

## ORIGINAL ARTICLE

# Identification of the potential biomarkers in patients with glioma: a weighted gene co-expression network analysis

Ting-Yu Chen<sup>1,†</sup>, Yang Liu<sup>1,†</sup>, Liang Chen<sup>2</sup>, Jie Luo<sup>1,2</sup>, Chao Zhang<sup>1,\*</sup> and Xian-Feng Shen<sup>1,3</sup>

<sup>1</sup>Center for Evidence-Based Medicine and Clinical Research and <sup>2</sup> Department of Neurosurgery and <sup>3</sup> Department of General Surgery, Taihe Hospital, Hubei University of Medicine, No.32, South Renmin Road, Shiyan 442000, China

\*To whom correspondence should be addressed. Tel: +86 0719-8801071; Fax: +86 0719-8801071; E-mail: [zhangchao0803@126.com](mailto:zhangchao0803@126.com)

Correspondence may also be addressed to Xian-Feng Shen. Tel: +86 0719-8801497; Fax: +86 0719-8801497; E-mail: [shenxianfeng2008@163.com](mailto:shenxianfeng2008@163.com)

<sup>†</sup>These authors contributed equally to this work.

## Abstract

Glioma is the most common brain tumor with high mortality. However, there are still challenges for the timely and accurate diagnosis and effective treatment of the tumor. One hundred and twenty-one samples with grades II, III and IV from the Gene Expression Omnibus database were used to construct gene co-expression networks to identify hub modules closely related to glioma grade, and performed pathway enrichment analysis on genes from significant modules. In gene co-expression network constructed by 2345 differentially expressed genes from 121 gene expression profiles for glioma, we identified the black and blue modules that associated with grading. The module preservation analysis based on 118 samples indicates that the two modules were replicable. Enrichment analysis showed that the extracellular matrix genes were enriched for blue module, while cell division genes were enriched for black module. According to survival analysis, 21 hub genes were significantly up-regulated and one gene was significantly down-regulated. What's more, *IKBIP*, *SEC24D*, and *FAM46A* are the genes with little attention among the 22 hub genes. In this study, *IKBIP*, *SEC24D*, and *FAM46A* related to glioma were mentioned for the first time to the current knowledge, which might provide a new idea for us to study the disease in the future. *IKBIP*, *SEC24D* and *FAM46A* among the 22 hub genes identified that are related to the malignancy degree of glioma might be used as new biomarkers to improve the diagnosis, treatment and prognosis of glioma.

## Introduction

Glioma is the most common tumor in the nervous system, and it is also the most harmful (1). It accounts for 26.5% of all central nervous system tumors and 80.7% of malignant tumors, with seven out of every 100,000 people suffer from this disease in the United States each year (2). According to the histological classification, it can be divided into diffuse and non-diffuse. Most gliomas are diffuse and extensive infiltration in the central nervous system. Diffuse gliomas tend to be high-grade malignant, including grade III (Anaplastic astrocytoma), grade IV (Glioblastoma), etc (1,3). High-grade gliomas with a median overall survival of about 16–18 months can occur at any age.

More importantly, there are challenges in distinguishing Grades II, III, and IV (4), and accurate grading is very important for the decision of patients' treatment plan, which is related to the prognosis of patients (5). Although gliomas have made progress in diagnosis and treatment, but the survival rates of them in general is not optimistic, thus, the painstaking search for new treatments continues. Currently, the research on glioma focuses on biological molecules, such as DNA, RNA and protein, and so on. Glial cell line-derived neurotrophic factor is a highly conserved neurotrophic factor, and its high expression is of great significance for the development and formation of glioma. Many

Received: 12 August 2019; Revised: 13 October 2019; Accepted: 22 November 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Abbreviations

ANOVA	analysis of variance;
CLIC1	Chloride intracellular channel 1;
DAVID	Database for Annotation, Visualization, and Integrated Discovery;
DEGs	differentially expressed genes;
GEO	Gene Expression Omnibus;
GO	gene ontology;
GS	gene significance;
IDH	Isocitrate dehydrogenase;
KEGG	Kyoto encyclopedia of genes and genomes;
ME	module eigengene;
MM	module membership;
MS	module significance;
PPI	protein-protein interaction;
RMA	multi-array averaging;
TOM	topological overlap matrix;
WGCNA	weighted gene co-expression analysis

studies have been conducted on Isocitrate dehydrogenase (IDH) (6) and the long noncoding RNAs (lncRNAs): IDH1 mutations are common in glioma, and lncRNAs (AC064875.2, HOTAIRM1, LINC00908, RP11-84A19.3, and LINC00319) has been reported to be closely related to the prognosis of glioma (7,8). In addition, miRNA have also been shown to be associated with glioma progression (9,10). However, the poor prognosis suggests that it is necessary to search for new biomarkers to evaluate the malignant degree and long-term prognosis of glioma.

Research methods based on biomolecular network are widely used, especially those based on gene co-expression network and protein-protein interaction (PPI) network are very important. In 2005, the Zhang and Horvath first systematically proposed weighted gene co-expression analysis (WGCNA) (11), which has since been used as a powerful data-driven tool in the study of many diseases. Gene co-expression networks and PPI networks were included in the WGCNA. The WGCNA method identified genes closely related to disease by constructing gene co-expression networks, and the PPI network provided scoring of candidate biomarkers from the protein level (12). Achievements in identifying potential biomarkers by constructing gene co-expression networks encourage researchers to study the direct possible relationship between modules and disease. The identification of modules related to grading is helpful to infer the mechanism of tumor, predict prognosis and establish new therapeutic ideas.

The purpose of this study is based on four separate microarray data set to build gene expression in the network to identify biomarkers are closely associated with glioma grading. The biological markers might provide new insights for the prognosis and treatment for patients, and the co-expression network constructed at the same time might provide novel ideas for the occurrence and progression mechanism of glioma.

