



# Expression Profile Analysis Identifies a Novel Five-Gene Signature to Improve Prognosis Prediction of Glioblastoma

Wen Yin<sup>1</sup>, Guihua Tang<sup>2</sup>, Quanwei Zhou<sup>1</sup>, Yudong Cao<sup>1</sup>, Haixia Li<sup>3</sup>, Xianyong Fu<sup>1</sup>, Zhaoping Wu<sup>1</sup> and Xingjun Jiang<sup>1\*</sup>

<sup>1</sup> Department of Neurosurgery, Xiangya Hospital of Central South University, Changsha, China, <sup>2</sup> Department of Clinical Laboratory, Hunan Provincial People's Hospital (First Affiliated Hospital of Hunan Normal University), Changsha, China, <sup>3</sup> Department of Operative Nursing, Xiangya Hospital of Central South University, Changsha, China

## OPEN ACCESS

### Edited by:

Monica Bianchini,  
University of Siena, Italy

### Reviewed by:

Nitish Kumar Mishra,  
University of Nebraska Medical  
Center, United States  
Sen Peng,  
Translational Genomics Research  
Institute, United States  
Max Shpak,  
St David's Medical Center,  
United States

### \*Correspondence:

Xingjun Jiang  
jiangxj@csu.edu.cn;  
jxyjz@163.com

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Genetics

Received: 20 December 2018

Accepted: 17 April 2019

Published: 03 May 2019

### Citation:

Yin W, Tang G, Zhou Q, Cao Y,  
Li H, Fu X, Wu Z and Jiang X (2019)  
Expression Profile Analysis Identifies  
a Novel Five-Gene Signature  
to Improve Prognosis Prediction  
of Glioblastoma.  
Front. Genet. 10:419.  
doi: 10.3389/fgene.2019.00419

Glioblastoma multiforme (GBM) is the most aggressive primary central nervous system malignant tumor. The median survival of GBM patients is 12–15 months, and the 5 years survival rate is less than 5%. More novel molecular biomarkers are still urgently required to elucidate the mechanisms or improve the prognosis of GBM. This study aimed to explore novel biomarkers for GBM prognosis prediction. The gene expression profiles from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets of GBM were downloaded. A total of 2241 overlapping differentially expressed genes (DEGs) were identified from TCGA and GSE7696 datasets. By univariate COX regression survival analysis, 292 survival-related genes were found among these DEGs ( $p < 0.05$ ). Functional enrichment analysis was performed based on these survival-related genes. A five-gene signature (PTPRN, RGS14, G6PC3, IGF2BP2, and TIMP4) was further selected by multivariable Cox regression analysis and a prognostic model of this five-gene signature was constructed. Based on this risk score system, patients in the high-risk group had significantly poorer survival results than those in the low-risk group. Moreover, with the assistance of GEPIA <http://gepia.cancer-pku.cn/>, all five genes were found to be differentially expressed in GBM tissues compared with normal brain tissues. Furthermore, the co-expression network of the five genes was constructed based on weighted gene co-expression network analysis (WGCNA). Finally, this five-gene signature was further validated in other datasets. In conclusion, our study identified five novel biomarkers that have potential in the prognosis prediction of GBM.

**Keywords:** glioblastoma, differentially expressed genes, gene signature, prognosis, TCGA, GEO

## INTRODUCTION

Glioblastoma multiforme (GBM) is the most common and aggressive primary central nervous system malignant tumor with high morbidity and mortality. According to genomic abnormalities and gene expression, GBM can be divided into four molecular subtypes: classical, mesenchymal, neural, and proneural, which lay a foundation for understanding its inherent heterogeneity (Verhaak et al., 2010; Ma et al., 2018). In the United States, the incidence of GBM is 2.96 cases/100,000 population/year (Jhanwar-Uniyal et al., 2015). Although there are several treatment options, including surgery, radiotherapy and chemotherapy, the median survival of GBM

patients remains 12–15 months, and the 5 years survival rate is less than 5% (Wen and Kesari, 2008; Ostrom et al., 2013).

With the development of next-generation sequencing technologies, many specific molecular signatures have been identified to better understand the molecular pathogenesis of GBM (Aldape et al., 2015). As a result, many potential diagnostic and prognostic biomarkers have been discovered that enable a more specific classification and a more precise outcome prediction of GBM. Some molecular markers including MGMT (O6-methylguanine DNA methyltransferase), IDH (isocitrate dehydrogenase), EGFR (epidermal growth factor receptor), and PTEN (phosphatase and tensin homolog) have been routinely tested in GBM patients clinically (van den Bent et al., 2017; Binabaj et al., 2018). More importantly, these molecular signatures have contributed to personalized therapeutic approaches and targeted anti-GBM therapies (Huang et al., 2017; Szopa et al., 2017). However, considering the poor prognosis of GBM, novel molecular biomarkers and new therapeutic strategies are still urgently required to elucidate the mechanisms of GBM or increase overall patient survival.

Previous studies have shown that gene expression profile analysis could detect gene signatures to predict the outcome for malignancy tumors (Luo et al., 2018; Mao et al., 2018; Zeng et al., 2018). Shergalis et al. (2018) discovered that 20 genes were overexpressed and correlated with poor survival outcomes in GBM patients by bioinformatics analysis using data from The Cancer Genome Atlas (TCGA) project. Bao et al. (2014) identified a nine-gene signature to predict the prognosis of glioma patients based on mRNA expression profiling from the Chinese Glioma Genome Atlas (CGGA) database. Therefore, it is necessary to understand the development and progression of GBM by identifying GBM-related genes and to investigate of their potential clinical roles and molecular mechanisms.

In this study, RNA-Seq data from TCGA and microarray data from the Gene Expression Omnibus (GEO) database of GBM were downloaded. Based on the overlapping differentially expressed genes (DEGs), the genes related to prognosis were screened. By using Cox regression, we developed a five-gene signature based risk score to demonstrate the association between gene expression and the prognosis of GBM. Moreover, we validated this signature in the GEO dataset and TCGA array dataset. These results might be able to provide new reference for the prognostic predication of GBM.

## MATERIALS AND METHODS

### Data Source

The GBM RNA sequencing (RNA-seq) dataset and corresponding clinical follow-up information were downloaded from TCGA database (March, 2018). Subtype data of GBM were downloaded from UCSC Xena<sup>1</sup>. A total of 159 patients, including 154 samples of primary GBM patients and five samples of normal brain tissue were extracted for subsequent analysis.

<sup>1</sup><http://xena.ucsc.edu/>

Gene expression microarray data GSE7696 (Lambiv et al., 2011), including 71 samples of primary GBM patients and four samples of normal brain tissue, were downloaded from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus<sup>2</sup>. The dataset was based on the GPL570 platform of [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, United States).

### Differential Expression Analyses

Then, gene profiles were standard normalized within and among samples, respectively. Because the numerical distribution of RPKM (reads per kilo-base per million mapped reads) is too wide, the final expression level of a gene was defined as the  $\log_2(x + 1)$  of the raw expression level. Next, the DEGs between the tumor and normal samples were calculated by the limma package ( $\text{Padj} < 0.05$  and  $|\log_2\text{FC}| > 1$ ). The Venn diagram was produced by the VennDiagram R package (Chen and Boutros, 2011).

### Identification and Selection of Survival-Related Genes

Only the patients with detailed follow-up times were extracted for subsequent survival analyses. Univariate Cox regression survival analysis using the Survival package in R was performed to identify survival-related genes (Yang et al., 2016). Genes were selected with a *p*-value of less than 0.05.

### Go and KEGG Annotation of Survival-Related Genes

Gene Ontology (GO) enrichment and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis were performed on the survival-related genes (Ogata et al., 1999; Wanggou et al., 2016; Li et al., 2018). DAVID (The Database for Annotation, Visualization, and Integrated Discovery) (Dennis et al., 2003) software and the clusterProfiler package (Yu et al., 2012) in R were used to annotate and visualize GO terms and KEGG pathways.

