

ORIGINAL ARTICLE

A four-gene signature-derived risk score for glioblastoma: prospects for prognostic and response predictive analyses

Mianfu Cao¹, Juan Cai², Ye Yuan¹, Yu Shi¹, Hong Wu¹, Qing Liu¹, Yueliang Yao¹, Lu Chen¹, Weiqi Dang¹, Xiang Zhang¹, Jingfang Xiao¹, Kaidi Yang¹, Zhicheng He¹, Xiaohong Yao¹, Yonghong Cui¹, Xia Zhang¹, Xiuwu Bian¹

¹Institute of Pathology and Southwest Cancer Center, Southwest Hospital, Third Military Medical University (Army Medical University), Key Laboratory of Tumor Immunopathology, Ministry of Education of China, Chongqing 400038, China;

²Department of Kidney, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing 400038, China

ABSTRACT

Objective: Glioblastoma (GBM) is the most common primary malignant brain tumor regulated by numerous genes, with poor survival outcomes and unsatisfactory response to therapy. Therefore, a robust, multi-gene signature-derived model is required to predict the prognosis and treatment response in GBM.

Methods: Gene expression data of GBM from TCGA and GEO datasets were used to identify differentially expressed genes (DEGs) through DESeq2 or LIMMA methods. The DEGs were then overlapped and used for survival analysis by univariate and multivariate COX regression. Based on the gene signature of multiple survival-associated DEGs, a risk score model was established, and its prognostic and predictive role was estimated through Kaplan-Meier analysis and log-rank test. Gene set enrichment analysis (GSEA) was conducted to explore high-risk score-associated pathways. Western blot was used for protein detection.

Results: Four survival-associated DEGs of GBM were identified: OSMR, HOXC10, SCARA3, and SLC39A10. The four-gene signature-derived risk score was higher in GBM than in normal brain tissues. GBM patients with a high-risk score had poor survival outcomes. The high-risk group treated with temozolomide chemotherapy or radiotherapy survived for a shorter duration than the low-risk group. GSEA showed that the high-risk score was enriched with pathways such as vasculature development and cell adhesion. Western blot confirmed that the proteins of these four genes were differentially expressed in GBM cells.

Conclusions: The four-gene signature-derived risk score functions well in predicting the prognosis and treatment response in GBM and will be useful for guiding therapeutic strategies for GBM patients.

KEYWORDS

Differentially expressed genes; gene set enrichment analysis; glioblastoma prognosis; radiotherapy; temozolomide chemotherapy

Introduction

Glioblastoma (GBM), which belongs to the World Health Organization (WHO) grade IV glioma, is the most common and aggressive primary malignant brain tumor^{1,2}. Despite advances in surgical resection, radiotherapy, and chemotherapy, the prognosis of GBM patients remains dismal, with a median survival time of less than 15 months^{3,4}. This poor outcome is attributable to malignant progression, drug resistance, and tumor recurrence, which are regulated by numerous oncogenes and tumor suppressor genes^{5,6}. Therefore, a multiple-gene signature derived model is

warranted for predicting the prognosis and treatment response more accurately and for more efficient management of GBM.

Recently, data-mining and bioinformatics analysis has been a powerful and promising tool in cancer research. A lot of high-throughput data generated by microarrays and next-generation sequencing are collected in public databases, of which The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) are the most widely used. Mining these data can be very helpful in exploring tumor expression characteristics and identifying prognostic indicators and predictive biomarkers⁷⁻⁹.

In the present study, we aimed to conduct a combination analysis on the gene expression data of GBM from TCGA and GEO to identify the survival-associated differentially expressed genes (DEGs), thereby establishing a risk model based on the gene signature of these identified genes for prognosis and treatment response prediction. We identified

Correspondence to: Xia Zhang and Xiuwu Bian

E-mail: zhangxia45@yahoo.com and bianxiuwu@263.net

Received August 22, 2018; accepted December 10, 2018.

