ORIGINAL RESEARCH PAPER
VOLUME-6 | ISSUE-7 | JULY-2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 78.46
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

META-ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES IN GLIOBLASTOMA BASED ON MICROARRAY DATA

Biological Science

Sweta P Tripathi Department of Bioinformatics, Applied Botany Center (ABC),University School of Sciences, Gujarat University, Ahmedabad-380009, India.

Vinal Upadhyay Department of Life Sciences, University School of Sciences, Gujarat University, Ahmedabd-380016, India.

Himanshu A. Pandya Department of Bioinformatics, Applied Botany Center (ABC),University School of Sciences, Gujarat University, Ahmedabad-380009, India.

Rakesh M Rawal Department of Life Sciences, University School of Sciences, Gujarat University, Ahmedabd-380016, India. - Corresponding author

ABSTRACT

To shorlist genes involved in development of glioblastoma (GB), we performed a meta-analysis of GB microarray data sets to indentify differentially expressed genes (DEGs) and annotate them to understand their biological functions in primary tumor. Publicly available seven GEO datasets of GB were used to perform a meta-analysis. Differentially expressed 537 DEGs across the GB population 160, were up-regulated and 377, down-regulated. GO terms for molecular functions significantly enriched were in protein binding and metal ion binding, while for biological processes, the enriched GO terms were signal transduction and apoptotic process, and for cellular component, the enriched GO terms were cytoplasm and nucleus. The most significant pathway in our KEGG analysis was pathways in cancer. PPI network analysis indicated that the significant hub proteins containing DOPEY1, NCBP2 and EIF4E2, USP46 and CDC37L1. Our meta-analysis identified DEGs and biological functions integrated with gene expression changes in GB, which is useful further identification and classification of GB among all brain neoplasm.

KEYWORDS:

Glioblastoma, Meta-analysis, Differentially expressed genes

1 Introduction

Glioblastoma (GB) is an aggressively progressing most common primary brain tumor in adults with morbidity and mortality globally including India, increased in adults mainly **(Jones C et al., 2012).** The standard treatment includes surgical treatment, immunotherapy and radiotherapy. Temozolomide is the commonly used chemotherapeutic drug **(Khosla 2016).** There are many hypothesis prevailing on glioblastoma histology in adults. Reported scenario patients survival have two major confounders one at age diagnosis and the other is at functional stage diagnosis; histologic grade and genetic markers **(Walid, M. S. 2008).**

High throughput technology eg. microarray demonstrated that metaanalysis based on multiple studies have the ability to identify novel gene signature, expressions profiles, to predict treatment response **(Ramaswamy S et al., 2002).** Despite this, many microarray-based studies have shown many times those results were not reproducible despite of this meta-analysis on large datasets required **(Ntzani EE et al., 2003).** Thus we performed meta-analysis on large GB datasets to indentify DEGs.

2. Materials and methods

2.1. Data retrieval from the Database

Seven datasets were used in this study are publicly available in ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) and downloaded from as the raw data (.CEL files) of microarray experimenta **(Parkinson et al., 2007)** as listed in Table 1.

Table 1 Summary of GB datasets

Datasets	Author and year	Array type	Total
GSE82009	Nutt CL et. al, 2003	HG_U95Av2	50
GSE43378	Kawaguchi A et.al, 2013	HG-U133_Plus_2	50
GSE51395	Chong YK et.al, 2016	HG-U133_Plus_2	20
GSE51062	Walsh AM et.al, 2015	HG-U133_Plus_2	52
GSE68850	No_publication	HG-U133_Plus_2	228
GSE60184	Li J et.al, 2014	HG-U133_Plus_2	23
GSE53733	Reifenberger G et.al, 2014	HG-U133_Plus_2	70

2.2. Gene expressions analysis of multiple microarray datasets

R Package version 3.3.1 (www.bioconductor.org/) with limma package **(Matthew E. Ritchie et al., 2015)** was used for future analysis. Retrieved data was converted to probe-level into expression values which have background correction; quantile normalization and probe summarization were performed. t-test **(Marine Jeanmougin et al., 2010)** was used to identify the significantly expressed DEGs in glioma patients, with a combination of $P<0.05$ and the $|\log2FC$ (fold change)$| >1$ used as the threshold. A heat map was also generated.

