

ORIGINAL RESEARCH

Identification of Prognostic Markers of Glioblastoma through Bioinformatics Analysis

Jieying Wen, MM; Haojie Zheng, MM; Xi Yuan, MM; Cuilan Huang, MM;
Xiaogang Yang, MM; Zhiying Lin, MM; Guanglong Huang, MD

ABSTRACT

Objective • Glioblastoma is the most common and aggressive type of the central nervous system cancers. Although radiotherapy and chemotherapy are used in the treatment of glioblastoma, survival rates remain unsatisfactory. This study aimed to explore differentially expressed genes (DEGs) based on the survival prognosis of patients with glioblastoma and to establish a model for classifying patients into different risk groups for overall survival.

Methods • DEGs from 160 tumor samples from patients with glioblastoma and 5 nontumor samples from other patients in The Cancer Genome Atlas database were identified. Functional enrichment analysis and a protein-protein interaction network were used to analyze the DEGs. The prognostic DEGs were identified by univariate Cox regression analysis. We split patient data from The Cancer Genome Atlas database into a high-risk group and a low-risk group as the training data set. Least absolute shrinkage and selection operator and multiple Cox regression were used to construct a prognostic risk model, which was validated in a test data set from The Cancer Genome Atlas database and was analyzed using external data sets from the Chinese Glioma Genome Atlas database

and the GSE74187 and GSE83300 data sets. Furthermore, we constructed and validated a nomogram to predict survival of patients with glioblastoma.

Results • A total of 3572 prognostic DEGs were identified. Functional analysis indicated that these DEGs were mainly involved in the cell cycle and focal adhesion. Least absolute shrinkage and selection operator regression identified 3 prognostic DEGs (*EFEMP2*, *PTPRN*, and *POM121L9P*), and we constructed a prognostic risk model. The receiver operating characteristic curve analysis showed that the areas under the curve were 0.83 for the training data set and 0.756 for the test data set. The predictive performance of the prognostic risk model was validated in the 3 external data sets. The nomogram showed that the prognostic risk model was reliable and that the accuracy of predicting survival in each patient was high.

Conclusion • The prognostic risk model can effectively classify patients with glioblastoma into high-risk and low-risk groups in terms of overall survival rate, which may help select high-risk patients with glioblastoma for more intensive treatment. (*Altern Ther Health Med*. 2025;31(1):401-407).

Jieying Wen, MM; Haojie Zheng, MM; Xi Yuan, MM; Guanglong Huang, MD; Department of Neurosurgery, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China; Cuilan Huang, MM; Xiaogang Yang, MM; Zhiying Lin, MM; Department of Intensive Care Unit, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, Jiangxi, China.

Corresponding author: Guanglong Huang, MD
E-mail: hgl1020@163.com

Corresponding author: Zhiying Lin, MM
E-mail: lzy1191528412@163.com

INTRODUCTION

Glioblastoma is a World Health Organization grade IV glioma, the most common and invasive primary malignant brain tumor globally, accounting for approximately 60% to

70% of malignant gliomas.¹ Despite the progress made in the past decade, glioblastoma is still one of the most difficult tumor types to treat. The median survival time of patients with glioblastoma is only 12 to 15 months.² Therefore, it is important to study the molecular mechanisms related to glioblastoma. Recently, tumor-targeted molecular therapy has been used to supplement conventional therapy in different cancer types.³ Compared with conventional chemotherapy, molecular therapy attacks malignant tumor cells more specifically with fewer side effects.⁴

It is imperative to develop prognostic biomarkers to closely monitor disease progression and shed light on treatment stratification. Differentially expressed genes (DEGs) can be used to accurately estimate the overall survival (OS) rate of a disease and are potential markers for clinical treatment.^{5,6} Recently, studies have shown that models constructed using data from the expression of

multiple genes can better predict and classify the survival outcomes of patients with glioblastoma.^{7,8} Nevertheless, these models have not been used in routine clinical practice, which may be related to a lack of consistent evidence, small sample sizes, and the large amount of data fitting present in these studies. Therefore, the most reliable biomarkers to predict and classify the prognosis of glioblastoma have the possibility of being identified using publicly accessible databases of gene expression patterns, such as The Cancer Genome Atlas (TCGA), the Chinese Glioma Genome Atlas (CGGA), and the Gene Expression Omnibus.

