Bioinformatics Analysis Reveals that IGFBP2, EMP3, PDPN, and MCUB Genes were Potential Biomarkers for Glioblastoma

Xiaobin Luo¹, Fubing Yang²

ABSTRACT

Introduction: The GSE4290 dataset in Gene Expression Omnibus was used for a weighted gene co-expression network analysis (WGCNA). In this study, we used the GSE4290 dataset in the Gene Expression Omnibus (GEO) database for a weighted gene co-expression network analysis (WGCNA) to identify differentially expressed genes between non-tumor tissues and tumor tissues.

Material and methods: Modules related to the GBM phenotype were identified and gene enrichment analyses, expression analyses, and survival analysis were performed.

Results: Genes in the module that was most closely related to GBM were enriched for functions in extracellular matrix organization, blood vessel development, response to growth factors and proteoglycans in gene enrichment analysis. Four genes significantly reduced the overall survival rate of patients. And IGFBP2, EMP3, PDPN, and MCUB with high sensitivity and specificity in ROC curve analysis.

Conclusion: These four genes play important roles in the growth and development of GBM and can be used as therapeutic biomarkers.

Keywords: Glioblastoma, Biomarkers, Bioinformatics, ROC Curve, Oncomine, Metascape.

INTRODUCTION

Glioblastoma (GBM) is the most malignant tumor in the central nervous system, accounting for 12–15% of intracranial tumors. The incidence is about 3/100,000 in the United States¹, about 4.64/100,000 in the United Kingdom (UK).² The prognosis of patients with GBM is extremely poor, with a median survival of approximately 15 months.³ Stupp et al⁴ found that patients who receive the standard treatment have 2-year and 5-year mortality rates of 70% and 90%, respectively. The invasiveness and infiltrative nature of GBM cells make complete removal during surgery difficult, contributing to a high rate of recurrence.⁵ Furnari et al⁶ found that approximately 70% of patients with low-grade glioma develop GBM within 5-10 years. Therefore, studies of the molecular mechanism underlying GBM are needed to guide the development of approaches for repressing the occurrence and development of GBM.

Current treatment methods for GBM mainly include surgery, radiotherapy, chemotherapy, and immunotherapy. Owing to the high degree of malignancy, genetic variation, and the lack of information regarding underlying mechanisms, the diagnosis and treatment of GBM is deadlocked. However, a bioinformatics approach can be used to clarify various mechanisms underlying GBM, providing a basis for targeted treatment. Bioinformatics analyses, which are reproducible and independent⁷, are widely used for molecular diagnosis, therapeutic evaluation, prognosis, and prediction of tumor sensitivity to drugs.⁸ Therefore, this approach is well-suited for studies of GBM.

In this study, we used the GSE4290 dataset in the Gene Expression Omnibus (GEO) database for a weighted gene co-expression network analysis (WGCNA) to identify differentially expressed genes between non-tumor tissues and tumor tissues. Genes in the module that was most closely related to GBM were evaluated by a survival analysis and an expression analysis. We identified four genes that are closely related to GBM (i.e., IGFBP2, EMP3, PDPN, and MCUB). We further compared the expression of these genes between GBM and non-tumor tissues. Based on the Receiver Operating Characteristics (ROC) curve, IGFBP2 had the largest area under the curve (AUC), suggesting that this gene is a therapeutic target for GBM.

MATERIAL AND METHODS

GSE4290 was downloaded from Gene Expression Omnibus (GEO, (https://www.ncbi.nlm.nih.gov/geo/). The dataset were submitted by Fine et al⁹ and were obtained using the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). A total of 180 samples were included in the GSE4290 dataset, including 23 non-tumor samples, 26 astrocytoma samples, 50 oligodendroglioma samples, and 81 GBM samples. An expression matrix was obtained by performing background correction and data processing using the original downloaded data with the range migration algorithm (RMA) implemented in the affy package. Then, the expression matrix was annotated (ID-converted) and repetitive genes were removed to obtain an expression matrix. Finally, the variance values of all genes were calculated and the corresponding genes with the first 25% variance were identified and gene enrichment analyses, expression analyses, and survival analysis were performed. We identified four genes that are closely related to GBM (i.e., IGFBP2, EMP3, PDPN, and MCUB). We further compared the expression of these genes between GBM and non-tumor tissues. Based on the Receiver Operating Characteristics (ROC) curve, IGFBP2 had the largest area under the curve (AUC), suggesting that this gene is a therapeutic target for GBM.