2.3. Gene ontology [GO] and Pathway enrichment analysis of DEGs

Gene Ontology (GO) enrichment analysis was also performed using online software GENECODIS (http://genecodis.cnb.csic.es) **(Tabas-Madrid D et al., 2012).**

2.4. Protein-Protein (PPI) network construction

In the present study, STRING tool (STRING; string-db.org/) was used to analyze the interactions between DEGs with the PPI required confidence (combined score, 0.9) and a PPI network was constructed **(Szklarczyk D et al., 2011).**

3. Results and Discussion

3.1. Identification and comparison of genes associated with GB

Differential gene expression study results the expression value logarithmically transformed to base 2 with 13,477 genes obtained. In the first screening of gene list top 100 genes were classified with a P values close to <0.05 and by applying a minimal fold change of 1.4, A total of we identified 537 DEGs from GBs in which primary tumor samples were compared, including 160 up and 377 down-regulated genes. A list of the top 10 most significantly up- or down-regulated DEGs are listed in Table 2.

The up-regulated gene with lowest P-value (P = 2.02E-16) was NCBP2 (Nuclear Cap Binding Protein Subunit 2), which previously reported to correlate with glioma low grade tumor growth, metastasis and patients survival, which is a part of nuclear cap binding complex (CBC) that is required for cell viability and poly (A) RNA. Alternate CBC is mainly plays role in mRNA biogenesis (Anna Gebhardt et al., 2015). We found that some genes have been closely related to the development of GB among top 10 up-regulated DEGs, such as USP46, DOPEY1, GFMB, MTCH2, C9orf91, IQGAP1, GNAQ and HADHA. According to previous studies USP46 is reported as flanking gene causes tumor progression or genesis in GB (Holtkamp et al., 2007). GMFB gene causes neural transmission and neurodevelopment (Gratacos M et al., 2007) in brain but it is not reported in GB. MTCH2 gene causes glioblastoma but its expression was not found up regulated in the published studies (Cristofanon, S., & Fulda, S. 2012). C9orf91 gene is involved in cell cycle progression due to its dysregulation of it in

G0/G1 phase cell-accumulation (Peter S. Linsley et al., 2007). IQGAP1 gene is involved in survival outcomes of glioma patients as a prognosis biomarker (McDonald KL et al., 2007). HADHA gene is down regulated and plays putative involvement in breast cancer (Manju Mamtani, 2012). GNAQ gene is primary melanocytic neoplasm of the central nervous system are uncommon neoplasms derived from melanocytes. and also affects pathogenesis affects mitogen-activated kinase signaling in cancer (Kusters-Vandevelde et al., 2012). MARCKS gene is present in nervous tissue where Jun N-terminal protein kinase (JNK) forms neural tube formation. This gene is phosphorylated by JNK on C-terminal residues but when its expression increased causes cancer types (Malchinkhuu E et al., 2005). PXPM2 is a peroxisomal membrane protein involved in oxidative stress response and also plays pivotal role in detoxification, fatty acid oxidation and regulation of oxygen. Due to interruption of this protein causes another proteins dyregulations (Zhang et al., 2012).

The down-regulated gene with the lowest P-value (P = 1.56E-12) was EIF4E2 (Eukaryotic Translation Initiation Factor 4E Family Member 2) which is also a cap-binding protein, that acts as wide variety of mRNA to active translation at polysomes (J. Uniacke et al., 2012).

Table 2 The top 10 most significantly expressed up- or down-regulated DEGs

Gene ID	Gene symbol	P-value	Fold change	Official full name
Up-regulated genes				
201517_at	NCBP2	2.02E-16	1.99	Nuclear Cap Binding Protein Subunit 2
203870_at	USP46	1.76E-13	2.01	Ubiquitin Specific Peptidase 46
222403_at	MTCH2	3.65E-13	2.34	Mitochondrial Carrier 2
202544_at	GMFB	6.44E-12	1.27	Glia Maturation Factor Beta
221865_at	C9orf91	1.53E-11	1.49	Chromosome 9 Open Reading Frame 91
210840_s_at	IQGAP1	1.76E-11	1.77	IQ Motif Containing GTPase Activating Protein 1
208629_s_at	HADHA	2.34E-11	1.42	Hydroxyacyl-CoA Dehydrogenase/3-Ketoacyl-CoA Thiolase/Enoyl-CoA Hydratase (Trifunctional Protein), Alpha Subunit
224861_at	GNAQ	1.97E-10	1.32	G Protein Subunit Alpha Q
200644_at	MARCKSL1	2.33E-10	2.49	MARCKS Like 1
219076_s_at	PXMP2	2.87E-10	1.39	Peroxisomal Membrane Protein 2
Down-regulated genes				
213571_s_at	EIF4E2	1.56E-12	-1.87	Eukaryotic Translation Initiation Factor 4E Family Member 2
219646_at	DEF8	1.56E-11	-1.72	Differentially Expressed In FDCP 8 Homolog
1554453_at	HNRNPLL	1.87E-11	-1.58	Heterogeneous Nuclear Ribonucleoprotein L like
228846_at	MXD1	1.14E-10	-2.30	MAX Dimerization Protein 1
222523_at	SENP2	1.30E-10	-0.95	SUMO1/Sentrin/SMT3 Specific Peptidase 2
218265_at	SECISBP2	1.94E-10	-1.28	SECIS Binding Protein 2
213271_s_at	DOPEY1	2.00E-10	-1.39	Dopey Family Member 1
219343_at	CDC37L1	2.14E-10	-1.20	Cell Division Cycle 37 Like 1
1556285_s_at	PPA2	6.01E-10	-0.91	Pyrophosphatase (Inorganic) 2
221903_s_at	CYLD	1.05E-09	-1.33	CYLD Lysine 63 Deubiquitinase