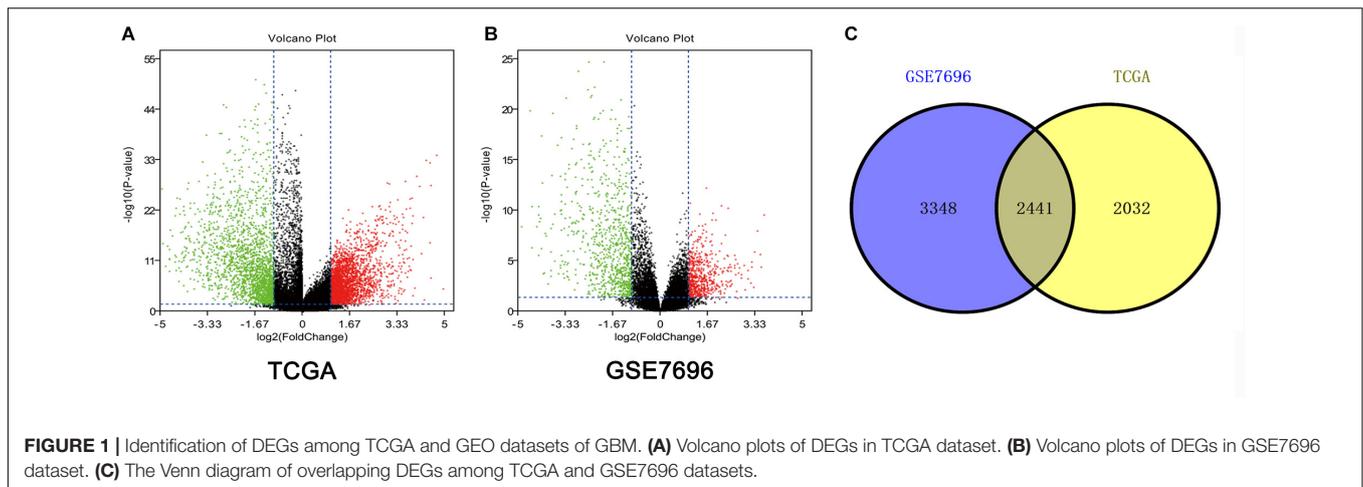
### Gene Signature Identification and Risk Score System Establishment

Based on the top 100 survival-related genes in TCGA dataset, multivariable Cox proportional hazard regression analysis was performed to establish a risk score formula (O'Quigley and Moreau, 1986). As previously reported, a prognosis risk score formula could be constructed on the basis of a linear combination of the expression level (exp) multiplied by a regression coefficient ( $\beta$ ) derived from the multivariate cox regression model.

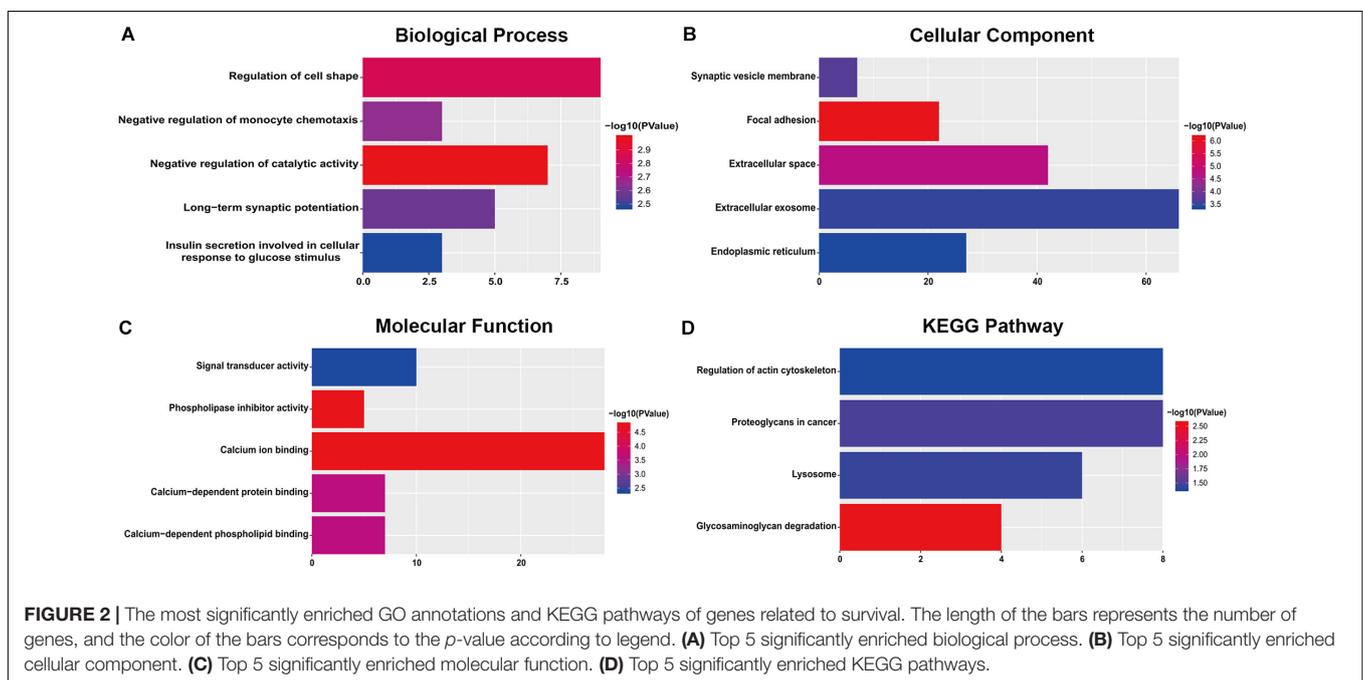
$$\text{Risk Score (RS)} = \text{expPTPRN} * \beta_{\text{PTPRN}} + \text{expRGS14} * \beta_{\text{RGS14}} \\ + \text{expG6PC3} * \beta_{\text{G6PC3}} + \text{expIGFBP2} * \beta_{\text{IGFBP2}} + \text{expTIMP4} * \beta_{\text{TIMP4}}$$

Based on the formula, the risk score of each GBM patient was calculated, and then GBM patients were divided into high-risk score and low-risk score groups. The receiver operating characteristic (ROC) curve analysis was conducted using the R

<sup>2</sup><http://www.ncbi.nlm.nih.gov/geo/>



**FIGURE 1 |** Identification of DEGs among TCGA and GEO datasets of GBM. **(A)** Volcano plots of DEGs in TCGA dataset. **(B)** Volcano plots of DEGs in GSE7696 dataset. **(C)** The Venn diagram of overlapping DEGs among TCGA and GSE7696 datasets.



**FIGURE 2 |** The most significantly enriched GO annotations and KEGG pathways of genes related to survival. The length of the bars represents the number of genes, and the color of the bars corresponds to the *p*-value according to legend. **(A)** Top 5 significantly enriched biological process. **(B)** Top 5 significantly enriched cellular component. **(C)** Top 5 significantly enriched molecular function. **(D)** Top 5 significantly enriched KEGG pathways.

package “pROC.” After choosing an optimal cut-off point with the maximal sensitivity and specificity, the survival differences between the low-risk and high-risk groups were assessed by the Kaplan–Meier analysis with log-rank test. Similarly, to evaluate the predictive power of the five-gene signature in internal dataset, we assessed the gene signature within each subtype (classical, mesenchymal, neural, and proneural).