In this study, we collected gene expression matrixes from the TCGA, CGGA, and Gene Expression Omnibus databases to investigate the clinical implications of DEGs on prognostic stratification and their potential as biomarkers for targeted glioblastoma therapy. We further established a 3-gene prognostic risk model to divide glioblastomas into 2 groups based on risk. The high-risk group had shorter OS than the low-risk group and showed different molecular features. The prognostic risk model was developed and validated in independent data sets, and underlying mechanisms were explored using bioinformatics analysis.

METHODS

Data source

The data used in this study were normalized RNA-sequencing data sets and corresponding clinical information from 160 tumor samples from patients with glioblastoma and 5 nontumor samples from other patients in the TCGA database, 374 patients with glioblastoma in the CGGA database, 60 patients with glioblastoma in the GSE74187 data set, and 50 patients with glioblastoma in the GSE83300 data set.

Data preprocessing and differential analysis

DEGs between glioblastoma tumor samples and normal tissues were identified using the Wilcoxon test after within-array replicate probes were replaced with their mean via the limma package of R version 3.6.1 (R Foundation for Statistical Computing).^{9,10} The *P* value was adjusted using the false discovery rate. False discovery rate less than .05 and $|\log_2(\text{fold change})|$ greater than 2 were considered significant.

The Kyoto Encyclopedia of Genes and Genomes (KEGG)¹¹ pathway enrichment gene ontology¹² function was analyzed with the prognostic DEGs using the clusterProfiler package of R.¹³ *P* < .05 was considered statistically significant.

Development and validation of a prognostic risk model

Patients from the TCGA data set were randomly assigned in a 5:5 ratio to a training data set and a test data set. Using the expression profiles of the identified prognostic DEGs, we conducted least absolute shrinkage and selection operator (LASSO) regression analysis in the training data set. The Akaike information criterion value was used for further analysis by multivariate Cox regression with LASSO penalty. Subsequently, we calculated individual risk scores with

coefficients and constructed a prognostic risk model that separated the high-risk group and low-risk group.

Risk scores were established in accordance with the LASSO regression coefficient, and the risk-score formula was constructed as follows:

$$\text{risk score} = \sum_{n=1}^I (\text{Exp}_n \times \text{COFF}_n)$$

In the formula, *I* is the number of selected genes, Exp_n is the expression value of each gene, and COFF_n is the LASSO regression coefficient.

Clinical relevance was validated using survival analysis between groups using the survival and survminer packages of R; receiver operating characteristic (ROC) analysis was performed using the survival ROC package of R, and the area under the curve (AUC) was calculated at multiple time points to evaluate discrimination.

Clinical characteristics and pathological features, including age, sex, isocitrate dehydrogenase status, the promoter methylation status of the O-6-methylguanine-DNA methyltransferase gene, and Karnofsky performance scale scores, were collected from the TCGA database and integrated with transcriptome profile data derived from the TCGA database.

A multivariate proportional hazards model (Cox) regression analysis was performed using clinical data and risk scores to demonstrate that the prognostic value of risk scores was the independent prognostic risk factor. *P* < .05 was considered statistically significant.

External validation of the prognostic risk model

The prognostic risk model with the same risk-score formula and cutoff value to separate the high-risk and low-risk groups was validated in the test data set and external data set from the CGGA database (*n* = 374), GSE74187 data set (*n* = 60), and GSE83300 data set (*n* = 50). Likewise, the prognostic risk model was presented as a risk plot for each data set that encompassed the expression level of the included genes, the distribution of risk scores, and the survival status of individuals.

Construction of a nomogram based on the prognostic risk model

A nomogram was constructed for the risk scores based on the expression of the prognostic risk model and clinicopathological factors by using the rms package in R. Discrimination of the nomogram was validated using the ROC package in R to analyze the data at 1-year and 2-year follow-up, and predictive accuracy was tested by presenting the difference between predicted survival and actual survival using a calibration plot.

RESULTS

Screening for DEGs and functional analysis

We identified 3572 DEGs from the TCGA database. The DEGs comprised 2571 upregulated genes and 956

downregulated genes, using the criterion of $|\log_2(\text{fold change})|$ greater than 2. The top 3 upregulated genes were *PRAME*, *ELDR*, and *HOXD9*, and the top 3 downregulated genes were *NEUROD6*, *PRKCG*, and *C1QL3* (Figure 1A).

We annotated the functions of the DEGs using KEGG enrichment pathway analysis. The results indicated that the DEGs were enriched in focal adhesion, cell cycle, cell adhesion molecules, and extracellular matrix–receptor interaction ($P < .05$; Figure 1B). Furthermore, gene ontology enrichment analysis for biological processes also demonstrated that the DEGs were mainly enriched in leukocyte migration, signal release, and regulation of exocytosis, regulation of ion transmembrane transport, and neurotransmitter transport (Figure 1C).