Available at www.cancerbiomed.org

Copyright © 2019 by Cancer Biology & Medicine

four aberrantly expressed genes in GBM that were significantly associated with survival. The risk score model based on the four-gene signature functioned well for estimating the prognosis and treatment response, and the high-risk score was enriched with multiple tumor-associated pathways. Our results demonstrate that the four-gene signature-derived risk score model can predict the prognosis and treatment response in GBM patients.

Materials and methods

Acquisition of gene expression data of GBM from TCGA and GEO databases

The non-normalized gene expression data of TCGA GBM generated by the Illumina HiSeq 2000 RNA sequencing platform, shown as counts, which were required for the subsequent DEseq2 analysis¹⁰, were downloaded from the TCGA official website (<https://cancergenome.nih.gov/>), while the corresponding normalized data, shown as $\log_2(x+1)$ transformed RSEM normalized counts, which were used for correlation analysis, survival analysis, and gene set enrichment analysis, were obtained from the TCGA hub in UCSC Xena (<https://tcga.xenahubs.net>). The GBM samples from TCGA after removing those without survival time formed the 165 cases for this study (**Supplementary Table S1**). Use of the TCGA data adhered to TCGA publication guidelines and policies. The microarray data of GEO GBM (GSE4290, GSE16011, GSE43378, GSE83300 and GSE90604) were acquired from the GEO official website (<http://www.ncbi.nlm.nih.gov/geo>) and normalized using the RMA method^{11,12}, whereas the normalized RNA sequencing data of GSE59612 were directly downloaded from the GEO website. These normalized GEO data were used for LIMMA analysis and survival analysis. The clinical information of GBM patients from TCGA database and GEO database is summarized in **Supplementary Table S2** and **Table S3** respectively.

Exploration of DEGs of GBM

For mRNA sequencing data of GBM from TCGA database, the R language package DEseq2 was used to identify DEGs between GBM and normal brain tissue¹⁰. For the mRNA microarray data of GBM from GEO database, the R language package LIMMA was applied to explore DEGs in GBM compared to the normal brain tissue¹³. Genes with adjusted *p* values below 0.05 and the absolute value of fold change above 2 were referred as DEGs in GBM. To search for the common

and robust DEGs of GBM, we overlapped the upregulated genes in the TCGA GBM cohort, GSE4290 and GSE16011, and also overlapped the downregulated genes in these cohorts.

Survival analysis of the overlapped DEGs in GBM

To evaluate whether the overlapped DEGs were associated with survival, univariate Cox proportional hazards regression analysis of each of them for overall survival in TCGA GBM cohort was conducted. Subsequently, the upregulated genes inversely correlated with survival (*P* < 0.01) and the downregulated genes positively correlated with survival (*P* < 0.01) were further analyzed through multivariate Cox proportional hazards regression to identify the key DEGs associated with survival.

Establishment of the risk score model for prognosis and treatment response prediction

Based on the gene signature of survival-associated DEGs identified via multivariate Cox proportional hazards regression analysis, a risk score model was established with the following formula.

$$\text{Risk score} = \sum_{i=1}^n \beta_i * E_i$$

In the above equation, “*n*” is the total number of key genes; “ β_i ” represents the regression coefficient of gene *i*; and “ E_i ” represents the expression value of gene *i*. In the validation GEO datasets, “ β_i ” used was the same as that in the TCGA cohort.

To explore the prognostic and predictive role of risk score, GBM patients were divided into high-risk and low-risk groups according to the cutoff of risk score determined by the receiver operating characteristic curve analysis. Then the two groups in all GBM patients, or the patients with therapy were analyzed for survival using the Kaplan-Meier method with the log-rank test. Univariate and multivariate Cox regression analysis of clinicopathological parameters and risk score were conducted to confirm the clinical significance of the risk score.