### Analysis in GEPIA and Exploring Co-expression by WGCNA

The expression levels of the five genes were acquired with the assistance of GEPIA<sup>3</sup>, which is a newly developed interactive web server for analyzing the RNA sequencing expression data of 23 types of cancers and normal samples from TCGA

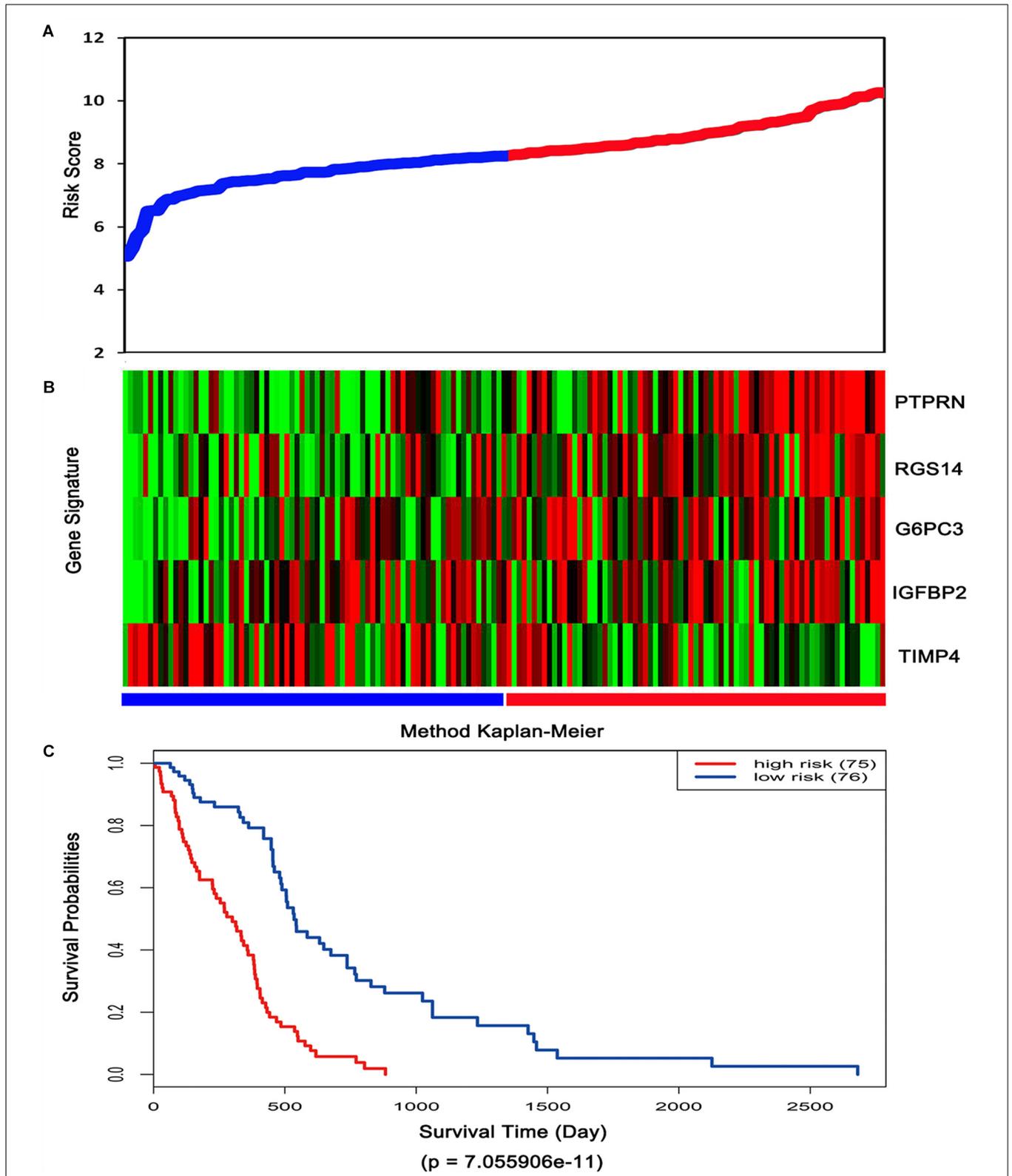
<sup>3</sup><http://gepia.cancer-pku.cn/>

**TABLE 1 |** Information about the five genes screened to build the risk score system.

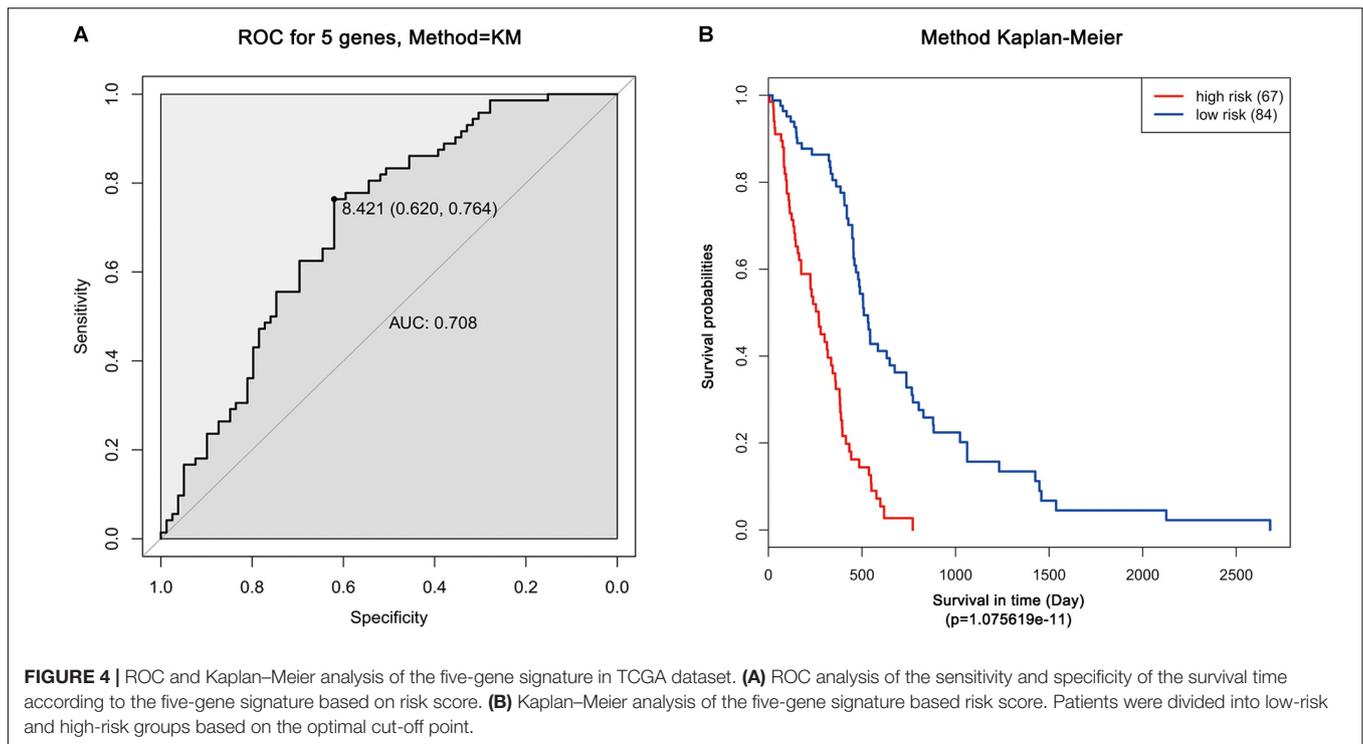
Genes	Coefficient	HR	95% CI	<i>P</i> -value
PTPRN	0.50894	1.66353	1.4010–1.9753	6.35e-09
RGS14	0.54671	1.72757	1.2026–2.4816	0.00309
G6PC3	1.20753	3.34520	1.9960–5.6063	4.57e-06
IGFBP2	0.25845	1.29492	1.1096–1.5112	0.00104
TIMP4	−0.20684	0.81315	0.6951–0.9513	0.00976

and the GTEx projects according to the standard processing pipeline (Tang et al., 2017).