How to cite this article: Xiaobin Luo, Fubing Yang. Bioinformatics analysis reveals that IGFBP2, EMP3, PDPN, and MCUB genes were potential biomarkers for glioblastoma. International Journal of Contemporary Medical Research 2020;7(10):J1-J9.

DOI: http://dx.doi.org/10.21276/ijcmr.2020.7.10.25
selected for WGCNA analysis.
Construction of a gene co-expression network and identification of modules associated with GBM.
The R (version was 3.5.2) package WGCNA \(^{10}\) was used to construct the gene co-expression network. First, genes and samples were selected as described above to construct the sample clustering tree. After eliminating outliers, a sample clustering tree was reconstructed with phenotypes. Second, by limiting the soft threshold, the module contained at least 30 genes, a cut-off line of 0.25 was set, and similar modules were merged. Finally, a correlation map between module genes and phenotype was drawn based on clinical information.

**Module screening and gene enrichment analysis**
Genes in the module that was most closely related to the GBM phenotype were selected for an enrichment analysis. Metascape (http://metascape.org/gp/index.html), a web-based bioaccumulation analysis tool, integrates the functions of (Gene Ontology) GO, (Kyoto Encyclopedia of Genes and Genomes) KEGG, UniProt, and DrugBank authoritative databases, which are frequently updated.\(^{11}\) Using Metascape, a hypergeometric distribution test was used to calculate P-values, the Benjamini–Hochberg method was used to calculate Q-values\(^{12}\), and Kappa scores were used as a similarity measure.\(^{13}\) Then, the protein-protein interaction (PPI) network was constructed using STRING (https://string-db.org/cgi/input.pl?sessionId=BBPrHBBKJPUey&input_page_show_search=on, version was 11.0) online tools, and key modules were obtained using the Cytoscape (http://www.cytoscape.org/, version was 3.7.1) plug-in molecular complex detection (MCODE, http://apps.cytoscape.org/apps/mcode, version was 1.5.1). Since the default K-core value is 2, a total of 19 modules were obtained. To filter the modules, the K-core value was changed to 6, and 4 modules were obtained.

**Screening of key genes**
The blue module, which was closely related to GBM phenotype, was selected, and module membership and gene importance were calculated according to the Pearson correlation coefficient. The P-value was obtained by the t-test, and P-values of less than 0.01 were significant. There was a clear correlation between genes and phenotypes. The top 10 genes related to the GBM phenotype were identified as key genes and used for subsequent analyses.

**Verifying key genes using the Oncomine database**
The Oncomine (https://www.oncomine.org/resource/login.html) database is currently the world's largest oncogene chip database, including 715 data sets and 86,733 samples.\(^{14–16}\) Using the key genes selected in the previous step, datasets with clinical outcomes were retrieved to evaluate expression. The survival status and follow-up time were retrieved separately, and a survival analysis was performed according to gene expression and survival status using Graphpad Prism (https://www.graphpad.com/, version was 7.0). Four genes closely related to the survival of patients with GBM were identified. The expression levels of these four genes in non-tumorous tissues and GBM tissues were sorted and analyzed by an receiver operating characteristic (ROC) curve, and the prediction efficiency of genes was evaluated based on the Area Under Curve (AUC).

![Figure-1:](image-url)
RESULTS

Data preprocessing
The original CEL data were normalized and background corrected by the RMA algorithm to obtain an expression matrix. Then, ID conversion was performed using the corresponding platform, and an expression matrix without duplicate genes was obtained by taking the median expression value for each gene and eliminating duplicate genes. A total of 20186 genes were obtained. Finally, the first 25% of the variance corresponds to 5,047 genes.