## Materials and methods

### Data collection and data preprocessing

Four raw public microarray datasets were downloaded from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Data sets GSE43378 (13), GSE51395 (14), GSE4290 (15), and GSE43289 (16) all come from the same platform: GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array. Four data sets were divided into two, 121 samples with grades II, III, and IV were used to construct co-expression networks to identify biomarkers closely related to glioma grading, and 118

samples with grades II, III, and IV were used for module preservation analysis and gene validation.

### Gene co-expression network construction

A WGCNA R software package summarized and standardized its methods including network construction, module detection, gene identification, topological property calculation, and visualization (17). After data normalization of 121 glioma gene expression profiles by using robust multi-array averaging (RMA) (18) algorithm, outliers was excluded by filtering the samples with reference to Z.K value (Z.K value  $< -2.5$ ). Moreover, the differentially expressed genes (DEGs) identified by “limma” (19) were used to construct co-expression networks, and  $|\log_2(\text{fold change})| \geq 0.5$  and  $P < 0.01$  were the cut-off thresholds for DEGs. The specific steps are as follows: the correlation matrix ( $S_{ij}$ ) was transformed into an adjacency matrix ( $A_{ij}$ ) based on soft threshold  $\beta$ :  $A_{ij} = |S_{ij}|^\beta$ . The soft threshold  $\beta$  is also known as the weighting coefficient, which is selected according to the scale-free topological criterion, and it can make the network meet the scale-free approximately ( $R^2 > 0.8$ ), thus leading to the network with higher biological signals. In order to visualize the network, the adjacency matrix was transformed into a topological overlap matrix (TOM) that allows visualization and identification of network modules. Dynamic tree cutting was used to identify modules in dendrogram (20,21), which is based on average linkage hierarchical clustering coupled with the TOM-based dissimilarity  $d_{w_{ij}}$ . Module is also known as sets of highly co-expressed genes, and the number of members in module is usually more than 30. Normally not all genes used in the study are co-expressed with other genes, all uncharacterized genes were assigned to the gray module according to the authors (11).

### Module preservation

To verify the existence of the module, we need to demonstrate that the module is replicable or preserved in gene co-expression network based on other independent microarray data sets. We use “modulePreservation” to calculate the preservation statistics of modules and determine whether the modules are preserved according to Z summary score and medianRank. Z summary score usually depends on the size of the module, while the medianRank has a relatively small dependence on the module (22), so whether a module is preserved or not depends on the Z summary score and the medianRank. It should be made clear that the modules with a higher Z summary score are more firmly preserved than those with a lower Z summary score, while the modules with a lower medianRank have stronger module preservation statistics than those with a higher medianRank.

### Identification of modules associated with different grade of glioma and functional annotations

The selection of hub modules requires reference to the Module-trait relationships diagram. Module eigengene (ME) is the first principal component of a module and the representative of gene expression profile in a module (17). Therefore, we analyzed the relationship between MEs and clinical traits to select interest modules for subsequent analysis. Furthermore, the three important parameters, including gene significance (GS), module membership (MM), and module significance (MS) were used for identifying hub modules. GS was defined as transform of the P-value from the linear regression between gene expression and clinical stage, and average GS in a module was defined as MS. MM was defined as the correlation between the expression of genes and the ME.

Moreover, genes in interest module were uploaded to the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (23) for gene ontology (GO) functional annotation and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis for funding the underlying mechanism and biological pathways.

### Hub genes identification, validation and Kaplan-Meier survival analysis

The hub genes are those with high connectivity in the module. The first condition for the hub genes in this study is that the connectivity ranking is in the top 10 percent among genes in the module, and second cut-off criteria is the absolute value of `geneModuleMembership`  $> 0.8$  and

geneTraitSignificance > 0.4. Furthermore, PPI network was constructed based on Search Tool for the retrieval of interacting genes (24) and visualized by Cytoscape (25). Genes in hub modules were projected into a PPI to further clarify the interaction between genes, which was one of the evidence supporting the status of hub genes. One hundred and eighteen samples was used for gene verification based on one-way analysis of variance (one-way ANOVA), which could make the selection of hub genes more convincing. If the  $P$ -value < 0.01, the selection of the gene is considered statistically significant and it is considered to be validated.

Kaplan–Meier survival analysis is a common non-parametric survival analysis method, which can intuitively show the relationship between one or more gene expression profiles and survival time by drawing survival curve (26). Glioma patients in GSE43378 were divided into high expression group and low expression group according to median expression value of each hub genes. After relevant calculation, survival curves were drawn for subsequent analysis. The survival curve is a step function, and the steeper the curve, the lower the survival rate. As a common way to compare survival curves, log-rank test can draw a conclusion that there is statistical significance between groups or not by analyzing the significance of differences between actual and theoretical values. A log-rank  $P$ -value < 0.01 was considered to indicate a statistically significant difference. We conducted the Kaplan–Meier survival analysis and log-rank test using the “survival” package of R software to assess the relationship between hub genes and glioma patients.