To explore the regulatory network of the five genes, all the overlapped DEGs were analyzed by WGCNA (Ahn et al., 2016; Chen et al., 2018). Finally, the co-expression network of the



**FIGURE 3 |** Risk score analysis, expression distribution and survival analysis of the five-gene signature in TCGA dataset. **(A)** The five-gene signature risk score distribution. **(B)** The heat-map of the five-gene expression profiles. Red indicates a higher expression and green indicates a lower expression. Blue bar: low-risk group. Red bar: high-risk group. **(C)** Kaplan-Meier analysis using the median risk score cut-off which divided patients into low-risk and high-risk groups.



five genes was constructed based on WGCNA and visualized by Cytoscape 3.6.1 (Shannon et al., 2003).

## Validation of the Five-Gene Prognostic Signature by the GEO Dataset and TCGA Microarray Dataset

Dataset GSE13041 from the GEO and TCGA microarray dataset were used to validate this five-gene prognostic signature (Lee et al., 2008). The GSE13041 dataset including 188 samples of GBM patients and the TCGA microarray dataset including 498 samples of GBM patients were both based on the Affymetrix Human Genome U133A Array platform (GPL97). The ROC curves and Kaplan–Meier analyses were used to validate the prognostic value of the five-gene for GBM patients.

## RESULTS

### Differentially Expressed Genes (DEGs) in TCGA and GSE7696

Altogether, 4473 DEGs in TCGA dataset (Figure 1A) and 5789 DEGs in the GSE7696 dataset (Figure 1B) were screened by the limma package. The 2241 overlapping DEGs were screened for further analysis (Figure 1C).

### Survival-Related Genes in GBM

In TCGA dataset, every overlapped DEG was evaluated by univariate Cox regression survival analysis. Altogether, 292 significantly changed genes were considered -survival-related

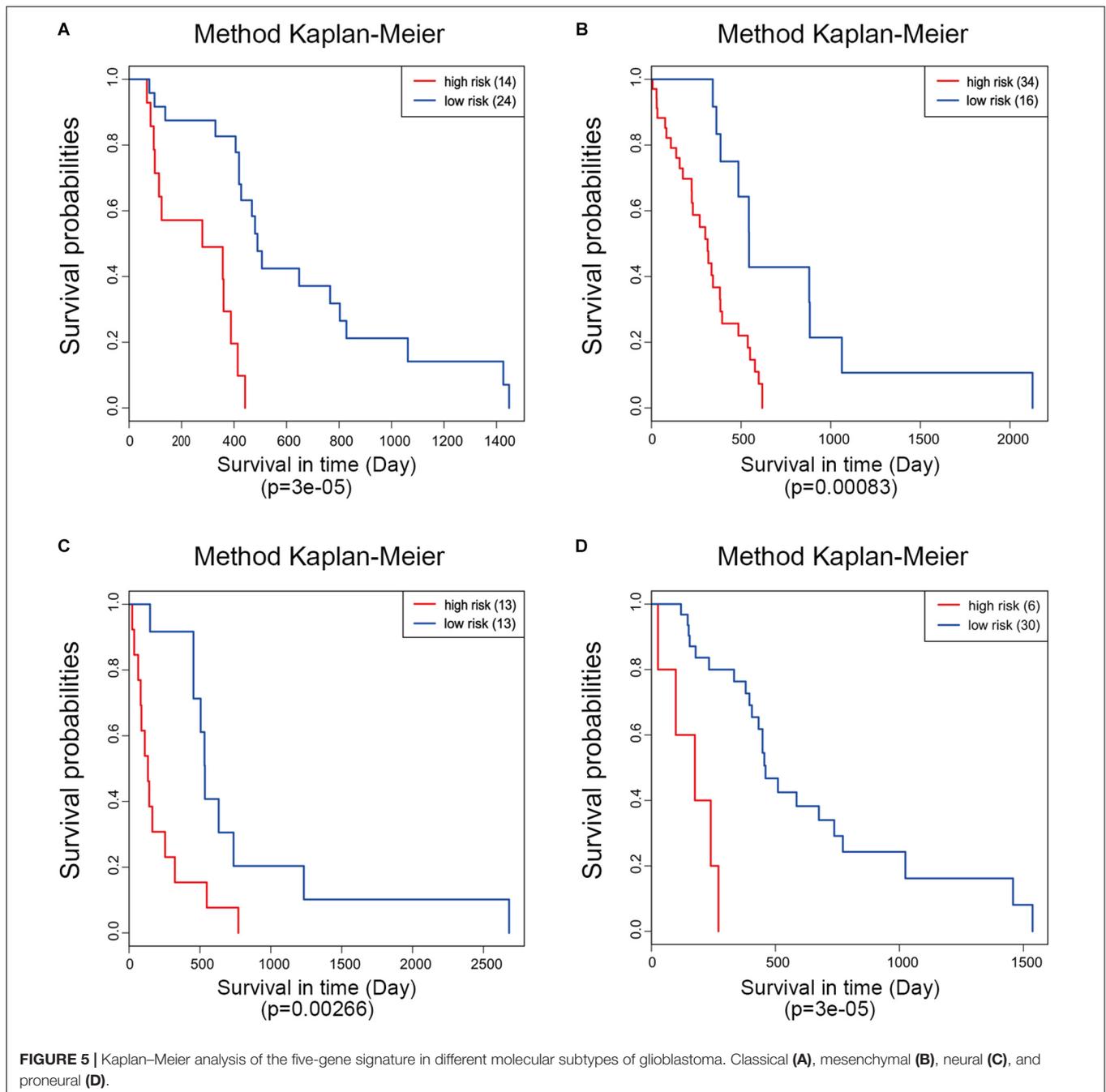
genes by the threshold of  $p < 0.05$ . The top 100 survival-related genes are shown in **Supplementary Table 1**.

### Go and KEGG Analysis of Survival-Related Genes

For the “biological processes” (BP), negative regulation of catalytic activity, regulation of cell shape, negative regulation of monocyte chemotaxis, long-term synaptic potentiation and insulin secretion involved in cellular response to glucose stimulus were the commonly enriched categories (Figure 2A). For the “cellular component” (CC), the enriched categories were correlated with focal adhesion, extracellular space, synaptic vesicle membrane, extracellular exosome, and endoplasmic reticulum (Figure 2B). For the “molecular function” (MF), those genes mainly showed enrichment in calcium ion binding, phospholipase inhibitor activity, calcium-dependent protein binding, calcium-dependent phospholipid binding, and signal transducer activity (Figure 2C). KEGG pathway enrichment analysis suggested that glycosaminoglycan degradation was the most significant pathway. These genes also participated in following pathways: proteoglycans in cancer, lysosome, and regulation of the actin cytoskeleton (Figure 2D).

### Risk Score System Based on Five-Gene Signature

After multivariate Cox regression analysis was conducted for these 100 genes, five genes (PTPRN, RGS14, G6PC3, IGFBP2, and TIMP4) were selected as signature genes that can optimally predict the overall survival of patients with GBM (Table 1). To comprehensively investigate the association between these five