Gene set enrichment analysis (GSEA)

GSEA was conducted to explore the high-risk score associated signaling pathways on GSEA software downloaded from the Broad Institute (<http://www.broadinstitute.org/gsea>)¹⁴. The gene expression data of the high-risk and

low-risk groups in TCGA GBM cohort together with the gene sets in Hallmark, and C2.CP and C5.BP MSigDB datasets from the Broad Institute were imported into the GSEA software. By running GSEA, normalized enrichment scores and *p*-value were generated. The Cytoscape software with the Enrichment map plug-in was used for visualizing the GSEA results¹⁵.

Cell culture

The normal human glial cell line HEB was kindly provided by Prof. Guang-Mei Yan (Department of Pharmacology, Sun Yat-Sen University, Guangzhou, China)¹⁶. Human GBM cell line LN229 was from the American Type Culture Collection (ATCC) (Manassas, Virginia) and authenticated by short tandem repeat testing (Beijing Microread Genetics Co., Ltd., Beijing, China)¹⁷. Primary human GBM cells, GBM-1 and GBM-2, were established using two surgical specimens from GBM patients at the Department of Neurosurgery, Southwest Hospital with written informed consent¹⁸, which was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Southwest Hospital. The information of the two GBM specimens is summarized in **Supplementary Table S4**. For primary GBM cell culture, the specimens were cut into $1 \times 1 \times 1 \text{ mm}^3$ pieces and cultured in 0.2 ml FBS (Hyclone, GE Healthcare Bio-Sciences, Pittsburgh, Pennsylvania) in a 60 mm dish inverted for 1 h. The dish was then turned the right side up and supplemented with 5 ml DMEM (Hyclone) containing 10% FBS (Hyclone). The primary GBM cells were detached with 0.25% trypsin upon formation of a monolayer of cells. All the cells were maintained in DMEM (Hyclone) supplemented with 10% FBS (Hyclone) without mycoplasma contamination, and incubated at 37°C in a humidified incubator with 5% CO₂/95% air.

Western blot

Western blot analysis was carried out as described previously¹⁶. Antibodies used in western blot are listed as follows: anti-OSMR (Abcam, ab210771, 1 : 1000) (Cambridge, United Kingdom), anti-HOXC10 (Abcam, ab153904, 1 : 1000), anti-SCARA3 (Abcam, ab96205, 1 : 000), anti-SLC39A10 (Abcam, ab83947, 1 : 1000), anti-β-Actin (Cell Signaling Technology, #8457, 1 : 1000) (Danvers, Massachusetts), and horseradish peroxidase-conjugated anti-rabbit IgG (Cell Signaling Technology, #7074, 1 : 5000). The relative expression of target proteins was quantified in comparison with β-Actin using Image Lab software (Bio-Rad

Laboratories, Berkeley, California) and was then normalized to the corresponding expression in HEB cells.

Statistical analysis

Unpaired Student's *t*-test and Mann–Whitney U test were used to compare two groups, whereas one-way ANOVA and Kruskal-Wallis H test were used to compare multiple groups. Pearson Chi-Square test and Fisher's exact test were conducted to analyze the correlation between risk score and clinicopathological parameters. Experiments were conducted independently three times and data were presented as mean ± SEM. *P* < 0.05 (*), *P* < 0.01 (**), and *P* < 0.001 (***) were considered statistically significant. All statistical analyses were performed on SPSS 19.0 (IBM, Armonk, New York), Graphpad Prism 6.0 (GraphPad Software, La Jolla, California), or R software (www.r-project.org).

Results

Multiple differentially expressed genes in GBM

Three GBM datasets (TCGA, GSE4290 and GSE16011) were used to identify the DEGs in GBM compared to normal brain tissues. As shown in **Figure 1A**, there were 3881 upregulated genes and 3908 downregulated genes in TCGA GBM dataset, 1048 upregulated genes and 1443 downregulated genes in the GSE4290 dataset, and 1705 highly expressed genes and 1440 low expressed genes in the GSE16011 cohort (**Figure 1A**). Overlapping the DEGs from three GBM cohorts resulted in 483 upregulated genes and 765 downregulated genes (**Figure 1B**), which could clearly discriminate GBM and normal brain tissues (**Figure 1C**). Collectively, these data demonstrate that there are multiple DEGs in GBM compared to normal brain tissue.