3.2. Functional annotation

To investigate the biological roles of DEGs from GB, GO categories are classified into the three groups eg. biological process, cellular components, and molecular functions. These groups are thus examined separately using GENECODIS. Gene that shows a nominal significance level of P < 0.01 were selected and tested against background set of all the genes with GO annotations. We found GO terms for molecular functions significantly enriched in protein binding (GO:0005515, P=4.95E-27), and metal ion binding (GO:0005509, P=1.09E-07), while for biological processes, the enriched GO terms were signal transduction (GO:0007165, P=2.03E-06) and apoptotic process (GO:0006915, P=1.42E-06), and for cellular compartments, the enriched GO terms were cytoplasm (GO:0005737, P=5.36E-32) and nucleus (GO:0005634, P=2.57E-23) (Figure 1).

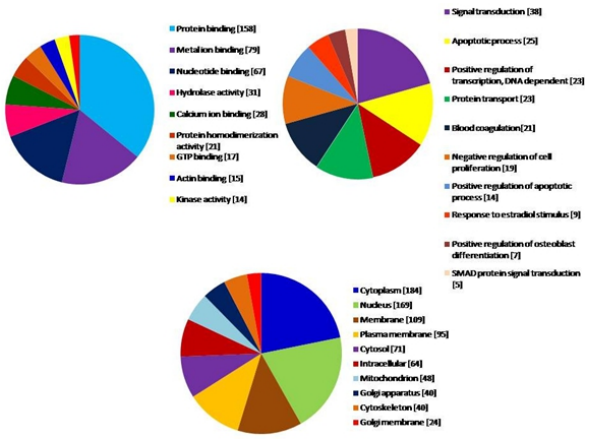

Figure 1 Top 15 GO enriched terms for differentially expressed genes. (A). molecular Functions for DEGs (B). biological process for DEGs (C). cellular component for DEGs

Further evaluation of biological significance for the DEGs, we also performed KEGG pathway enrichment analysis. This analysis was based on hypergeometric test with P value < 0.05 was used as the criteria for pathways identification. The most significant pathway present in KEGG analysis were Pathways in cancer (P = 3.36E-05), Amoebiasis (P = 9.79E-05) and Regulation of actin cytoskeleton (P = 1.76E- 06) (Table 3). The most significant pathways in cancer is associated with several cancers and its causing processes such as cell cycle progression, growth, apoptosis etc. However, this pathway is observed in human by cell cycle desregulation, genetic alterations to show contradictory changes in normal to human cells (Kelley et al., 2014).

Table 3 The top 15 enriched KEGG pathway of DEGs

KEGG ID	KEGG Name	No. of genes	P value
hsa05200	Pathways in cancer	19	3.36E-05
hsa05146	Amoebiasis	10	9.79E-05
hsa04810	Regulation of actin cytoskeleton	14	1.76E-06
hsa05211	Renal cell carcinoma	8	6.24E-06
hsa04510	Focal adhesion	12	2.48E-05
hsa04666	Fc gamma R-mediated phagocytosis	8	4.70E-05
hsa05142	Chagas disease (American trypanosomiasis)	8	9.82E-05
hsa04010	MAPK signaling pathway	13	9.61E-05
hsa03015	mRNA surveillance pathway	7	0.0002
hsa04540	Gap junction	7	0.0002
hsa05100	Bacterial invasion of epithelial cells	6	0.0005
hsa04144	Endocytosis	10	0.0005
hsa04380	Osteoclast differentiation	8	0.0006
hsa05220	Chronic myeloid leukemia	6	0.0006
hsa04660	T cell receptor signaling pathway	7	0.0008