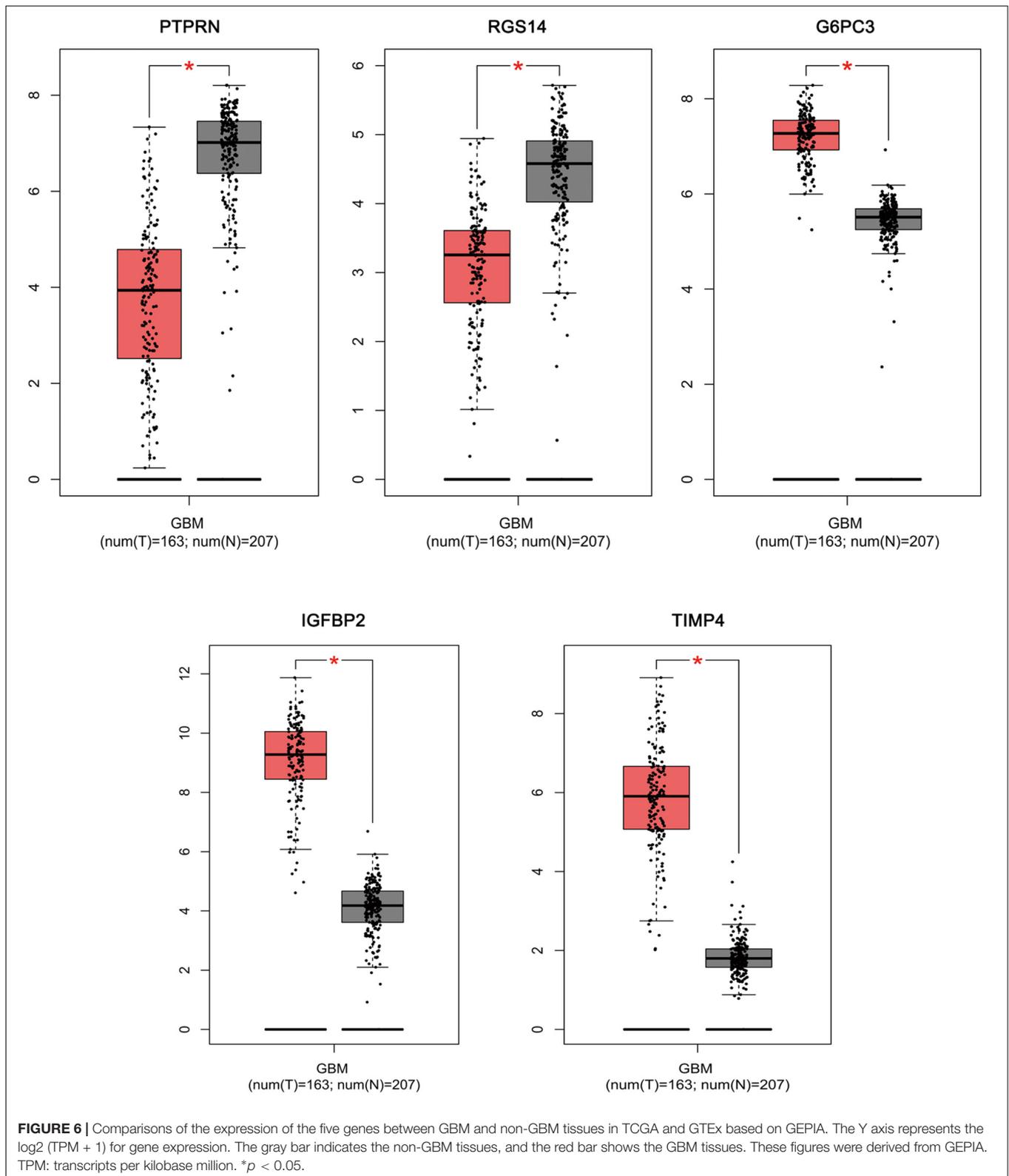
genes and the prognosis of GBM, a five-gene survival risk score system was established based on their Cox coefficients.

$$\text{Risk Score (RS)} = 0.50894 \cdot \text{exp}_{\text{PTPRN}} + 0.54671 \cdot \text{exp}_{\text{PRGS14}} + 1.20753 \cdot \text{exp}_{\text{G6PC3}} + 0.25845 \cdot \text{exp}_{\text{IGFBP2}} - 0.20684 \cdot \text{exp}_{\text{TIMP4}}$$

Then, the risk score for each patient was calculated in TCGA dataset and ranked according to the risk scores. Thus, patients were divided into a high-risk group ( $n = 75$ ) and a low-risk group ( $n = 76$ ). The survival time of GBM patients was adversely

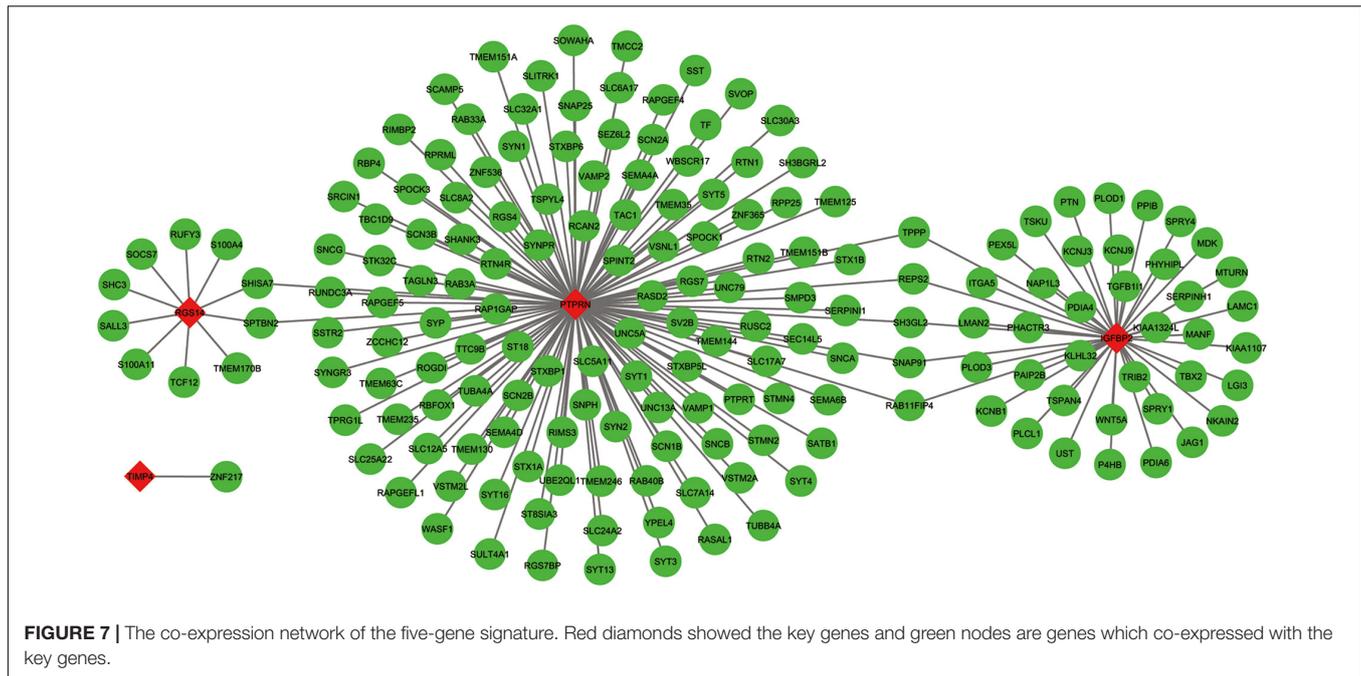
associated with their risk scores (Figure 3A). A remarkably lower expression was noted for TIMP4 in the high-risk groups, while a higher expression was observed for the other genes in the high-risk groups (Figure 3B). The Kaplan–Meier analysis and log-rank test showed that patients in the low-risk group had a significantly positive overall survival time compared to the high-risk group ( $p = 7.055906e-11$ ) (Figure 3C).

Moreover, ROC analysis was performed for this risk score system. Figure 4A shows that the area under the ROC Curves (AUC) was 0.704. The optimal cutoff point was selected as 8.421. With this cutoff point, the patients were further divided into a



high-risk group and a low-risk group. The Kaplan–Meier analysis and log-rank test further indicated a significant difference in overall survival between the two groups ( $p = 1.075619e-11$ )

(Figure 4B). Similarly, with different cutoff points, the patients in each subtype were divided into a high-risk group and a low-risk group. The Kaplan–Meier analysis and log-rank test also



indicated a significant difference between the two groups in each subtype (**Figures 5A–D**).

### Analysis in GEPIA and Exploring Co-expression by WGCNA

Based on the results derived from GEPIA, the expression of G6PC3, IGFBP2, and TIMP4 were significantly up-regulated in GBM, while the expression of PTPRN and RGS14 were significantly down-regulated (**Figure 6**). By using GEPIA, the selected five genes were verified as DEGs in GBM with amplified normal sample sizes.