Survival analysis of the overlapped DEGs

To evaluate the prognostic significance of the overlapped DEGs in GBM, univariate Cox regression analysis of each for overall survival was conducted. As shown in **Table 1**, 15 upregulated genes inversely correlated with survival and 3 downregulated genes positively correlated with survival were identified (*P* < 0.01) (**Table 1**). Subsequently, multivariate stepwise Cox regression analysis of these 18 genes was applied and 4 genes were eventually proved to be correlated with overall survival (**Table 1**). These 4 genes are oncostatin M receptor (OSMR), homeobox C10 (HOXC10), scavenger receptor class A member 3 (SCARA3), and solute

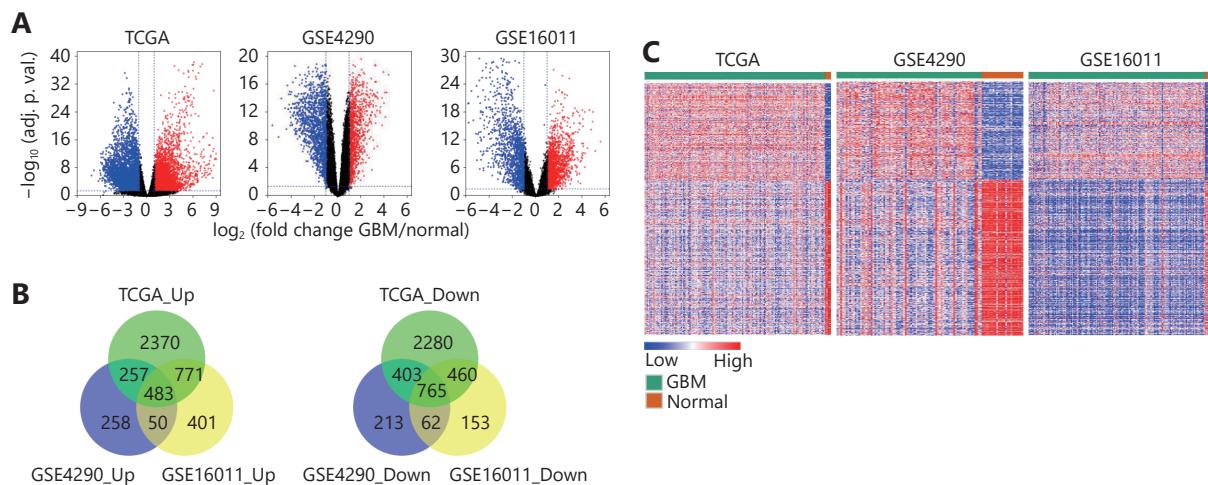


Figure 1 Differentially expressed genes between GBM and normal brain tissues. (A) Volcano plots showing the \log_2 (fold change) of mRNA in GBM compared to normal brain tissues, and the corresponding $-\log_{10}$ (adjusted P value) in TCGA, GSE4290 and GSE16011 datasets. Genes with adjusted P value below 0.05 and fold change above 2 (below -2) were marked with red (blue) dots. (B) Venn diagrams showing the gene numbers of the upregulated genes (left) and the downregulated genes (right) of GBM in TCGA, GSE4290 and GSE16011 cohorts. (C) Heatmaps of the overlapped genes in TCGA, GSE4290 and GSE16011 datasets. The overlapped genes that are from **Figure 1B** include 483 upregulated genes and 765 downregulated genes.