## Results

### Weighted co-expression network construction

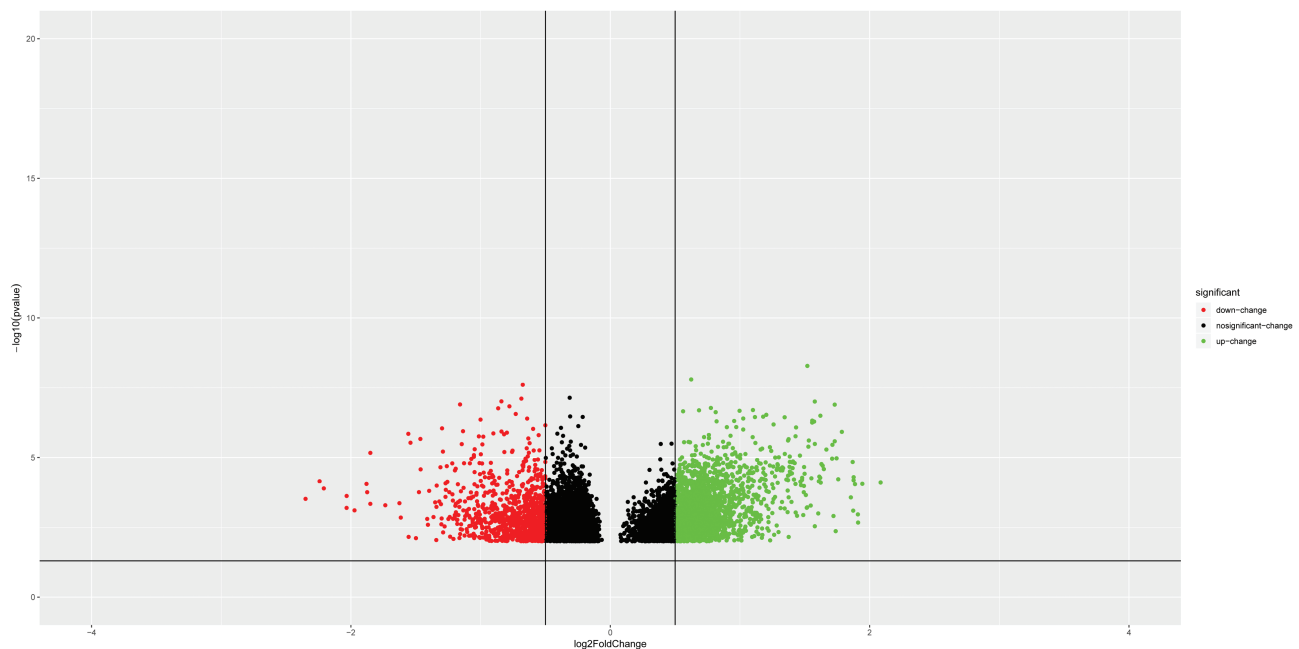
After gene expression profiles of 121 samples were used for RMA algorithm normalization, GSM61679 samples with Z.K value < -2.5 were excluded as outliers and did not participate in subsequent analysis (Supplemental Figure 1). A total of 2345 DEGs with  $|\log_2(\text{fold change})| \geq 0.5$  and  $P < 0.01$  including 1644 up-regulation and 701 down-regulation were selected (Figure 1). The correlation matrix composed of 2345 DEGs is transformed into an adjacency matrix based on soft threshold  $\beta = 10$  (scale free  $R^2 = 0.82$ ) (Supplemental Figure 2). We identified 10 gene modules based on average hierarchical clustering coupled with dissimilarity  $d\omega_{ij}$ .

### Module preservation

Since the gray module is a non-functional module that does not participate in subsequent analysis, nine gene modules participate in the module preservation analysis. For Z summary score, Z summary score < 5 is considered not to be preserved, and the value between 5 and 10 is considered to be moderately preserved, while the Z summary score > 10 module is considered to be highly preserved, and the higher the Z summary score is, the more stable the module is preserved. We could maintain that 7/9 modules were highly preserved: turquoise (34.2), blue (16.4) yellow (15.8), red (13.9), pink (11.2), black (13.3), and brown module (10.2). For the medianRank, the lower the medianRank is, the higher the score of module preservation is. As shown in Supplemental Figure 3, both the black module (Z summary score = 13.3; medianRank = 1) and the blue module (Z summary score = 16.4; medianRank = 5) showed good results.

### Identification of modules associated with different grade of glioma and functional annotations

The ME of each gene module represents the gene expression of the entire module, so the correlation between each module and clinical characteristics such as type and grade is measured by the correlation between each ME and clinical characteristics such as type and grade. We note that black ( $r = 0.51$ ,  $P = 2e-09$ ), blue ( $r = 0.62$ ,  $P = 4e-14$ ), and magenta ( $r = 0.64$ ,  $P = 4e-15$ ) all have high correlation with glioma clinical grade (Figure 2). However, magenta and brown modules were excluded from the subsequent analysis for their performance was not outstanding in the module preservation analysis. Incidentally, the correlation between the turquoise module and clinical characteristics was not significant, although the turquoise modules were well preserved in the module preservation analysis (Z summary score = 34.2; medianRank = 4). Additionally, the black module ( $\text{cor} = 0.57$ ,  $P = 1.2e-08$ ) and the blue module ( $\text{cor} = 0.76$ ,  $P = 5.2e-50$ ) showed high genetic significance and module membership (Supplemental Figure 4). In short, black and blue module



**Figure 1.** Volcano figure. Note: DEGs were defined as the genes with  $|\log_2(\text{fold change})| \geq 0.5$  and  $P < 0.01$ . Green dots represent up-regulated genes, and red dot represent down-regulated genes.

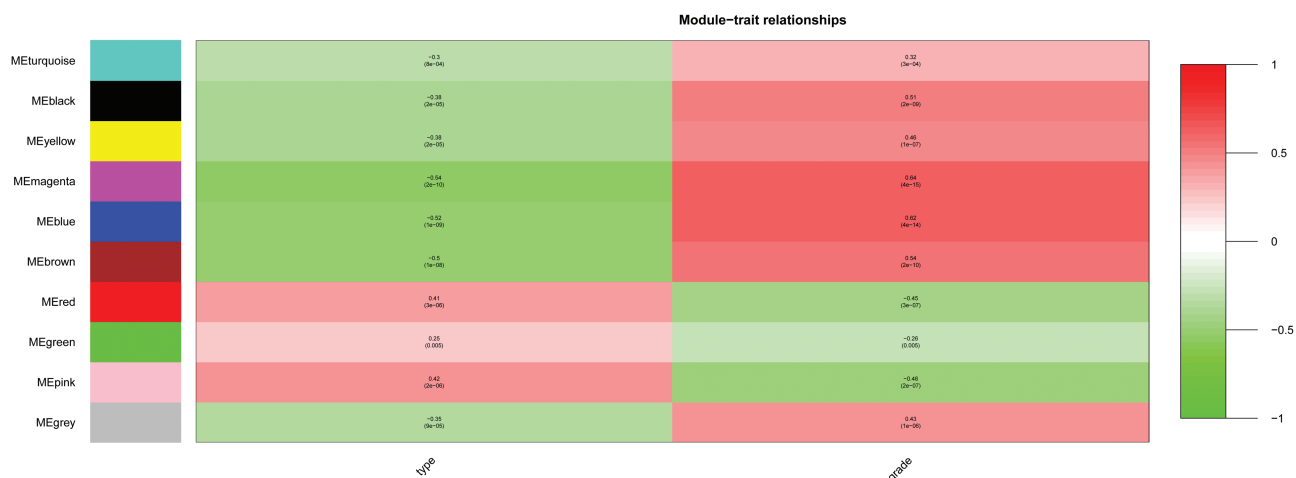


Figure 2. Heatmap of the correlation between ME and clinical traits of glioma.