Gene co-expression module analysis
A gene co-expression analysis was performed using the R-based WGCNA package. Two abnormal expression samples were found based on the gene clustering tree, GSM97895 and GSM97932. After deleting the abnormal samples, 5047 genes and 178 samples were used to construct a co-expression network (Fig. 1). When the soft threshold of the adjacent matrix was set to 7, SFT.R2 was 0.857 (Fig. 2).
A total of 11 modules were obtained for the dynamic cut tree (Fig. 2C). After color-coding the modules, 127 non-co-expressed genes in the grey module were found, and the ratio of grey genes to the total genes was about 2.5%. After merging similar modules, the relationship between the sample clustering tree and the gene module was obtained (Fig. 2C). The gene modules and traits were then linked to obtain a correlation map for both (Fig. 2D). There was a very high correlation between the blue module and the GBM phenotype (Fig. 2E). Although the green module was also highly correlated with GBM, the correlation was negative and the number of genes was small. In a two-module integration analysis (STRING-PPI-CytoHubba-hubgenes), the first ten hub genes were found. There was only one green module gene, and it had a low correlation with GBM. Accordingly, subsequent analyses focused on the blue module gene.

**Module gene enrichment analysis**

The correlation coefficient for the relationship between the

---

**Figure-3:**

- **A**: Distribution of module sizes with a -log10(P) scale.
- **B**: Visualization of the correlation network.
- **C**: Module networks MCODE-1 and MCODE-2.
- **D**: Module networks MCODE-3 and MCODE-4.
blue gene module and GBM was 0.65 and the module included 860 genes. We performed an enrichment analysis of these 860 genes using databases, such as GO Biological Processes, KEGG Pathway, Reactome Gene Sets, Canonical Pathways, and CORUM. The results of this analysis are summarized in Figure 3A, and an enrichment pathway map (Fig. 3B) was constructed. A GO Biological Processes analysis showed enrichment for extracellular matrix organization, blood vessel development, and response to growth factors. A KEGG pathway analysis showed enrichment for proteoglycans in cancer. Enrichment for ECM proteoglycans was detected using the Reactome Gene Sets database. Enrichment for the PID INTEGRIN3 PATHWAY in the Canonical Pathways database was detected. Based on a STRING online analysis, a PPI network was constructed. When the K-core value was set to 6, a total of 4 modules were obtained (Fig. 3C). The results of the first three enrichment analyses for each module are shown in Table 1. The MCODE1 analysis indicated
enrichment for extracellular matrix organization, integrin cell surface interactions, and blood vessel development. The genes in MCODE2 were enriched in extracellular matrix organization, post-translational protein phosphorylation, and non-integrin membrane-ECM interactions. The genes in MCODE3 were enriched in cellular zinc ion homeostasis, and metallothioneins bind metals. The genes in MCODE4 were enriched in smooth muscle contraction, RHO GTPases activate PKNs, and neutrophil degranulation.

Screening of key genes
The first ten genes (Table 2) with the highest correlations with the GBM phenotype (highest GS value) in the blue gene module were obtained using the WGCNA package. They were Insulin-like growth factor-binding protein 2 (IGFBP2), Vacuole membrane protein 1 (VMP1), Epithelial membrane protein 3 (EMP3), Serpin H1 (SERPINH1), Chloride intracellular channel protein 1 (CLIC1), Collagen alpha-2(V) chain (COL5A2), Podoplanin (PDPN), Calcium uniporter regulatory subunit MCUb (MCUB), Proteolipid protein 2 (PLP2), and Collagen alpha-1(IV) chain (COL4A1).

Verification of key genes
Datasets in the Oncomine database were retrieved, and The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) Brain dataset was identified as suitable for further analysis. It included mRNA data for 547 tumor tissues and 10 normal tissues as well as data for age, gender, and survival. A gene found in the search was not expressed in the data set, SERPINH1. Expression levels of the remaining genes differed significantly between normal and tumor tissues (Fig. 4A, C, E, G). However, only 4 genes were significant in the survival analysis, i.e., IGFBP2, EMP3, PDPN, and MCUB (Fig. 4B, D, F, H). In the ROC curve analysis (Fig. 5), the AUC of IGFBP2 was 0.9683, the AUC of EMP3 was 0.9018, the AUC of PDPN was 0.9152, and the AUC of MCUB was 0.8884.