Table 1 Univariate and multivariate Cox regression analysis of the overlapped DEGs of GBM

Gene name	Up/Down	Univariate analysis			Multivariate analysis		
		β	HR (95%CI)	P	β	HR (95%CI)	P
OSMR	UP	0.312	1.366 (1.181, 1.581)	< 0.001	0.279	1.321 (1.136, 1.537)	< 0.001
PDIA4	UP	0.511	1.667 (1.246, 2.231)	0.001			
GJB2	UP	0.151	1.164 (1.063, 1.273)	0.001			
FKBP9	UP	0.320	1.378 (1.135, 1.672)	0.001			
STEAP3	UP	0.222	1.248 (1.086, 1.434)	0.002			
HOXC10	UP	0.100	1.105 (1.037, 1.177)	0.002	0.089	1.093 (1.024, 1.167)	0.008
ISG20	UP	0.297	1.346 (1.112, 1.629)	0.002			
SCARA3	UP	0.337	1.401 (1.127, 1.742)	0.002	0.238	1.268 (1.018, 1.580)	0.034
ZNF540	Down	-0.348	0.706 (0.558, 0.893)	0.004			
IKBIP	UP	0.404	1.498 (1.131, 1.984)	0.005			
KDELc2	UP	0.326	1.385 (1.104, 1.738)	0.005			
SPAG4	UP	0.160	1.174 (1.048, 1.315)	0.006			
C1RL	UP	0.216	1.241 (1.065, 1.446)	0.006			
SLC39A10	Down	-0.494	0.611 (0.430, 0.868)	0.006	-0.424	0.655 (0.445, 0.963)	0.031
GNS	UP	0.456	1.578 (1.140, 2.185)	0.006			
KHDRBS2	Down	-0.147	0.863 (0.775, 0.961)	0.007			
SLC2A10	UP	0.241	1.273 (1.062, 1.525)	0.009			
DENND2A	UP	0.263	1.301 (1.066, 1.588)	0.010			

Abbreviations: β , regression coefficient; HR, hazard ratio; CI, confidence interval.

carrier family 39 member 10 (SLC39A10). Among these, OSMR, HOXC10, and SCARA3 were highly expressed in GBM and inversely correlated with survival ($\beta > 0$, HR > 1), whereas SLC39A10 was lowly expressed and positively correlated with survival ($\beta < 0$, HR < 1) (Table 1). Taken together, these data suggest that the four key DEGs of GBM are significantly correlated with the overall survival of GBM patients.

The four-gene signature-derived risk score model predicts the survival and treatment response in TCGA GBM cohort

As GBM is regulated by multiple genes, we established a risk score model based on the gene expression signature of the four key DEGs to predict survival. The calculation formula of risk score is listed as follows: risk score = $\beta_{OSMR} * E_{OSMR} + \beta_{HOXC10} * E_{HOXC10} + \beta_{SCARA3} * E_{SCARA3} + \beta_{SLC39A10} * E_{SLC39A10}$. “ β ” is the regression coefficient derived from multivariate COX stepwise regression analysis based on TCGA GBM cohort and “E” represents the expression value of the corresponding

gene (Figure 2A). Compared to normal brain tissue, the risk score was higher in GBM (Figure 2B). Analysis of the correlation between the risk score and clinicopathological characteristics showed that a high-risk score was positively correlated with IDH-wild type (Figure 2A and 2C, Supplementary Table S5), and mesenchymal subtype (Figure 2A and 2D, Supplementary Table S5). The IDH-wild type or mesenchymal subtype GBM was associated with unfavorable outcome (Figure S1), which was also demonstrated in previous studies^{7,19}, suggesting that a high-risk score might be related to poor prognosis.

To confirm the prognostic significance of the risk score, Kaplan-Meier survival analyses of high-risk and low-risk GBM patients were conducted. As shown in Figure 3A, high-risk GBM patients survived shorter compared to low-risk patients (Figure 3A). Survival analysis by a combination of risk score and IDH status showed that GBM patients with high-risk score and IDH-wild type had the worst outcomes (Figure 3B), and survival analysis by a combination of risk score and expression subtypes revealed that GBM patients in the mesenchymal subgroup (or in non-mesenchymal