The co-expressed genes of the five genes were determined by WGCNA. Finally, 129 genes were discovered to be co-expressed with PTPRN, 41 genes were co-expressed with IGFBP2, 10 genes with RGS14 and 1 gene with TIMP4. However, no gene was co-expressed with G6PC3. The co-expression network of the four genes is visualized by WGCNA in **Figure 7**.

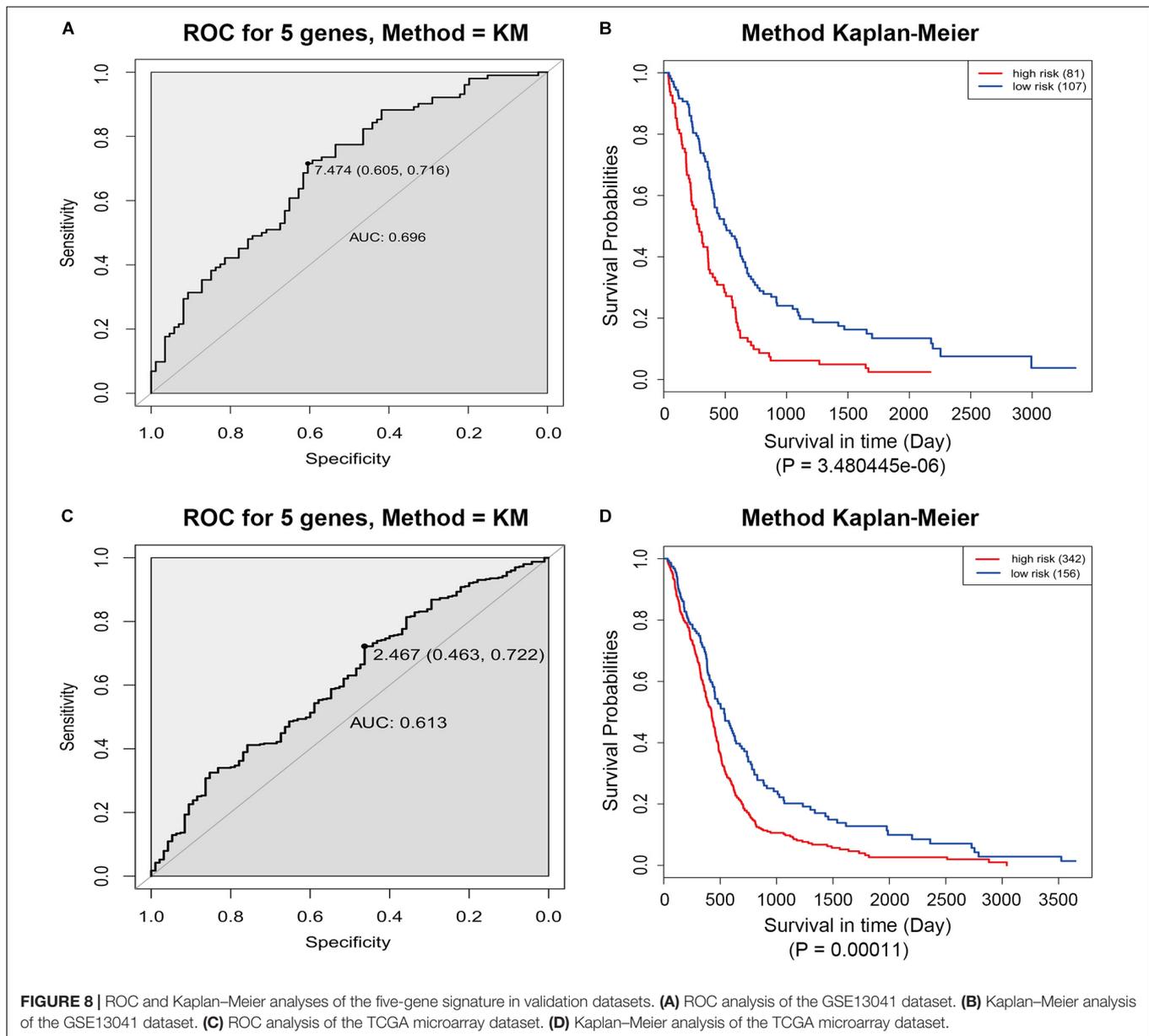
### Validation of the Five-Gene Prognostic Signature by GEO Dataset and TCGA Microarray Dataset

The GSE13041 dataset including 188 GBM patients and the TCGA microarray dataset including 498 GBM patients were used for the validation of the five-gene signature separately. Similarly, the risk score for each patient was calculated. ROC analyses were used to identify the optimal cutoff points (**Figures 8A,C**). Then, we divided the patients into a high-risk group and a low-risk group using the selected optimal cut-off points, respectively. The Kaplan–Meier analyses suggested a significantly prolonged survival time in the low-risk patients compared to that in the high-risk patients ( $p = 3.480445e-06$  and  $p = 0.00011$ ) (**Figures 8B,D**).

## DISCUSSION

GBM is the most aggressive brain tumor associated with poor prognosis. By analyzing TCGA and GSE7696 datasets, we identified 2241 significantly overlapping DEGs. A total of 292 survival-related DEGs were selected from the overlapping DEGs. Functional analyses demonstrated that these genes are mainly associated with following pathways: glycosaminoglycan degradation, proteoglycans in cancer, lysosome, and regulation of the actin cytoskeleton. More importantly, based on multivariate Cox regression analysis of TCGA dataset, five genes which could predict overall survival were screen out, namely PTPRN, RGS14, G6PC3, IGFBP2, and TIMP4. According to their Cox coefficients derived from cox regression, a risk score system based on the five genes was established. Additionally, after identifying the optimal cut-off point by ROC analysis, patients were classified into high-risk and low-risk groups. This five-gene signature was further successfully validated as a prognostic marker in each subtype of GBM, another independent GEO dataset (GSE13041) and TCGA microarray dataset. Furthermore, differential expression analysis of the five genes in GEPIA validated that three genes (G6PC3, IGFBP2, and TIMP4) were significantly up-regulated and two genes (PTPRN and RGS14) were significantly down-regulated in GBM. Co-expression network analysis revealed the regulation network of the five genes. These results suggest that these genes may play an important role in the molecular pathogenesis, progression and prognosis of GBM.

Based on GO and KEGG enrichment analyses of the survival-related DEGs among different studies, “negative regulation of catalytic activity” was the most significant enrichment in BP. This indicated that inhibiting the catalytic activity of some genes may be critical for cancer progression. Coincidentally, Zhao et al. (2009) found that IDH1 mutation could inhibit



IDH1 catalytic activity and contribute to the tumorigenesis of glioma. Other BPs such as regulation of cell shape and negative regulation of monocyte chemotaxis were also enriched. For the CC category, focal adhesion was the most significant enrichment which has been shown to be as a major determinant of cell migration and an essential process in tumor invasion (Garzon-Muvdi et al., 2012). The following three kinds of CCs, extracellular space, synaptic vesicle membrane and extracellular exosome, may also play important roles in tumor development and its micro-environmental manipulation (Wei et al., 2017). Regarding the MF category, calcium ion binding was the most affected MF.  $\text{Ca}^{2+}$ -mediated cell connectivity and plasticity are unique features of the central nervous system, and the  $\text{Ca}^{2+}$ /calmodulin-dependent process is able to regulate cell cycle progression and inhibit proliferation of malignant glioma

(Cheng et al., 1995; Liu et al., 2011). For KEGG pathway enrichment analysis, glycosaminoglycan degradation was the most significant pathway. Extracellular proteoglycans play critical roles in driving oncogenic pathways in tumor cells and promoting critical tumor-microenvironment interactions (Wade et al., 2013). The other KEGG pathways, proteoglycans in cancer, lysosome, and regulation of actin cytoskeleton, were also closely related to oncogenesis (Liu et al., 2012; Terakawa et al., 2013; Wade et al., 2013).