Table 1. Seven hub genes in black module related with pathological grade glioma

Gene symbol	Gene Module Membership	Gene Trait Significance	PPI connectivity degree	Up/down
BUB1	0.935022695	0.47172381	55	Up
CCNB2	0.949060871	0.448905865	55	Up
KIF20A	0.942235258	0.496177154	52	Up
NUSAP1	0.946448433	0.492814704	51	Up
PTTG1	0.953153568	0.505052177	50	Up
RRM2	0.935055347	0.512808509	54	Up
TOP2A	0.95505808	0.505171661	54	Up

as grade-associated gene modules were identified in glioma based on WGCNA.

To explore the biological functions of the black and blue modules, the genes in the modules were uploaded to DAVID for GO functional annotation and KEGG enrichment analysis. GO and KEGG indicated that black module was mainly enriched to cell division, mitotic nuclear division and nucleoplasm (Supplemental Figure 5), while blue module was mainly enriched to extracellular matrix organization and extracellular matrix, ECM-receptor interaction (Supplemental Figure 6).

### Hub genes identification, validation and Kaplan-Meier survival analysis

There were 85 genes in the black module and 259 genes in the blue module, and 34 genes were screened due to the connectivity ranking of the top 10 percent and those gene with geneModuleMembership > 0.8 and geneTraitSignificance > 0.4. According to the cut-off threshold of gene survival analysis, seven genes in the black module were identified as hub genes (Table 1). The PPI network of black module showed significant connectivity, and the overall effect of the network graph was good (Supplemental Figure 7). For the selection of hub genes in the blue module, PPI network (Supplemental Figure 8) and survival analysis are important screening references in addition to the restriction on the ranking of connectedness. Overall, 15 hub genes were identified in the blue module (Table 2). Finally, 22 hub genes related to clinical grade of glioma were identified in this study, and these hub genes showed significant statistical significance in survival analysis. Figure 3 intuitively shows the verification results, the *P* values of 22 hub genes are significant ( $P < 0.01$ ), which indicates that the selection of our 22 hub genes is statistically significant, scientific and convincing.

As shown in Figure 4, the survival curve generally shows a downward trend with the increase of time, and the slope is larger, which means lower survival rate. The image of the high-expression group of 21 genes was steeper than that of the low-expression group, indicating that the high expression of these hub genes was closely related to the poor prognosis of patients. However, we found that the low expression group based on KCNB1 showed a stronger association with poor prognosis, and this gene attracted our attention for further analysis and discussion.

### Discussion

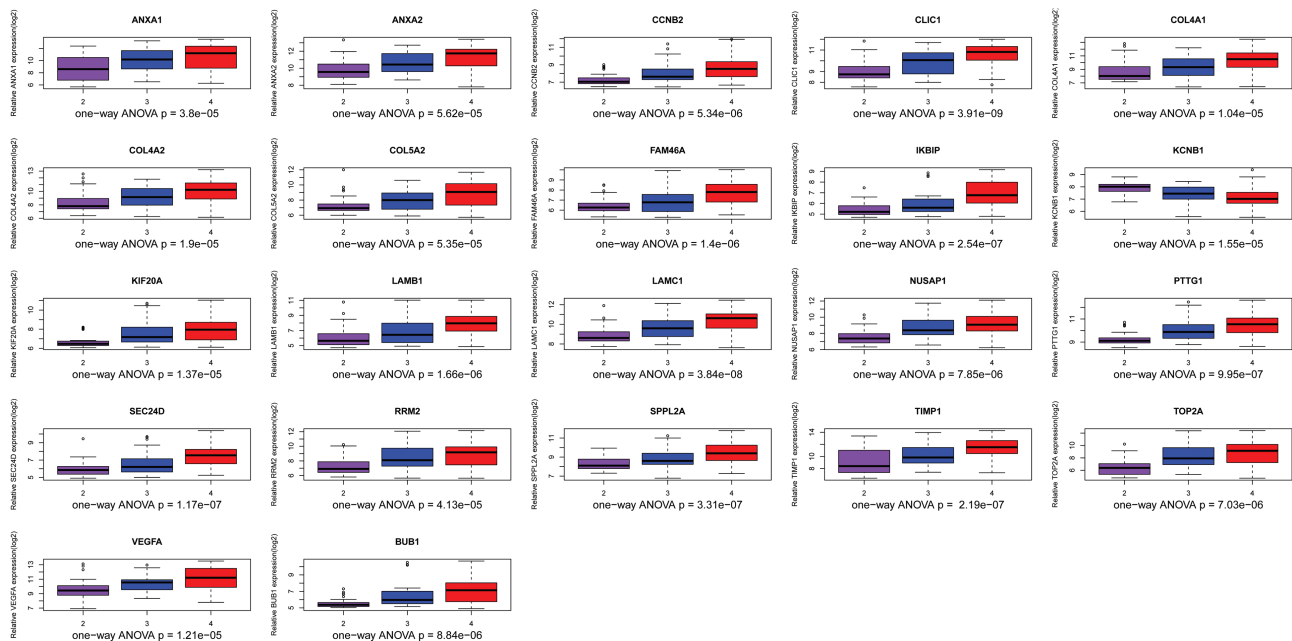
Glioma is a fatal brain tumor, which is the fourth leading cause of tumor-related death due to its invasive nature, molecular signals and the location of the central nervous system (27). In addition to causing death, seizures are common symptoms that seriously affect the lives of patients and their families (28). The researchers' exploration of the mechanism of progression, diagnosis and treatment, and prognosis prediction are positively correlated with its harm to humans. Although important changes have been found in the glioma genome, there is still a big gap in understanding the mechanism of the disease based on the current survival status of patients. Although some studies have used the WGCNA method to construct a co-expression network to identify potential biomarkers associated with glioma, our study might contribute to the establishment of a more complete set of molecular markers for the pathological grade of glioma.