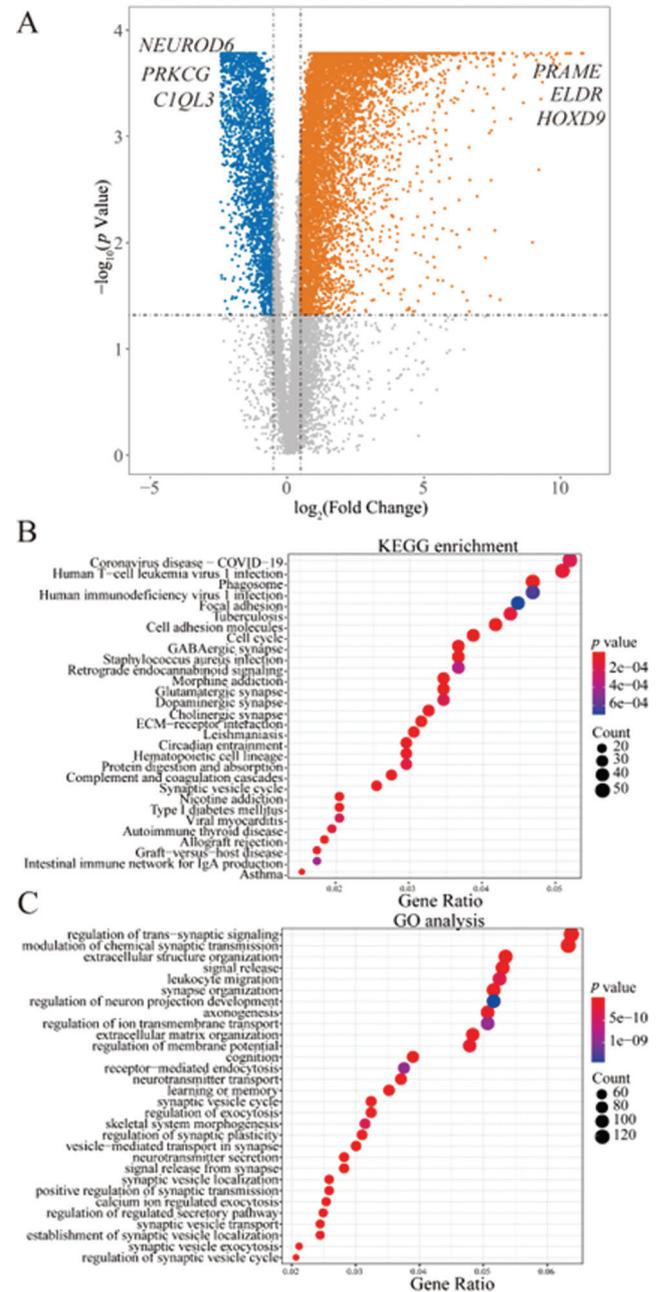
Construction of a prognostic model in the training data set

DEGs, which are closely related to tumor initiation and progression, are potential prognostic biomarkers. To identify potential prognostic DEGs, the TCGA cohort was randomly divided into a training data set ($n=80$) and an internal test data set ($n=80$) with an approximate ratio of 1:1. A univariate Cox regression analysis was used to screen out the prognostic DEGs that were related to OS, and 270 prognostic DEGs were identified. The complex protein-protein interaction network of these prognostic DEGs consisted of 41 nodes and 116 edges. We used the MCODE (Molecular Complex Detection) plug-in of Cytoscape software version 3.9.1 (Cytoscape Consortium) to conduct a module analysis of the network. We found 2 modules in the protein-protein interaction network. The first module, composed of 13 nodes that interacted most intensively, the cluster's computed score was 3.15 (Figure S1).

From the 270 prognostic DEGs, 7 survival-related genes were identified by LASSO penalized regression (Figure 2A and 2B, Table 1). Then, we screened these 7 genes with the minimum Akaike information criterion value and constructed a prognostic risk model composed of 3 potential prognostic genes as core genes: *EFEMP2*, *PTPRN*, and *POM121L9P* (Figure 2C).