<table>
<thead>
<tr>
<th>MCODE</th>
<th>GO and KEGG</th>
<th>Description</th>
<th>Log10(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCODE-1</td>
<td>R-HSA-1474244</td>
<td>Extracellular matrix organization</td>
<td>-40.85</td>
</tr>
<tr>
<td>MCODE-1</td>
<td>R-HSA-216083</td>
<td>Integrin cell surface interactions</td>
<td>-23.70</td>
</tr>
<tr>
<td>MCODE-1</td>
<td>GO:0001568</td>
<td>blood vessel development</td>
<td>-22.69</td>
</tr>
<tr>
<td>MCODE-2</td>
<td>R-HSA-1474244</td>
<td>Extracellular matrix organization</td>
<td>-31.97</td>
</tr>
<tr>
<td>MCODE-2</td>
<td>R-HSA-8957275</td>
<td>Post-translational protein phosphorylation</td>
<td>-27.79</td>
</tr>
<tr>
<td>MCODE-2</td>
<td>R-HSA-3000171</td>
<td>Non-integrin membrane-ECM interactions</td>
<td>-11.37</td>
</tr>
<tr>
<td>MCODE-3</td>
<td>GO:0006882</td>
<td>cellular zinc ion homeostasis</td>
<td>-19.15</td>
</tr>
<tr>
<td>MCODE-3</td>
<td>R-HSA-5661231</td>
<td>Metallothioneins bind metals</td>
<td>-15.42</td>
</tr>
<tr>
<td>MCODE-4</td>
<td>R-HSA-445355</td>
<td>Smooth Muscle Contraction</td>
<td>-12.71</td>
</tr>
<tr>
<td>MCODE-4</td>
<td>R-HSA-5625740</td>
<td>RHO GTPases activate PKNs</td>
<td>-11.83</td>
</tr>
<tr>
<td>MCODE-4</td>
<td>R-HSA-6796895</td>
<td>Neutrophil degranulation</td>
<td>-8.73</td>
</tr>
</tbody>
</table>

Table-1: Analysis of the blue gene module.
DISCUSSION

GBM is an intracranial tumor with very poor prognosis. Current research is focused on the treatment of GBM. Altinoz et al. found that noscapine blocks the inhibitory effects of NF-κB on tumor growth and enhances sensitivity to chemotherapy. Stupp et al. found that temozolomide combined with radiotherapy and chemotherapy can significantly improve progression-free survival and overall survival. Drugs targeting O6-methylguanine DNA methyltransferase (MGMT) and Epidermal growth factor receptor (EGFR) are currently undergoing Phase 3 clinical trials, with the potential to revolutionize the treatment of GBM. We identified 860 genes in the blue module, and these genes were enriched for extracellular matrix organization, blood vessel development, response to growth factors, the proteoglycans in cancer pathway, and ECM proteoglycans. Proteoglycans are a major component of the extracellular environment of the brain and play important roles in tumor proliferation, metastasis, and angiogenesis. In an analysis of canonical pathways, there was enrichment for the PID INTEGRIN3 PATHWAY. An MCODE module analysis also indicated that the genes were enriched for functions in extracellular matrix organization, blood vessel development, and integrin cell surface interactions. Among the ten key genes, four genes were associated with reductions in overall survival. Based on the ROC curve analysis, these four genes have high sensitivity and may be biomarkers of GBM. The IGFBP2 gene encodes Insulin Like Growth Factor Binding Protein 2 (IGFBP2), which promotes the self-renewal and proliferation of mouse neural stem cells. It is also involved in GBM invasion, migration, and progression as well as the maintenance of GBM cell phenotype. IGFBP2 is highly expressed in tumor tissues and is expressed at low levels in normal tissues. Dunlap et al. found that both IGFBP2 and platelet-derived growth factor beta (PDGFB) drive tumor development. The IGFBP2/integrin/ILK/NF-κB network drives the progression of gliomas in vivo. Insulin like growth factor 1 receptor (IGFIR) activates HIF1α and up-regulates the expression of Insulin-like growth factor 2 (IGF2) and IGFBP, and interference with HIF1α can inhibit the growth of GBM. IGFBP2 is involved in immunosuppression and is a biomarker of poor prognosis. Multiple studies have shown that IGFBP2 is associated with the prognosis of GBM, and high expression can reduce patient survival. IGFBP2 also enhances CD144 and Matrix Metalloproteinase 2 (MMP2) expression and is associated with angiogenic mimicry, which can make GBM resistant to Vascular Endothelial Growth Factor (VEGF) therapy. In summary, IGFBP2 is related to various biological behaviors of GBM and significantly reduces the overall survival rate of patients, indicating that it is a potential target for GBM treatment.