In the study, 121 samples without stage I were used to construct the co-expression network. After dynamic tree cutting, we identified 10 modules, among which the black module and blue module showed the strongest correlation with pathological

**Table 2.** Fifteen hub genes in blue module related with pathological grade glioma

Gene symbol	Gene Module Membership	Gene Trait Significance	PPI Connectivity Degree	Up/down
ANXA1	0.827343404	0.561923969	/	Up
ANXA2	0.875803843	0.548157845	/	Up
CLIC1	0.888418433	0.566527724	/	Up
COL4A1	0.866139195	0.54930032	25	Up
COL4A2	0.804525568	0.52982224	17	Up
COL5A2	0.852490117	0.623360253	15	Up
FAM46A	0.875552138	0.546436775	/	Up
IKBIP	0.887210744	0.516772095	/	Up
KCNB1	0.886305109	0.572156038	/	Down
LAMB1	0.895113867	0.57687226	19	Up
LAMC1	0.925793134	0.618090561	17	Up
SEC24D	0.888379385	0.567315315	/	Up
SPPL2A	0.873493419	0.536895048	/	Up
TIMP1	0.852549775	0.547783447	27	Up
VEGFA	0.82564312	0.559968118	43	Up

Note: “/” means that the connection degree of this gene in PPI network is < 15 or the gene is not projected into PPI network.



**Figure 3.** Hub gene validation was based on one-way ANOVA. Note: Grades II, III and IV (x-axis) and relative expression (log2) for each gene (y-axis) were displayed, and the P-values of 22 genes were calculated and showed.

grade. At the same time, the Z summary score and medianRank indicate that the two modules have good stability based on 118 samples. Moreover, in order to better illustrate how genes work, we also carried out GO functional annotation and KEGG enrichment analysis for black and blue module. Extracellular matrix organization genes were enriched for blue module, while cell division genes were enriched for black module. A total of 22 hub genes were finally identified in two hub modules based on the setting of cut-off thresholds for GS, MM, and PPI network connectivity. The PPI chart of 22 hub genes showed the ideal degree of connectivity, and we could intuitively see the hub genes interacting with each other. Both survival analysis and gene validation yielded satisfactory results.

Glioma is a brain tumor with a high degree of infiltration of surrounding tissue, and its ability to infiltrate normal brain tissue is a marker of poor prognosis (29). Because the brain lacks the special structure of the true lymphatic system, gliomas

migrate along blood vessels and invade surrounding normal brain tissues (30). These physiological activities are based on the tumor cells expressing different programming genes, which affect the cell cycle, modify the surrounding environment, and change the extracellular matrix (31). Changes in extracellular matrix are common in cancer, and there are numerous reports that its changes may promote cancer invasion (32). Hub genes COL4A2, COL4A1, COL5A2, LAMC1, and LAMB1 in blue module were enriched into extracellular matrix. COL4A1/2 encodes type IV collagen, which is an important part of basement membrane. If mutated, it can damage the secretion of extracellular matrix collagen and thus lead to weak blood vessels (33). Therefore, it is not difficult to understand that the mutation of these two genes might help tumor invasion. COL5A2 codes for type V collagen, which is mutated in most patients with classical Ehlers-Danlos syndrome (34). COL5A2 is considered as one of the effective biomarkers for the diagnosis and prognosis of bladder cancer (35),

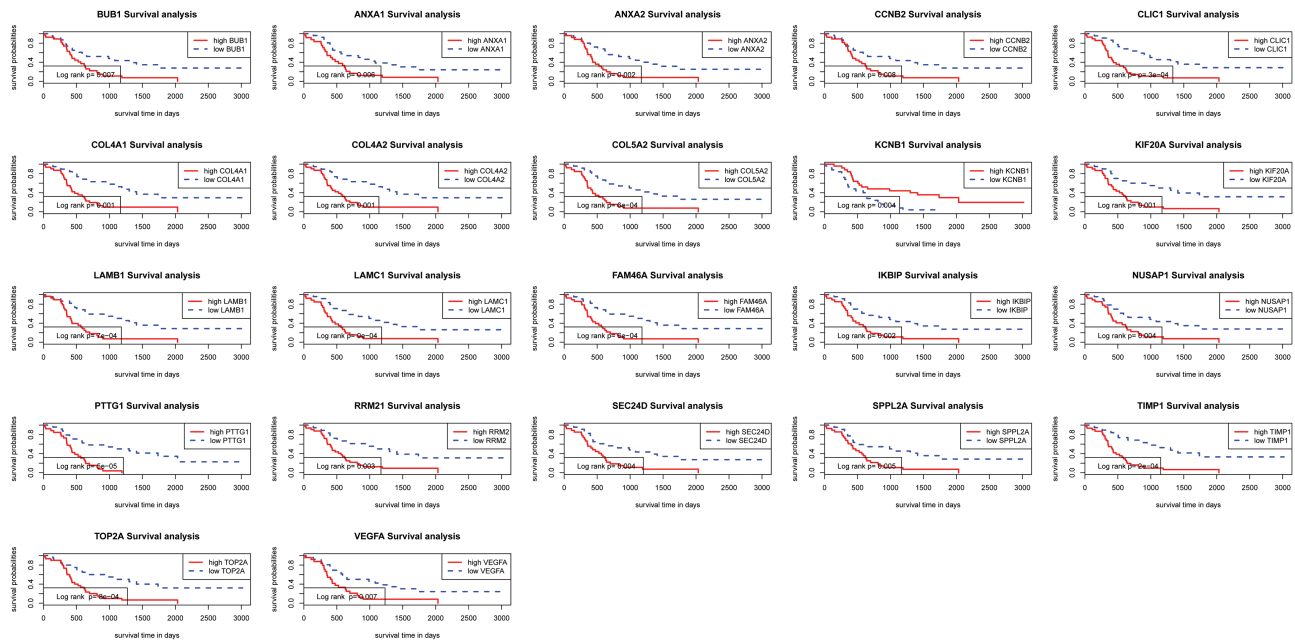


Figure 4. Survival curves for patients in different groups. Note: Red lines represent high expression of hub genes, while blue lines represent low expression of hub genes.