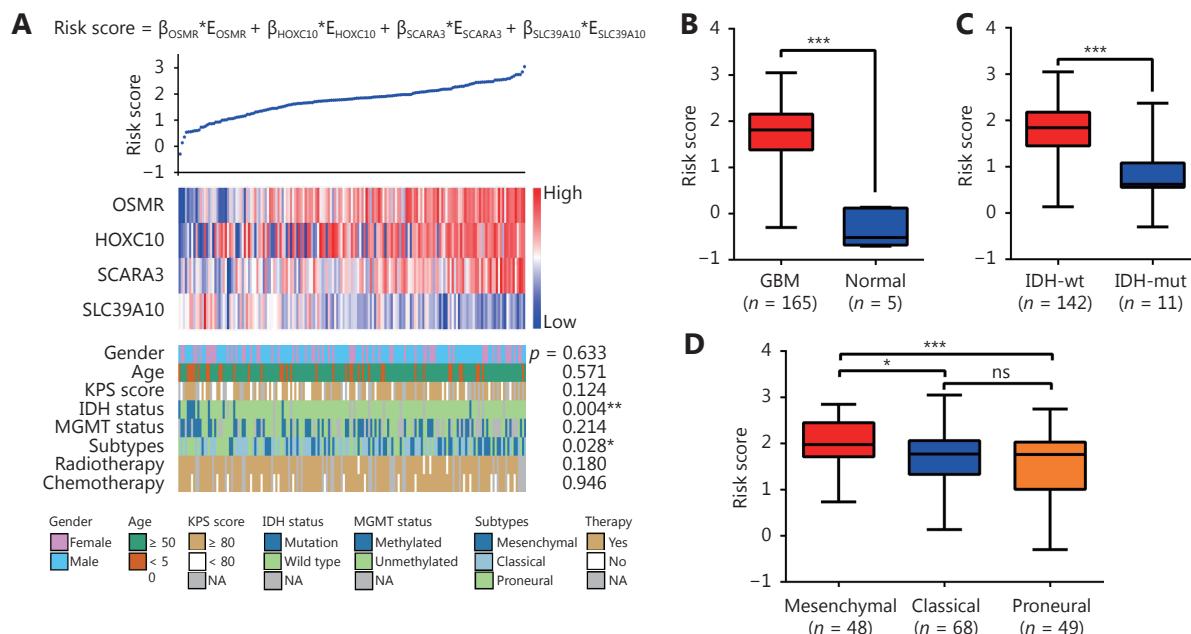


Figure 2 The four-gene signature-derived risk score is high in GBM. (A) The calculation formula and the value of risk score (top), the corresponding expression of four key genes (middle), and the associated clinicopathological parameters (bottom). “ β ” indicates the regression coefficient derived from multivariate COX stepwise regression in TCGA GBM cohort; “E” represents the expression value of the corresponding gene. The P-value indicates the correlation between the risk score and the clinicopathological parameters (Supplementary Table S5). KPS, Karnofsky performance status; IDH, isocitrate dehydrogenase; MGMT, O(6)-methylguanine-DNA methyltransferase; NA, not available. (B-D) Risk scores in GBM and normal brain tissues (B), in IDH-wt GBM and IDH-mut GBM (C), and in mesenchymal GBM and non-mesenchymal (proneural and classical) GBM (D) in TCGA GBM cohort. IDH, isocitrate dehydrogenase; IDH-wt, IDH-wild type; IDH-mut, IDH-mutation. Data are shown as mean \pm SEM, *P < 0.05, ***P < 0.001, ns, not significant.