The EMP3 gene is a member of the Epithelial Membrane Protein (EMP) family and is expressed in embryonic tissues, ovaries, intestines, and leukocytes. It is differentially expressed between tumor tissues and non-tumor tissues. It is related to cell proliferation, differentiation, apoptosis, and migration. High expression in oesophageal squamous cell carcinoma and non-small cell lung cancer can inhibit tumor cell growth. However, this gene does not have inhibitory effects in all tumors. In low-grade glioma, the methylation of EMP3 affects the survival rate. Moreover, the EMP3 gene is methylated in GBM, and hypermethylation leads to an increase in patient survival and mortality. Recent studies have shown that high expression of EMP3 in GBM affects prognosis. Our results indicated that EMP3 is highly expressed in GBM cells and leads to a decrease in overall survival. Accordingly, the gene is a potential biomarker for GBM treatment.

PDNP is a type I transmembrane mucin-like glycoprotein expressed in the lymphatic endothelium. It is related to the invasion and proliferation of GBM. Studies have shown that gene silencing and inhibitors result in a significant reduction in GBM proliferation and migration. A decrease in expression can also prolong the survival rate of patients. The MCUB gene is a member of the mitochondrial transporter family, which has not been found to have a correlation with GBM but can affect mitochondrial uptake of calcium ions. Our results showed that the overexpression of the MCUB gene in GBM significantly reduces overall survival, suggesting that it may be a biomarker for GBM.

CONCLUSION

In conclusion, we used a bioinformatics approach to identify differentially expressed genes between GBM and non-tumor tissues and further explored the correlation between survival and the expression of the key genes IGFBP2, EMP3, PDNP, and MCUB in patients with GBM. Combined with the results of previous studies, these four genes may be potential targets for GBM treatment. We will study the roles of these four genes in GBM cells in the future; we expected these genes to provide a reference for the diagnosis and treatment of GBM.

REFERENCES

5. Koo S, Martin GS, Schulz KJ, Ronck M, Toussaint LG. Serial selection for invasiveness increases expression of miR-143/miR-145 in glioblastoma cell lines. BMC


39. Li KKW, Pang JC-S, Chung NY-F et al. EMP3 overexpression is associated with oligodendroglial...
tumors retaining chromosome arms 1p and 19q. Int. J.
40. Alaminos M, Dávalos V, Ropero S et al. EMP3, a
myelin-related gene located in the critical 19q13.3
region, is epigenetically silenced and exhibits features
of a candidate tumor suppressor in glioma and
41. Guo X-X, Su J, He X-F. A 4-gene panel predicting the
2019;23:30-38.
T, Matsutani M. Increased expression of podoplanin
in malignant astrocytic tumors as a novel molecular
2006;111:483–488.
43. Raica M, Cimpean AM, Ribatti D. The role of
podoplanin in tumor progression and metastasis.
44. Cortez MA, Nicoloso MS, Shimizu M et al. miR-
29b and miR-125a regulate podoplanin and suppress
invasion in glioblastoma. Genes Chromosomes Cancer
45. Peterziel H, Müller J, Danner A et al. Expression
of podoplanin in human astrocytic brain tumors
is controlled by the PI3K-AKT-AP-1 signaling
pathway and promoter methylation. Neuro-oncology
Recombinant anti-podoplanin (NZ-1) immunotoxin for
the treatment of malignant brain tumors. Int. J. Cancer
47. Jovčevska I. Genetic secrets of long-term glioblastoma
mitochondrial calcium uniporter is a multimer that can
include a dominant-negative pore-forming subunit.

Source of Support: Nil; Conflict of Interest: None
Submitted: 03-09-2020; Accepted: 01-10-2020; Published: 31-10-2020