thyroid cancer (36), and gastric cancer (37). In addition to an important association with cancer, COL5A2 is also associated with Alzheimer's disease (38), a disease of the central nervous system. COL4A2, COL4A1, and COL5A2 have been shown to be strongly associated with glioma in a study (39). LAMC1 and LAMB1 are family members of extracellular matrix glycoproteins, and are also the main component of basement membrane, which is involved in cell adhesion differentiation, migration, signal transduction, and so on. The up-regulated expression of LAMB1 in glioma was related to the miR-124-5p down-regulation table based on quantitative PCR and western blot for samples (40). LAMB1 has been identified as a potential biomarker for liver (41), colon (42), and stomach cancers (43). The up-regulated expression of LAMC1 might lead to the down-regulation of miRNA-181a (44) and miR-124a (45), which may be associated with tumor invasion and lead to poor prognosis. Although these genes are closely related to glioma both individually and through enrichment pathways, we do not know exactly how they play a role in tumor progression and infiltration. The gene co-expression network and PPI network constructed in our study may provide clues for studying the complex regulation between them. The up-regulated expression of ANXA1 and ANXA2 in glioma is considered to be the therapeutic target and prognostic marker of the cancer (46). ANXA1 is a direct transcription target of FOXM1 may be involved in cell proliferation, differentiation and apoptosis (47). Chloride intracellular channel 1 (CLIC1) has obvious overexpression in glioma, which is negatively correlated with patient survival, and is a promising potential therapeutic target for the tumor (48). CLIC1 might be related to the mechanism that the Metformin used to treat diabetes inhibit tumor proliferation (49). KCNB1 is a member of voltage-gated potassium channel, which cannot only regulate excitatory cell signals but also play a role in the occurrence of brain tumors (50). The moderate oxidation of this channel is thought to affect neurons by affecting synapses based on mouse experiments (51). In vitro studies have shown that KCNB1 can induce autophagy and increase apoptosis while inhibiting tumor growth, which is related

to the prognosis of glioma (52). This is consistent with the KCNB1 down-regulation obtained in this study. The mechanism of formation and development of glioma by KCNB1 has not been clearly studied, and the co-expression network we constructed may provide some new insights for researchers.

It is worth mentioning that three hub genes *IKBIP*, *SEC24D*, and *FAM46A* in the blue module are special. *IKBIP* (I kappa B kinase interacting protein), also known as *IKIP*, is on human chromosome 12. Researchers have paid little attention to this gene. Currently, we know that this gene is a gene that promotes the function of apoptosis (53). *SEC24D* encoded proteins belong to the *SEC24* family member and *SEC24D* is an important component of coat protein complex type II that mediated transportation of newly synthesized proteins from the endoplasmic reticulum to the Golgi (54). Mutations in the gene are associated with Cole-Carpenter syndrome (55), but the link between the gene and cancer has been poorly studied. So far, we know that *SEC24D* may be one of the candidate genes for the treatment of T-cell lymphoma (56), and this gene may activate the transformation of hepatic stellate cells into myofibroblasts to contribute to cirrhosis (57). *FAM46A* proteins belong to the nucleotidyl transferase fold superfamily. *FAM46A* plays an important role in bone development and has been mentioned in studies of bone-related diseases. A variable number of tandem repeat polymorphism in *FAM46A* gene is associated with osteoarthritis (58) and tuberculosis (59). In addition, there is a link between *FAM46A* and 5-fluorouracil in the treatment of breast cancer (60). Simply to say, the understanding of certain genes is still incomplete, and our co-expression networks might provide new clues to the complex regulation of these different molecules for researchers.

The hub genes in the black module received more attention from researchers than those in the blue module. Seven hub genes in the black module have been further studied. Hub genes *CCNB2*, *BUB1*, and *PTTG1* were enriched into the cell cycle. *BUB1*, a mitotic checkpoint serine kinase, had been reported in many cancers, including glioma, and it is associated

with tumor proliferation, growth, metastasis, and prognosis (61). Cancer genes such as *TOP2A*, *KIF20A*, and *PTTG1* were also identified in this study. Nucleolar and spindle-associated protein 1 is an important protein involved in mitosis, and its abnormal expression is associated with different types of tumors, which has been detected to be overexpressed in glioma. In a study that also constructed a co-expression network to identify glioma-related genes, this gene was suggested to be a potential gene therapy target for the neoplasm (20). Ribonucleotide Reductase Regulatory Subunit M2 (*RRM2*) expression was significantly up-regulated based on TCGA data, western blot and immunohistochemistry, which could promote tumor migration and proliferation (62). Studies have shown that a *RRM2* inhibitor triapine can treat some gliomas (63). Recent studies have provided some new insights into the regulation mechanism of *RRM2* in glioma. *RRM2* may be a potential therapeutic target for cancer, and its inhibitor DHS (*trans*-4,4'-dihydroxystilbene) has been shown to be effective in mouse models of tumor xenotransplantation (64).

*IKBIP*, *SEC24D*, and *FAM46A* were selected from 22 hub genes that related to glioma grading based on a series of related gene analysis. These three genes have received less attention than other genes and to our knowledge these three genes are the first to be shown to be involved in the mechanism of glioma progression. In addition, due to the small number of samples included in this study, there might be bias, and more studies are needed to confirm our results. As far as we know, the samples of published studies are mainly from TCGA (65–67), while our samples are from GEO and these sample sets are rarely or not used. Moreover, our study focused on grades II, III, and IV. More importantly, our study identified new genes and our study provides new insights into the biomarkers of glioma, which may provide new ideas for the diagnosis, treatment and prognosis of patients.