The five-gene signature provides a wealth of potential biological and therapeutic information about GBM. PTPRN (protein tyrosine phosphatase, receptor type N), located on the long arm of human chromosome 2 (2q35) (Lan et al., 1996), is an integral transmembrane protein of dense core vesicles and plays an important role in the secretion of hormones and

neurotransmitters (Xu et al., 2016). PTPRN has been confirmed to be negatively related to the survival of hepatocellular carcinoma patients and closely related to liver tumorigenesis (Zhangyuan et al., 2018). Moreover, the hypermethylation of PTPRN is also associated with shorter survival in ovarian cancer patients (Bauerschlag et al., 2011). A high expression of PTPRN in small cell lung cancer is associated with tumor growth and proliferation. Interestingly, Shergalis et al. also found that a high PTPRN expression is strongly associated with a poor prognosis in GBM patients, which was consistent with our finding (Shergalis et al., 2018). RGS14 is a member of the regulator of the G-protein signaling (RGS) protein family and is highly expressed in the caudate nucleus of the brain, spleen and thymus (Cho et al., 2005; Gerber et al., 2016). Previous study found that RGS14 is important for centrosome function, transcriptional regulation and stress-induced cellular responses (Cho et al., 2005). However, little work has been done to elucidate the role of RGS14 in cancer. Interestingly, PTPRN and RGS14 expressed at low levels in GBM tissue, but their increased expression was associated with poor prognosis. The reason may be that they have different functions in normal and tumor tissues. More work is needed elucidate their functions in GBM. G6PC3, namely, glucose-6-phosphatase isoform  $\beta$ , is a catalysis subunit of G6PC (Gao et al., 2017). G6PC (glucose-6-phosphatase) is a key enzyme that regulates glucose homeostasis and glycogenolysis, which has been reported as a specific enzyme regulating proliferation and invasiveness in several tumors, such as liver, kidney and ovarian cancer (Gao et al., 2017). Furthermore, a previous study revealed that G6PC is a key enzyme regulating glioblastoma invasion (Abbadi et al., 2014). Our study demonstrated that G6PC3 was significantly up-regulated in GBM samples compared with normal brain tissue, and the high expression of G6PC3 was closely related to a poor prognosis in GBM patients. IGFBP2 (Insulin-like growth factor binding protein 2), an important member of the Insulin-like growth factor binding protein family, modulates cell growth, differentiation, migration, and invasion in neoplasms (Fukushima and Kataoka, 2007). IGFBP2 is involved in immunosuppressive activities and is a potential immunotherapeutic target for GBM (Cai et al., 2018). Our study confirmed that IGFBP2 was significantly up-regulated in GBM and predicted a worse outcome for patients, which was consistent with the previous study (Cai et al., 2018). TIMP4 is a member of tissue inhibitors of matrix metalloproteinases (TIMPs), which are involved in several processes of tumorigenesis including proliferation, migration, and invasion (Boufraqech et al., 2016). A high-expression of TIMP4 has been found in patients with breast, cervical, and prostate cancers, whereas a low expression has been observed in patients with pancreatic cancer (Boufraqech et al., 2016).

## REFERENCES

- Abbadi, S., Rodarte, J. J., Abutaleb, A., Lavell, E., Smith, C. L., Ruff, W., et al. (2014). Glucose-6-phosphatase is a key metabolic regulator of glioblastoma invasion. *Mol. Cancer Res.* 12, 1547–1559. doi: 10.1158/1541-7786.MCR-14-0106-T
- Ahn, R., Gupta, R., Lai, K., Chopra, N., Arron, S. T., and Liao, W. (2016). Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs. *BMC Genomics* 17:841. doi: 10.1186/s12864-016-3188-y
- Aldape, K., Zadeh, G., Mansouri, S., Reifenberger, G., and von Deimling, A. (2015). Glioblastoma: pathology, molecular mechanisms and markers. *Acta Neuropathol.* 129, 829–848. doi: 10.1007/s00401-015-1432-1

Interestingly, our study found that TIMP4 was high-expressed in GBM patients, however, its high expression was associated with a good prognosis in patients with GBM. More work is also needed elucidate its functions in GBM. In summary, the five-gene signature not only is robust for predicting the overall survival for GBM, but also has promising practical value in the treatment of GBM.

There are some limitations in our work. First of all, there were only very limited normal samples included in our differential expression analyses, which might neglect some potential mRNAs. Moreover, the efficiency of the five-gene signature should be confirmed in more GBM patients. Furthermore, the molecular mechanisms how the five-gene signature affected the prognosis of GBM patients should be further elucidated by a series of experiments.

## CONCLUSION

In conclusion, our study identified five novel biomarkers that have potential for the prognosis prediction in GBM. Moreover, our findings provide new insights into the pathogenesis and prognosis of GBM.

## AUTHOR CONTRIBUTIONS

WY and XJ conceived and designed the study. GT, QZ, YC, HL, XF, and ZW performed the analysis procedures. GT, WY, and XJ analyzed the results. WY and XJ wrote the manuscript. All authors contributed to the editing of the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (No. 81472355).

## ACKNOWLEDGMENTS

We sincerely acknowledge the public databases: TCGA, GEO, and GEPIA.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00419/full#supplementary-material>