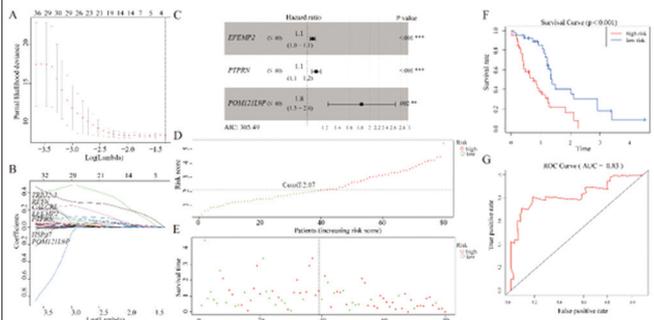
A risk score was calculated for each patient in the TCGA training data set by combining the relative expression of the 3 DEGs in the prognostic risk model and the LASSO coefficients. Patients with a risk score of 2.07 or greater (median cutoff) were classified as high risk, and the remaining patients were classified as low risk (Figure 2D). The expression of *EFEMP2*, *PTPRN*, and *POM121L9P* was significantly higher in the high-risk group than in the low-risk group in the training data set (Figure S2). Patients in the high-risk group had a worse prognosis than patients in the low-risk group in the training data set (Figure 2E). To investigate the relationship between risk score and the OS of patients with glioblastoma, a Kaplan-Meier analysis and a log-rank test were performed using the training data set. The median survival time of patients was 5.95 months in the high-risk group and 12.66 months in the low-risk group; the survival time was significantly shorter ($P < .001$) for patients in the

Figure 1. DEGs in Glioblastoma From The Cancer Genome Atlas Database A, DEGs were screened using the limma package of R. The red points represent significantly upregulated DEGs, and the green points represent significantly downregulated DEGs. *NEUROD6*, *PRKCG*, *C1QL3* were the top 3 downregulated genes. *PRAME*, *ELDR*, and *HOXD9* were the top 3 upregulated genes. B, KEGG enrichment pathway analysis showed the 3572 DEGs are involved in focal adhesion, cell cycle, cell adhesion molecules, and ECM-receptor interaction. C, GO enrichment analysis showed that the 3572 DEGs are enriched in leukocyte migration, signal release, regulation of exocytosis, regulation of ion transmembrane transport, and neurotransmitter transport.



Abbreviations: DEG, differentially expressed gene; ECM, extracellular matrix; GO, gene ontology.

Figure 2. Construction of a Prognostic Model Based on the 270 Prognostic Differentially Expressed Genes in the Training Data Set. A, B, LASSO penalized regression identified 7 survival-related genes. C, Multiple Cox regression with LASSO penalty identified 3 potential prognostic genes. D-F, Prognostic classifier analysis of the patients in the internal test data set. G, ROC curve of the 3 potential prognostic genes for glioblastoma.



** represent $P < .01$
 *** represent $P < .001$.

Abbreviations: AIC, Akaike information criterion; AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

Table 1. Features Selected by Multivariate Cox Proportional Hazard Regression Model With LASSO Penalty

Gene	LASSO Coefficient	HR	HR.95L	HR.95H	P value
<i>EFEMP2</i>	0.062	1.06	1.03	1.10	<.001
<i>PTPRN</i>	0.101	1.11	1.06	1.16	<.001
<i>POM121L9P</i>	0.599	1.82	1.26	2.64	<.001

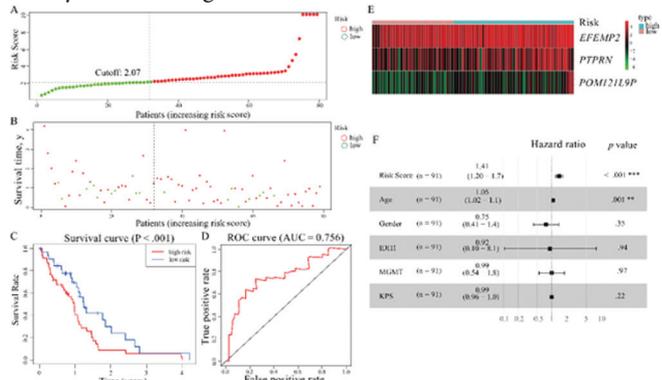
Abbreviations: HR, hazard ratio; HR.95H: higher value of 95% confidence interval of hazard ratio; HR.95L: lower value of 95% confidence interval of hazard ratio; LASSO, least absolute shrinkage and selection operator.

high-risk group than for patients in the low-risk group (Figure 2F). The AUC value was 0.83 for the time-dependent ROC curve (Figure 2G).