3.3. Protein-Protein interaction (PPI) network construction

The PPI network was established for the top 10 up-regulated and

down-regulated DEGs by Cytoscape software included 196 nodes and 188 edges. The significant hub proteins containing DOPEY1 (Dopey Family Member 1, Degree= 18), NCBP2 (Nuclear Cap Binding Protein Subunit 2) and EIF4E2 (Eukaryotic Translation Initiation Factor 4E Family Member 2) have same Degree=12. Both proteins are involved in mRNA initiation causes cell proliferation, apoptosis, cell migration and cell death also reported in glioma studies **(WEI et al., 2015).** Other two hub proteins namely USP46 (Ubiquitin Specific Peptidase 46) and CDC37L1 (Cell Division Cycle 37 Like 1) have same Degree= 11 **(Figure 2).** In the further more analysis results from PPI network of the top 10 up-regulated and down-regulated DEGs indicates the significant hub protein containing DOPEY1, NCBP2, EIF4E2, USP46 and CDC37L1. DOPEY1, a member of DOPEY family involved in the traffic of myelin formation **(Tanaka M et al., 2012).** Results of PPI suggested that these described hub proteins may also play important role in the development of GB.

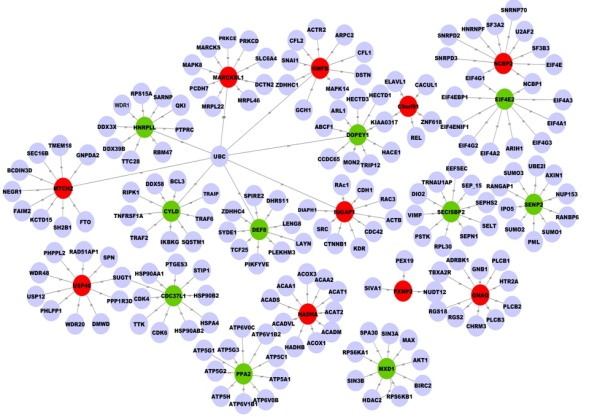

Figure 2. Protein-protein interaction network for top 10 up and down regulated DEGs. Red coloured nodes are up-regulated, Green coloured nodes are down regulated, and others nodes are predicted PPI interactions.

4. Conclusion

In conclusion, using this meta-analysis based gene expression data of glioblastoma we have found the underlying molecular mechanism differences between various tumor cells and histology. Another, important aspect is including DEGs and their biological function which may contribute to the successful identification of therapeutic targets for glioblastoma. Further functional annotation studies may provide additional knowledge to understand role of differentially expressed genes in the pathophysiology of glioblastoma.