- Bao, Z. S., Li, M. Y., Wang, J. Y., Zhang, C. B., Wang, H. J., Yan, W., et al. (2014). Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling. *CNS Neurosci. Therapeut.* 20, 112–118. doi: 10.1111/cns.12171
- Bauerschlag, D. O., Ammerpohl, O., Brautigam, K., Schem, C., Lin, Q., Weigel, M. T., et al. (2011). Progression-free survival in ovarian cancer is reflected in epigenetic DNA methylation profiles. *Oncology* 80, 12–20. doi: 10.1159/000327746
- Binabaj, M. M., Bahrami, A., ShahidSales, S., Joodi, M., Hassanian, S. M., Anvari, K., et al. (2018). The prognostic value of MGMT promoter methylation in glioblastoma: a meta-analysis of clinical trials. *J. Cell. Physiol.* 233, 378–386. doi: 10.1002/jcp.25896
- Boufraqeh, M., Zhang, L., Nilubol, N., Sadowski, S. M., Kotian, S., Quezado, M., et al. (2016). Lysyl oxidase (LOX) transcriptionally regulates SNAI2 expression and TIMP4 secretion in human cancers. *Clin. Cancer Res.* 22, 4491–4504. doi: 10.1158/1078-0432.CCR-15-2461
- Cai, J., Chen, Q., Cui, Y., Dong, J., Chen, M., Wu, P., et al. (2018). Immune heterogeneity and clinicopathologic characterization of IGF2BP2 in 2447 glioma samples. *Oncoimmunology* 7:e1426516. doi: 10.1080/2162402X.2018.1426516
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. doi: 10.1186/1471-2105-12-35
- Chen, J., Wang, X., Hu, B., He, Y., Qian, X., and Wang, W. (2018). Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. *PeerJ* 6:e4692. doi: 10.7717/peerj.4692
- Cheng, E. H., Gorelick, F. S., Czernik, A. J., Bagaglio, D. M., and Hait, W. N. (1995). Calmodulin-dependent protein kinases in rat glioblastoma. *Cell Growth Differ.* 6, 615–621.
- Cho, H., Kim, D. U., and Kehrl, J. H. (2005). RGS14 is a centrosomal and nuclear cytoplasmic shuttling protein that traffics to promyelocytic leukemia nuclear bodies following heat shock. *J. Biol. Chemistry* 280, 805–814. doi: 10.1074/jbc.m408163200
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:3.
- Fukushima, T., and Kataoka, H. (2007). Roles of insulin-like growth factor binding protein-2 (IGFBP-2) in glioblastoma. *Anticancer Res.* 27, 3685–3692.
- Gao, Y., Li, L., Xing, X., Lin, M., Zeng, Y., Liu, X., et al. (2017). Coronin 3 negatively regulates G6PC3 in HepG2 cells, as identified by label-free mass spectrometry. *Mol. Med. Rep.* 16, 3407–3414. doi: 10.3892/mmr.2017.7002
- Garzon-Muvdi, T., Schiapparelli, P., Smith, C., Kim, D. H., Kone, L., Farber, H., et al. (2012). Regulation of brain tumor dispersal by NKCC1 through a novel role in focal adhesion regulation. *PLoS Biol.* 10:e1001320. doi: 10.1371/journal.pbio.1001320
- Gerber, K. J., Squires, K. E., and Hepler, J. R. (2016). Roles for regulator of G protein signaling proteins in synaptic signaling and plasticity. *Mol. Pharmacol.* 89, 273–286. doi: 10.1124/mol.115.102210
- Huang, J., Liu, F., Liu, Z., Tang, H., Wu, H., Gong, Q., et al. (2017). Immune checkpoint in glioblastoma: promising and challenging. *Front. Pharmacol.* 8:242. doi: 10.3389/fphar.2017.00242
- Jhanwar-Uniyal, M., Labagnara, M., Friedman, M., Kwasnicki, A., and Murali, R. (2015). Glioblastoma: molecular pathways, stem cells and therapeutic targets. *Cancers* 7, 538–555. doi: 10.3390/cancers7020538
- Lambiv, W. L., Vassallo, I., Delorenzi, M., Shay, T., Diserens, A. C., and Misra, A. (2011). The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence. *Neurooncology* 13, 736–747. doi: 10.1093/neuonc/nor036
- Lan, M. S., Modi, W. S., Xie, H., and Notkins, A. L. (1996). Assignment of the IA-2 gene encoding an autoantigen in IDDM to chromosome 2q35. *Diabetologia* 39, 1001–1002. doi: 10.1007/s001250050545
- Lee, Y., Scheck, A. C., Cloughesy, T. F., Lai, A., Dong, J., and Farooqi, H. K. (2008). Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. *BMC Med. Genomics* 1:52. doi: 10.1186/1755-8794-1-52
- Li, J., Lan, C. N., Kong, Y., Feng, S. S., and Huang, T. (2018). Identification and analysis of blood gene expression signature for osteoarthritis with advanced feature selection methods. *Front. Genet.* 9:246. doi: 10.3389/fgene.2018.00246
- Liu, S., Han, Y., Zhang, T., and Yang, Z. (2011). Protective effect of trifluoperazine on hydrogen peroxide-induced apoptosis in PC12 cells. *Brain Res. Bull.* 84, 183–188. doi: 10.1016/j.brainresbull.2010.12.008
- Liu, Y., Zhou, Y., and Zhu, K. (2012). Inhibition of glioma cell lysosome exocytosis inhibits glioma invasion. *PLoS One* 7:e45910. doi: 10.1371/journal.pone.0045910
- Luo, D., Deng, B., Weng, M., Luo, Z., and Nie, X. (2018). A prognostic 4-lncRNA expression signature for lung squamous cell carcinoma. *Artif. Cells Nanomed. Biotechnol.* 46, 1207–1214. doi: 10.1080/21691401.2017.1366334
- Ma, Q., Long, W., Xing, C., Chu, J., Luo, M., Wang, H. Y., et al. (2018). Cancer stem cells and immunosuppressive microenvironment in Glioma. *Front. Immunol.* 9:2924. doi: 10.3389/fimmu.2018.02924
- Mao, X., Qin, X., Li, L., Zhou, J., Zhou, M., Li, X., et al. (2018). A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecol. Oncol.* 149, 181–187. doi: 10.1016/j.ygyno.2017.12.011
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34.
- O'Quigley, J., and Moreau, T. (1986). Cox's regression model: computing a goodness of fit statistic. *Comput. Methods Programs Biomed.* 22, 253–256. doi: 10.1016/0169-2607(86)90001-5
- Ostrom, Q. T., Gittleman, H., Farah, P., Ondracek, A., Chen, Y., Wolinsky, Y., et al. (2013). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neurooncology* 15(Suppl. 2), iii–iii56. doi: 10.1093/neuonc/nou151
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shergalis, A., Bankhead, A. III, Luesakul, U., Muangsin, N., and Neamati, N. (2018). Current challenges and opportunities in treating glioblastoma. *Pharmacol. Rev.* 70, 412–445. doi: 10.1124/pr.117.014944
- Szopa, W., Burley, T. A., Kramer-Marek, G., and Kaspera, W. (2017). Diagnostic and therapeutic biomarkers in glioblastoma: current status and future perspectives. *BioMed Res. Int.* 2017:8013575. doi: 10.1155/2017/8013575
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi: 10.1093/nar/gkx247
- Terakawa, Y., Agnihotri, S., Golbourn, B., Nadi, M., Sabha, N., and Smith, C. A. (2013). The role of drebrin in glioma migration and invasion. *Exp. Cell Res.* 319, 517–528. doi: 10.1016/j.yexcr.2012.11.008
- van den Bent, M. J., Weller, M., Wen, P. Y., Kros, J. M., Aldape, K., and Chang, S. (2017). A clinical perspective on the 2016 WHO brain tumor classification and routine molecular diagnostics. *Neurooncology* 19, 614–624. doi: 10.1093/neuonc/now277
- Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110. doi: 10.1016/j.ccr.2009.12.020
- Wade, A., Robinson, A. E., Engler, J. R., Petritsch, C., James, C. D., and Phillips, J. J. (2013). Proteoglycans and their roles in brain cancer. *FEBS J.* 280, 2399–2417. doi: 10.1111/febs.12109
- Wangguo, S., Feng, C., Xie, Y., Ye, L., Wang, F., and Li, X. (2016). Sample level enrichment analysis of KEGG pathways identifies clinically relevant subtypes of Glioblastoma. *J. Cancer* 7, 1701–1710. doi: 10.7150/jca.15486
- Wei, Z., Batagov, A. O., Schinelli, S., Wang, J., Wang, Y., El Fatimy, R., et al. (2017). Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat. Commun.* 8:1145. doi: 10.1038/s41467-017-01196-x
- Wen, P. Y., and Kesari, S. (2008). Malignant gliomas in adults. *New Engl. J. Med.* 359, 492–507. doi: 10.1056/nejmra0708126
- Xu, H., Cai, T., Carmona, G. N., Abuhatzira, L., and Notkins, A. L. (2016). Small cell lung cancer growth is inhibited by miR-342 through its effect of the target gene IA-2. *J. Transl. Med.* 14:278.

- Yang, R., Xiong, J., Deng, D., Wang, Y., Liu, H., Jiang, G., et al. (2016). An integrated model of clinical information and gene expression for prediction of survival in ovarian cancer patients. *Transl. Res.* 172, 84.e11–95.e11.
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zeng, W. J., Yang, Y. L., Liu, Z. Z., Wen, Z. P., Chen, Y. H., Hu, X. L., et al. (2018). Integrative analysis of DNA methylation and gene expression identify a three-gene signature for predicting prognosis in lower-grade gliomas. *Cell. Physiol. Biochem.* 47, 428–439. doi: 10.1159/000489954
- Zhangyuan, G., Yin, Y., Zhang, W., Yu, W., Jin, K., Wang, F., et al. (2018). Prognostic value of phosphotyrosine phosphatases in hepatocellular carcinoma. *Cell. Physiol. Biochem.* 46, 2335–2346. doi: 10.1159/000489625
- Zhao, S., Lin, Y., Xu, W., Jiang, W., Zha, Z., Wang, P., et al. (2009). Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* 324, 261–265. doi: 10.1126/science.1170944

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yin, Tang, Zhou, Cao, Li, Fu, Wu and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.