Validation of the prognostic model using the TCGA test data set

To further explore the relationship between risk score and OS of patients with glioblastoma, a Kaplan-Meier analysis and log-rank test were performed on the TCGA test data set. We used the same prognostic risk model; using the median of all risk scores, patients with a risk score of 2.07 or greater were classified as high risk, and the remaining patients were classified as low risk (Figure 3A and 3B). The OS was significantly shorter for patients with a higher risk score than for patients with a lower risk score ($P < .001$; Figure 3C). As most deaths occurred within 5 years after diagnosis, we used a time-dependent ROC curve to assess prognosis (Figure 3D); the AUC value was 0.756. The expression of *EFEMP2*, *PTPRN*, and *POM121L9P* was significantly higher in the high-risk group than in the low-risk group in the TCGA test data set (Figure 3E). By excluding the influence of age, gender, isocitrate dehydrogenase status, methylation status of the *O*-6-

Figure 3. Validation of the 3 Potential Prognostic Genes for Glioblastoma in the TCGA Test Data Set. A-C, Survival status of high-risk and low-risk patients with glioblastoma by risk score (A), survival time (B), and survival rate (C). D, The ROC curve of the 3 potential prognostic genes for glioblastoma. E, Expression profile of 3 potential prognostic genes for glioblastoma in the test data set. F, Multivariable analyses of the risk score, age, gender, isocitrate dehydrogenase status, promoter methylation status of the *O*-6-methylguanine-DNA methyltransferase gene, and KPS scores.



** represent $P < .01$
 *** represent $P < .001$

Abbreviations: AUC, area under the curve; IDH1, isocitrate dehydrogenase; KPS, Karnofsky performance scale; MGMT, *O*-6-methylguanine-DNA methyltransferase; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas.

Figure 4. Gene Set Enrichment Analysis. Gene set enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the high-risk group (A) and the low-risk group (B).

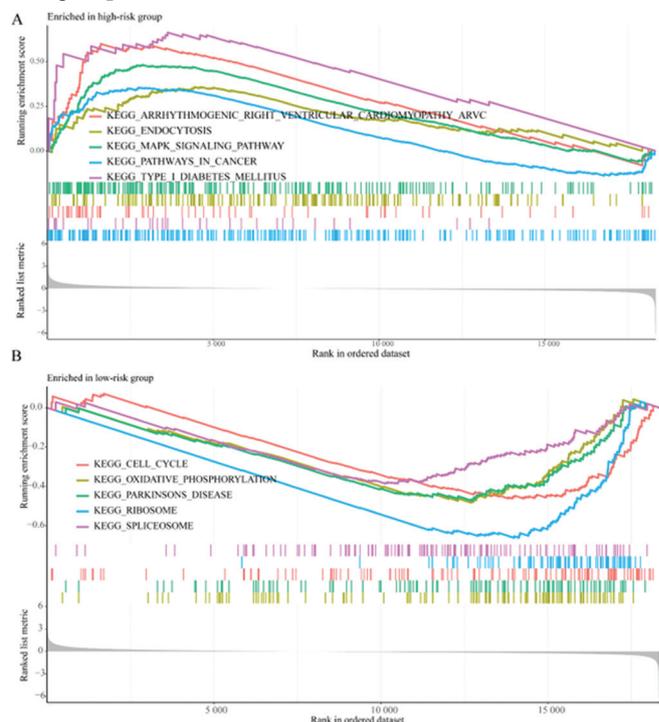
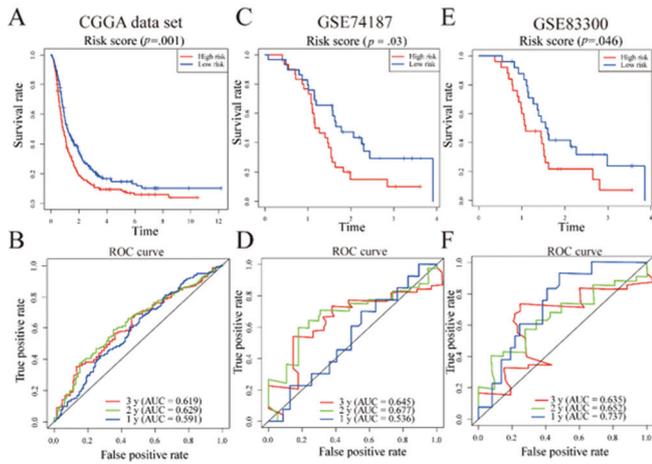
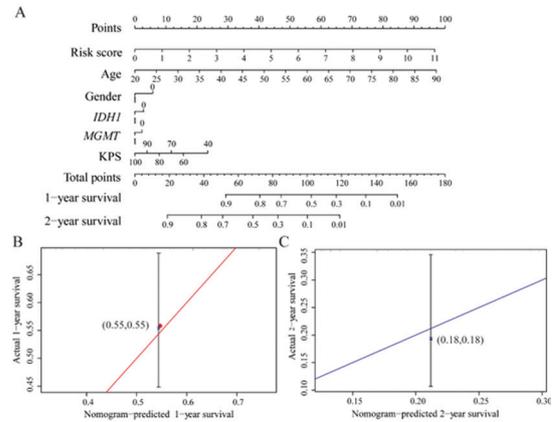