## Conclusions

To sum up, we established the co-expression network to identify the 22 hub genes related to the pathological grade of glioma, the 22 genes are closely related to survival time based on survival analysis. As potential biomarkers, *IKBIP*, *SEC24D*, and *FAM46A* might provide new ideas for more timely and accurate diagnosis, more effective treatment and better prognosis prediction of glioma patients. We need more research to validate our study.

## Supplementary material

Supplementary data are available at *Carcinogenesis* online.

## Funding

None.

## Acknowledgements

None.

*Conflicts of Interest Statement*: The authors declare no conflict of interest.

## Authors' Contributions

CZ, XFS and JL conceived and designed this study. TYC and YL carried out the analysis procedure, CZ and YL analyzed the results, JL and LC contributed analysis tools, TYC, XFS and YL

participated in the manuscript writing. All the authors reviewed the manuscript.

Ethical Approval: Not required.

Consent for Publication: All the authors agreed to publish.

## References

- Wesseling, P. et al. (2018) WHO 2016 Classification of gliomas. *Neuropathol. Appl. Neurobiol.*, 44, 139–150.
- Ostrom, Q.T. et al. (2017) CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro. Oncol.*, 19(suppl\_5), v1–v88.
- Louis, D.N. et al. (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.*, 114, 97–109.
- Sengupta, A. et al. (2019) Glioma grading using a machine-learning framework based on optimized features obtained from T1 perfusion MRI and volumes of tumor components. *J. Magn. Reson. Imaging.*, 50, 1295–1306.
- Wang, X. et al. (2018) Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front. Neurosci.*, 12, 1046.
- Yan, H. et al. (2009) IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.*, 360, 765–773.
- Li, Q. et al. (2019) LncRNA LINC00319 is associated with tumorigenesis and poor prognosis in glioma. *Eur. J. Pharmacol.*, 861, 172556.
- Song, L. et al. (2019) Genome-wide identification of lncRNAs as novel prognosis biomarkers of glioma. *J. Cell. Biochem.*, 120, 19518–19528.
- Chen, Y. et al. (2019) Long non-coding RNA HCG11 modulates glioma progression through cooperating with miR-496/CPEB3 axis. *Cell Prolif.*, 52, e12615.
- Chen, W.Y. et al. (2019) miR143 acts as a novel Big mitogenactivated protein kinase 1 suppressor and may inhibit invasion of glioma. *Oncol. Rep.*, 42, 1194–1204.
- Zhang, B. et al. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, 4, Article17.
- Jeong, H. et al. (2001) Lethality and centrality in protein networks. *Nature*, 411, 41–42.
- Kawaguchi, A. et al. (2013) Gene expression signature-based prognostic risk score in patients with glioblastoma. *Cancer Sci.*, 104, 1205–1210.
- Chong, Y.K. et al. (2016) ST3GAL1-associated transcriptomic program in glioblastoma tumor growth, invasion, and prognosis. *J. Natl. Cancer Inst.*, 108, pii: djv326.
- Sun, L. et al. (2006) Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell*, 9, 287–300.
- Vital, A.L. et al. (2010) Gene expression profiles of human glioblastomas are associated with both tumor cytogenetics and histopathology. *Neuro. Oncol.*, 12, 991–1003.
- Langfelder, P. et al. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.
- Irizarry, R.A. et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4, 249–264.
- Ritchie, M.E. et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43, e47.
- Yang, Q. et al. (2018) Candidate biomarkers and molecular mechanism investigation for glioblastoma multiforme utilizing WGCNA. *Biomed. Res. Int.*, 2018, 4246703.
- Langfelder, P. et al. (2008) Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics*, 24, 719–720.
- Langfelder, P. et al. (2011) Is my network module preserved and reproducible? *PLoS Comput. Biol.*, 7, e1001057.
- Dennis, G. Jr et al. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, 4, P3.
- Szklarczyk, D. et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, 43(Database issue), D447–D452.
- Kohl, M. et al. (2011) Cytoscape: software for visualization and analysis of biological networks. *Methods Mol. Biol.*, 696, 291–303.