5. References

1. Jones C, Perryman L, Hargrave D., 2012. Paediatric and adult malignant glioma: close relatives or distant cousins?. Nat Rev Clin Oncol. 9(7):400-13.
2. Khosla, D, 2016. "Concurrent therapy to enhance radiotherapeutic outcomes in glioblastoma.". Annals of translational medicine. 4 (3): 54.
3. Walid, M. S., 2008. Prognostic Factors for Long-Term Survival after Glioblastoma. The Permanente Journal. 12(4), 45–48.
4. Ramaswamy S, Golub TR., 2002. DNA microarrays in clinical oncology. Journal of Clinical Oncology. 20:1932.
5. Ntzani EE, Ioannidis JP., 2003. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet. 362:1439-1444.
6. Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, et al., 2007. ArrayExpress—a public database of microarray experiments and gene expression profiles. Nucleic Acids Research, 35(Database issue), D747–D750.
7. Nutt CL, Mani DR, Betensky RA, Tamayo P et al., 2003. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res. 63(7):1602-7.
8. Kawaguchi A, Yajima N, Tsuchiya N, Homma J et al., 2013. Gene expression signature-based prognostic risk score in patients with glioblastoma. Cancer Sci. 104(9):1205-10.
9. Chong YK, Sandanaraj E, Koh LW, Thangaveloo M et al., 2016. ST3GAL1-Associated Transcriptomic Program in Glioblastoma Tumor Growth, Invasion, and Prognosis. J Natl Cancer Inst. 108(2).
10. Reifenberger G, Weber RG, Riehmer V, Kaulich K et al., 2014. Molecular characterization of long-term survivors of glioblastoma using genome- and transcriptome-wide profiling. Int J Cancer. 135(8):1822-31.
11. Matthew E. Ritchie, Belinda Phipson, Di Wu, et al., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43 (7): e47.
12. Marine Jeanmougin ,Aurelien de Reynies,Laetitia Marisa, et al., 2010. Should We Abandon the t-Test in the Analysis of Gene Expression Microarray Data: A Comparison of Variance Modeling Strategies. PLoS ONE 5(9): e12336.
13. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A., 2015. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res.15(Web Server issue):W478–W483.
14. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al., 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39(Database issue): D561–D568. 2011.
15. Anna Gebhardt, Matthias Habjan, Christian Benda, Arno Meiler, Darya A. Haas, Marco Y. Hein et al., 2015. mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. Nature Communications 6, Article number: 8192.
16. Holtkamp, N., Ziegenhagen, N., Malzer, E., Hartmann, C., Giese, A., & von Deimling, A., 2007. Characterization of the amplicon on chromosomal segment 4q12 in glioblastoma multiforme. Neuro-Oncology. 9(3), 291–297.
17. Gratacos M, Costas J, de Cid R, Bayes M, Gonzalez JR, et al., 2009. The Psychiatric Genetics Network Group. Identification of New Putative Susceptibility Genes for Several Psychiatric Disorders by Association Analysis of Regulatory and Non-Synonymous SNPs of 306 Genes Involved in Neurotransmission and Neurodevelopment. Am J Med Genet Part B. 150B:808–816.
18. Cristofanon, S., & Fulda, S., 2012. ABT-737 promotes tBid mitochondrial accumulation to enhance TRAIL-induced apoptosis in glioblastoma cells. Cell Death & Disease, 3(11), e432.
19. Peter S. Linsley, Janell Schelter, Julja Burchard, Miho Kibukawa, et al., 2007. Transcripts Targeted by the MicroRNA-16 Family Cooperatively Regulate Cell Cycle Progression. Mol Cell Biol. 27(6): 2240–2252.
20. McDonald KL, O'Sullivan MG, Parkinson JF, Shaw JM, Payne CA, et al., 2007. IQGAP1 and IGFBP2: valuable biomarkers for determining prognosis in glioma patients. J Neuropathol Exp Neurol. 66(5):405-17.
21. Manju Mamtani and Hemant Kulkarni, 2012. Association of HADHA expression with the risk of breast cancer: targeted subset analysis and meta-analysis of microarray data. BMC Research Notes5:25.
22. Kusters-Vandevelde, H.V.N., Klaasen, A., Küsters, B. et al., 2010. Activating mutations of the GNAQ gene: a frequent event in primary melanocytic neoplasms of the central nervous system. Acta Neuropathol. 119: 317.
23. Malchinkhuu E, et al., 2005. Role of p38 mitogen-activated kinase and c-Jun terminal kinase in migration response to lysophosphatidic acid and sphingosine-1-phosphate in glioma cells.Oncogene. 24: 6676– 6688
24. Zhang, Y., Baker, S. S., Baker, R. D., Zhu, R., & Zhu, L., 2012. Systematic Analysis of the Gene Expression in the Livers of Nonalcoholic Steatohepatitis: Implications on Potential Biomarkers and Molecular Pathological Mechanism. PLoS ONE, 7(12), e51131.
25. James Uniacke, J. Kishan Perera, Gabriel Lachance, Camille B. Francisco and Stephen Lee, 2014. Cancer Cells Exploit eIF4E2-Directed Synthesis of Hypoxia Response Proteins to Drive Tumor Progression. Microenvironment and Immunology. Cancer.Res. 74(5); 1379–89.
26. Chunlong Zhan, Chunquan Li, Jing Li, Junwei Han, Desi Shang, Yunpeng Zhang, Wei Zhang et al., 2014. Identification of miRNA-Mediated Core Gene Module for Glioma Patient Prediction by Integrating HighThroughput miRNA, mRNA Expression and Pathway Structure. 9(5): e96908.
27. Kelley, M. R., Logsdon, D., & Fishel, M. L., 2014. Targeting DNA repair pathways for cancer treatment: what's new? Future Oncology (London, England). 10(7), 1215–1237.
28. WEI, S., WANG, Y., XU, H., & KUANG, Y., 2015. Screening of potential biomarkers for chemoresistant ovarian carcinoma with miRNA expression profiling data by bioinformatics approach. Oncology Letters. 10(4), 2427–2431.
29. Tanaka M, Izawa T, Yamate J, Franklin RJ, Kuramoto T, Serikawa T, et al. (2014) The VF rat with abnormal myelinogenesis has a mutation in Dopey1. J Glia 62(9):1530–1542.