Figure 5. Validation of the 3 Potential Prognostic Genes for Glioblastoma. Validation of the 3 potential prognostic genes for glioblastoma in the CGGA data set (A, B), GSE74187 data set (C, D), and GSE83300 data set (E, F).



Abbreviations: AUC, area under the curve; ROC, receiver operating characteristic.

Figure 6. Construction of a Nomogram Based on the 3 Potential Prognostic Genes for Glioblastoma. A, A nomogram was constructed based on risk scores and clinical information. Calibration plots for the predictive accuracy of the nomogram at 1-year survival (B) and at 2-year survival (C).



Abbreviations: IDH1, isocitrate dehydrogenase; KPS, Karnofsky performance scale; MGMT, O-6-methylguanine-DNA methyltransferase.

methylguanine-DNA methyltransferase gene promoter, and Karnofsky performance scale score on prognosis, we showed that the risk score was an independent prognostic risk factor for patients with glioblastoma. Independent prognostic risk analysis showed that the hazard ratio was 1.41 (95% CI, 1.20-1.70) in the TCGA test data set.

External validation of the prognostic risk model

To verify that our prognostic risk model can be applied universally, we applied the 3-gene prognostic risk model to the CGGA, GSE74187, and GSE83300 data sets.

A total of 374 samples were obtained from the CGGA database, and the prognostic risk model successfully subdivided the patients into a high-risk group and a low-risk group. The OS was significantly different between the 2 groups, and the median overall survival time of patients was 9.72 months in the high-risk group and 12.48 months in the low-risk group. The survival rate of patients in the high-risk group was low (Figure 4A). The AUCs of the 1-year, 2-year, and 3-year ROC curves were 0.591, 0.629, and 0.619, respectively (Figure 4B).

Furthermore, we verified in the GSE74187 data set (Figure 4C) that patients with GBM in the high-risk group had poor prognosis. In the GSE74187 data set, the AUCs of the 1-year, 2-year, and 3-year ROC curves were 0.536, 0.677, and 0.645, respectively (Figure 4D). We also verified in the GSE83300 data set (Figure 4E). In the GSE83300 data set, the AUCs of the 1-year, 2-year, and 3-year ROC curves were 0.737, 0.652, and 0.635, respectively (Figure 4F). Therefore, we verified using 3 data sets that the 3-gene prognostic risk model has good prediction performance.

We analyzed the molecular mechanism that promotes the malignant progression of GBM, and gene set enrichment analysis was used to predict the possible biological functions

of the products of the DEGs of each risk groups in GBM. CGGA data showed that the products of the DEGs of each risk group messenger RNA was significantly correlated with the mitogen-activated protein kinase signaling pathway, and the cell cycle (Figure 5A and 5B).

Developing and validating a predictive nomogram based on the 3-gene prognostic risk model

To confirm the prognostic value of the 3-gene prognostic risk model, we constructed a nomogram based on the prognostic risk model and determined the clinical relevance and prognostic value of age, gender, isocitrate dehydrogenase status, the methylation status of the O-6-methylguanine-DNA methyltransferase gene promoter, and the Karnofsky performance scale score. The 1-year, 3-year, and 5-year survival rates were estimated from the total points, which are the sum of the points for each item, as shown in the nomogram (Figure 6A). Analysis of the nomogram showed that the prognostic risk model was reliable and the accuracy of predicting survival in each patient was high. By comparing the factors in the nomogram, we also found that the prognostic risk model accounts for a high risk score and plays an important role in predicting survival. The calibration curves showed that the predicted and actual survival rates matched well, which indicates that the nomogram can accurately predict patient survival (Figure 6B and 6C).

