategies. PLoS ONE. 2010Mar;5(9).
26. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Research. 2012Sep;40(W1).

27. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 1999Jan;27(1):29–34.
28. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Research. 2010Feb;39(Database).
29. Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003Jan;13(11):2498–504.
30. Varghese S, Whipple R, Martin SS, Alexander HR. Multipotent Cancer Stem Cells Derived from Human Malignant Peritoneal Mesothelioma Promote Tumorigenesis. PLoS ONE. 2012;7(12).
31. Charalambous C, Chen TC, Hofman FM. Characteristics of tumor-associated endothelial cells derived from glioblastoma multiforme. Neurosurgical Focus. 2006;20(4).
32. Roy A, Coum A, Marinescu VD, Põlajeva J, Smits A, Nelander S, et al. Glioma-derived plasminogen activator inhibitor-1 (PAI-1) regulates the recruitment of LRP1 positive mast cells. Oncotarget. 2015;6(27).
33. Kijewska M, Kocyk M, Kloss M, Stepniak K, Korwek Z, Polakowska R, et al. The embryonic type of SPP1 transcriptional regulation is re-activated in glioblastoma. Oncotarget. 2016;8(10).