26. Goel, M.K. et al. (2010) Understanding survival analysis: Kaplan-Meier estimate. *Int. J. Ayurveda Res.*, 1, 274–278.
27. Kundu, M. et al. (2019) Prospect of natural products in glioma: a novel avenue in glioma management. *Phytother. Res.*, 33, 2571–2584.
28. Samudra, N. et al. (2019) Seizures in glioma patients: an overview of incidence, etiology, and therapies. *J. Neurol. Sci.*, 404, 80–85.
29. Mitchell, C.B. et al. (2019) Tropomyosin Tpm 2.1 loss induces glioblastoma spreading in soft brain-like environments. *J. Neurooncol.*, 141, 303–313.
30. Schiera, G. et al. (2017) Molecular determinants of malignant brain cancers: from intracellular alterations to invasion mediated by extracellular vesicles. *Int. J. Mol. Sci.*, 18, pii: E2774.
31. Tejero, R. et al. (2019) Gene signatures of quiescent glioblastoma cells reveal mesenchymal shift and interactions with niche microenvironment. *EBioMedicine*, 42, 252–269.
32. Kai, F. et al. (2019) The extracellular matrix modulates the metastatic journey. *Dev. Cell*, 49, 332–346.
33. Vitale, G. et al. (2019) Cortical malformations and COL4A1 mutation: three new cases. *Eur. J. Paediatr. Neurol.*, 23, 410–417.
34. Duong, J. et al. (2019) A family with Classical Ehlers-Danlos Syndrome (cEDS), mild bone fragility and without vascular complications, caused by the p.Arg312Cys mutation in COL1A1. *Eur. J. Med. Genet.*, 14, 103730.
35. Zhu, H. et al. (2019) Collagen stiffness promoted non-muscle-invasive bladder cancer progression to muscle-invasive bladder cancer. *Oncol. Targets. Ther.*, 12, 3441–3457.
36. Tang, X. et al. (2019) Identifying gene modules of thyroid cancer associated with pathological stage by weighted gene co-expression network analysis. *Gene*, 704, 142–148.
37. Hao, S. et al. (2019) Identification of key genes and circular RNAs in human gastric cancer. *Med. Sci. Monit.*, 25, 2488–2504.
38. Pang, C. et al. (2019) Identification and analysis of Alzheimer's candidate genes by an amplitude deviation algorithm. *J. Alzheimers Dis. Parkinsonism*, 9, pii: 460.
39. Vastrad, B. et al. (2017) Molecular mechanisms underlying gliomas and glioblastoma pathogenesis revealed by bioinformatics analysis of microarray data. *Med. Oncol.*, 34, 182.
40. Chen, Q. et al. (2014) MiR-124-5p inhibits the growth of high-grade gliomas through posttranscriptional regulation of LAMB1. *Neuro. Oncol.*, 16, 637–651.
41. Zhang, L. et al. (2018) Is integrin subunit alpha 2 expression a prognostic factor for liver carcinoma? A validation experiment based on bioinformatics analysis. *Pathol. Oncol. Res.*, 25, 1545–1552.
42. Lin, Q. et al. (2015) Analysis of colorectal cancer glyco-secretome identifies laminin  $\beta$ -1 (LAMB1) as a potential serological biomarker for colorectal cancer. *Proteomics*, 15, 3905–3920.
43. Choi, M.R. et al. (2015) Laminin gene LAMB4 is somatically mutated and expressionally altered in gastric and colorectal cancers. *APMIS*, 123, 65–71.
44. Huang, S.X. et al. (2016) The correlation of microRNA-181a and target genes with poor prognosis of glioblastoma patients. *Int. J. Oncol.*, 49, 217–224.
45. Fowler, A. et al. (2011) miR-124a is frequently down-regulated in glioblastoma and is involved in migration and invasion. *Eur. J. Cancer*, 47, 953–963.
46. Schittenhelm, J. et al. (2009) Comparative analysis of annexin-1 in neuroepithelial tumors shows altered expression with the grade of malignancy but is not associated with survival. *Mod. Pathol.*, 22, 1600–1611.
47. Cheng, S.X. et al. (2013) FoxM1 promotes glioma cells progression by up-regulating Anxa1 expression. *PLoS One*, 8, e72376.
48. Setti, M. et al. (2013) Functional role of CLIC1 ion channel in glioblastoma-derived stem/progenitor cells. *J. Natl. Cancer Inst.*, 105, 1644–1655.
49. Gritti, M. et al. (2014) Metformin repositioning as antitumoral agent: selective antiproliferative effects in human glioblastoma stem cells, via inhibition of CLIC1-mediated ion current. *Oncotarget*, 5, 11252–11268.
50. Wang, H.Y. et al. (2016) A three ion channel genes-based signature predicts prognosis of primary glioblastoma patients and reveals a chemotherapy sensitive subtype. *Oncotarget*, 7, 74895–74903.
51. Yu, W. et al. (2019) Oxidation of KCNB1 potassium channels in the murine brain during aging is associated with cognitive impairment. *Biochem. Biophys. Res. Commun.*, 512, 665–669.
52. Wang, H.Y. et al. (2017) Role of KCNB1 in the prognosis of gliomas and autophagy modulation. *Sci. Rep.*, 7, 14.
53. Hofer-Warbinek, R. et al. (2004) A highly conserved proapoptotic gene, IKIP, located next to the APAF1 gene locus, is regulated by p53. *Cell Death Differ.*, 11, 1317–1325.
54. Wang, B. et al. (2018) The COPII cargo adapter SEC24C is essential for neuronal homeostasis. *J. Clin. Invest.*, 128, 3319–3332.
55. Takeyari, S. et al. (2018) Japanese patient with cole-carpenter syndrome with compound heterozygous variants of SEC24D. *Am. J. Med. Genet. A*, 176, 2882–2886.
56. Refaat, A. et al. (2018) Retrospective screening of microarray data to identify candidate IFN-inducible genes in a HTLV-1 transformed model. *Oncol. Lett.*, 15, 4753–4758.
57. Tomoishi, S. et al. (2017) CREB3L2-mediated expression of Sec23A/Sec24D is involved in hepatic stellate cell activation through ER-Golgi transport. *Sci. Rep.*, 7, 7992.
58. Etokebe, G.E. et al. (2015) Susceptibility to large-joint osteoarthritis (hip and knee) is associated with BAG6 rs3117582 SNP and the VNTR polymorphism in the second exon of the FAM46A gene on chromosome 6. *J. Orthop. Res.*, 33, 56–62.
59. Etokebe, G.E. et al. (2014) Association of variable number of tandem repeats in the coding region of the FAM46A gene, FAM46A rs11040 SNP and BAG6 rs3117582 SNP with susceptibility to tuberculosis. *PLoS One*, 9, e91385.
60. Tsao, D.A. et al. (2010) Gene expression profiles for predicting the efficacy of the anticancer drug 5-fluorouracil in breast cancer. *DNA Cell Biol.*, 29, 285–293.
61. Yu, H. et al. (2019) Serine/threonine kinase BUB1 promotes proliferation and radio-resistance in glioblastoma. *Pathol. Res. Pract.*, 215, 152508.
62. Sun, H. et al. (2019) RRM2 is a potential prognostic biomarker with functional significance in glioma. *Int. J. Biol. Sci.*, 15, 533–543.
63. Rasmussen, R.D. et al. (2016) BRCA1-regulated RRM2 expression protects glioblastoma cells from endogenous replication stress and promotes tumorigenicity. *Nat. Commun.*, 7, 13398.
64. Chen, C.W. et al. (2019) DHS (trans-4,4'-dihydroxystilbene) suppresses DNA replication and tumor growth by inhibiting RRM2 (ribonucleotide reductase regulatory subunit M2). *Oncogene*, 38, 2364–2379.
65. Wang, Y. et al. (2019) A risk classification system with five-gene for survival prediction of glioblastoma patients. *Front. Neurol.*, 10, 745.
66. Xu, P. et al. (2018) Identification of glioblastoma gene prognosis modules based on weighted gene co-expression network analysis. *BMC Med. Genomics*, 11, 96.
67. Zhao, J. et al. (2019) A 6-gene risk signature predicts survival of glioblastoma multiforme. *Biomed Res. Int.*, 2019, 1649423.