DISCUSSION

GBM is the most common malignant tumor type in the central nervous system,¹⁴ and there is no targeted therapy to ensure the maximum survival rate of patients with glioma.^{3,15} In recent years, a large number of researchers have used bioinformatics to analyze the data of thousands of expressed genes in the human genome through high-throughput

sequencing and microarray analysis, which can be used to identify the gene characteristics existing in GBM and reveal its potential mechanism.¹⁶

In this study, we identified 270 prognostic DEGs by analyzing the TCGA database. After multivariate Cox regression with LASSO penalty, we identified 3 DEGs (*EFEMP2*, *PTPRN*, and *POM121L9P*), and validation analyses were performed using 3 independent data sets, showing good reproducibility.

Epidermal growth factor (EGF)-containing fibulin-like extracellular matrix protein 2 (*EFEMP2*), containing 4 EGF2 domains and 6 calcium-binding EGF2 domains, is a member of the fibulin family.¹⁷ *EFEMP2* is overexpressed in glioma and significantly positively correlated with growth and cell-cycle progression.¹⁸ *EFEMP2* might play a role in maintaining the oncogenesis of GBM cell lines in vitro.¹⁹ Receptor-type tyrosine-protein phosphatase-like N (*PTPRN*), also known as islet cell antigen 512 (*ICA512*), is a tyrosine phosphatase-like intrinsic membrane protein.²⁰ Its activation promotes cell migration and is involved in immune infiltration of lung adenocarcinoma and natural killer-cell activation.²¹ *PTPRN* may be a potential marker in lung adenocarcinoma. The *POM121L9P* locus encodes for *POM121* transmembrane nucleoporin-like 9, pseudogene. Douglas et al²² found that the overexpression of *POM121L9P* was associated with a poorer prognostic outcome in patients with epithelial ovarian cancer.

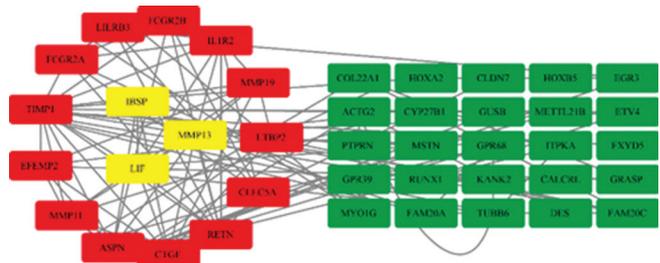
Our study had some differences compared with previous research.²³ Our strategy involved LASSO penalized regression, which can analyze all independent variables as well as the most influential variables. When dealing with large data sets such as gene-expression profiles, this method is much more accurate than the stepwise regression method of multivariate Cox regression models. Moreover, we then used a TCGA test data set for internal validation, and we used data from the CGGA database and the GSE74187 and GSE83300 data sets to validate the 3 prognostic DEGs.

This study has the following limitations. First, this study was based on publicly available databases. Although we carried out external validation among different databases, validation from clinical trials is still lacking. Second, we only used selected representative databases and data sets, and other databases could be analyzed and other data sets could be collected to verify the results of this study.

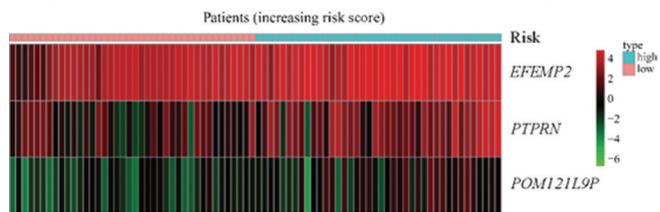
CONCLUSION

In summary, 7 survival-related DEGs were identified by LASSO penalized regression. We constructed a prognostic risk model composed of 3 core genes and developed a 3-gene signature that is associated with OS in patients with GBM. Our findings suggest that the 3-gene prognostic risk model can serve as an independent biomarker for predicting survival prognosis. We are poised for further investigation and eagerly anticipate the verification of our findings in a larger cohort of patients to assess whether the 7 genes we identified are likely to become new targets of drug treatments.

Supplementary Figure 1. The Complex Protein-Protein Interaction Network of the 270 Prognostic Differentially Expressed Genes. Plot showed the module composed of 13 nodes with the 3.15 cluster's computed score.



Supplementary Figure 2. Expression Profile of 3 Potential Prognostic Genes for Glioblastoma in the Training Data Set



CONFLICT OF INTEREST

The authors declare that they have no competing interests.

FUNDING

This study was supported by the Guangdong Science and Technology Department (2016A040403053).

AUTHORS' CONTRIBUTIONS

Conception and design of the work: JYW. Acquisition, analysis, and interpretation of data: JYW, HJZ, XY, CLH, and XGY. Drafting of the manuscript: ZYL and GLH. Critical revision for important intellectual content: ZYL and GLH. All authors approved the final version of this manuscript.

ACKNOWLEDGEMENTS

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data sets generated for this study used data from the TCGA database at <https://www.cancer.gov/ccg/research/genome-sequencing/tcga> and <http://www.cgga.org.cn/>.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

REFERENCES

- Weller M, Wick W, Aldape K, et al. Glioma. *Nat Rev Dis Primers*. 2015;1(1):15017. doi:10.1038/nrdp.2015.17
- Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310(17):1842-1850. doi:10.1001/jama.2013.280319
- Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803-820. doi:10.1007/s00401-016-1545-1
- Parker NR, Khong P, Parkinson JF, Howell VM, Wheeler HR. Molecular heterogeneity in glioblastoma: potential clinical implications. *Front Oncol*. 2015;5:55. doi:10.3389/fonc.2015.00055
- Aldape K, Zadeh G, Mansouri S, Reifenberger G, von Deimling A. Glioblastoma: pathology, molecular mechanisms and markers. *Acta Neuropathol*. 2015;129(6):829-848. doi:10.1007/s00401-015-1432-1
- Szopa W, Burley TA, Kramer-Marek G, Kaspera W. Diagnostic and therapeutic biomarkers in glioblastoma: current status and future perspectives. *BioMed Res Int*. 2017;2017:8013575. doi:10.1155/2017/8013575
- Bao ZS, Li MY, Wang JY, et al. Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling. *CNS Neurosci Ther*. 2014;20(2):112-118. doi:10.1111/cns.12171
- Wang JJ, Wang H, Zhu BL, et al. Development of a prognostic model of glioma based on immune-related genes. *Oncol Lett*. 2021;21(2):116. doi:10.3892/ol.2020.12377
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
- Yue C, Ma H, Zhou Y. Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma. *PeerJ*. 2019;7:e8128. doi:10.7717/peerj.8128
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27
- Harris MA, Clark J, Ireland A, et al; Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(database issue):D258-261.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118

14. Davis ME. Epidemiology and overview of gliomas. *Semin Oncol Nurs*. 2018;34(5):420-429. doi:10.1016/j.soncn.2018.10.001
15. Filbin MG, Suvà ML. Gliomas genomics and epigenomics: arriving at the start and knowing it for the first time. *Annu Rev Pathol*. 2016;11(1):497-521. doi:10.1146/annurev-pathol-012615-044208
16. Zhou L, Tang H, Wang F, et al. Bioinformatics analyses of significant genes, related pathways and candidate prognostic biomarkers in glioblastoma. *Mol Med Rep*. 2018;18(5):4185-4196. doi:10.3892/mmr.2018.9411
17. Katsanis N, Venable S, Smith JR, Lupski JR. Isolation of a paralog of the Doyme honeycomb retinal dystrophy gene from the multiple retinopathy critical region on 11q13. *Hum Genet*. 2000;106(1):66-72. doi:10.1007/s004399900224
18. Wang L, Chen Q, Chen Z, et al. EFEMP2 is upregulated in gliomas and promotes glioma cell proliferation and invasion. *Int J Clin Exp Pathol*. 2015;8(9):10385-10393.
19. Huang L, Wang Z, Chang Y, et al. EFEMP2 indicates assembly of M0 macrophage and more malignant phenotypes of glioma. *Aging (Albany NY)*. 2020;12(9):8397-8412. doi:10.18632/aging.103147
20. Toledo PL, Torkko JM, Müller A, et al. ICA512 RESP18 homology domain is a protein-condensing factor and insulin fibrillation inhibitor. *J Biol Chem*. 2019;294(21):8564-8576. doi:10.1074/jbc.RA119.007607
21. Song X, Jiao X, Yan H, et al. Overexpression of PTPRN promotes metastasis of lung adenocarcinoma and suppresses NK cell cytotoxicity. *Front Cell Dev Biol*. 2021;9:622018. doi:10.3389/fcell.2021.622018
22. Oliveira DVNP, Prahm KP, Christensen IJ, Hansen A, Høgdall CK, Høgdall EV. Gene expression profile association with poor prognosis in epithelial ovarian cancer patients. *Sci Rep*. 2021;11(1):5438. doi:10.1038/s41598-021-84953-9
23. Dimitrieva S, Schlapbach R, Rehrauer H. Prognostic value of cross-omics screening for kidney clear cell renal cancer survival. *Biol Direct*. 2016;11(1):68. doi:10.1186/s13062-016-0170-1