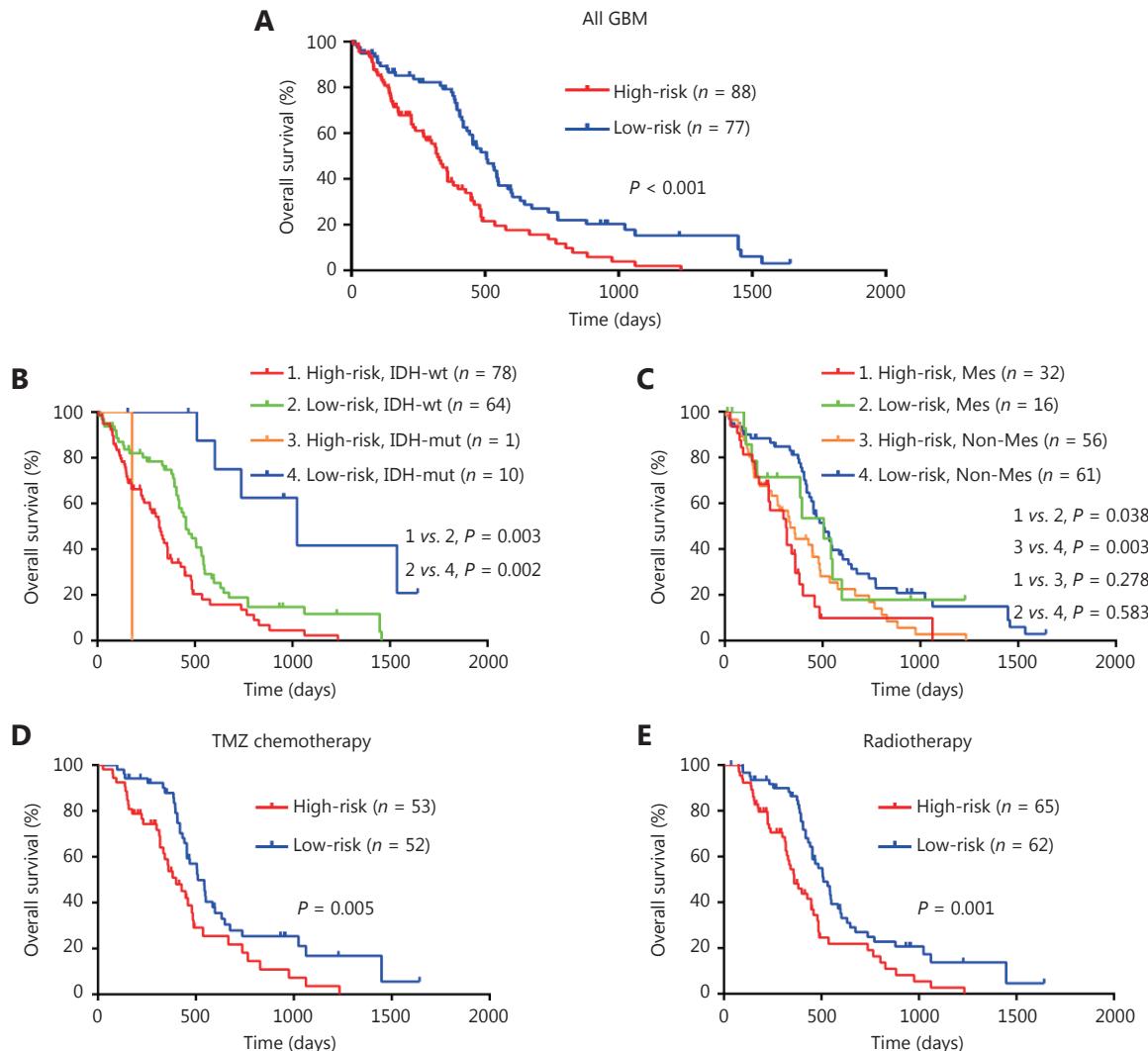


Figure 3 Risk score predicts the prognosis and treatment response in TCGA GBM cohort. (A-C) Kaplan–Meier overall survival analysis among TCGA GBM patients stratified by risk score only (A), combined with IDH status (B), and combined with expression subtypes (C). Mes, mesenchymal; Non-Mes, non-mesenchymal, including proneural and classical subtypes. (D-E) Kaplan–Meier overall survival analysis of TCGA GBM patients with TMZ chemotherapy (D), or radiotherapy (E) according to the risk score. TMZ, temozolomide.

subgroup) with a high-risk score had a shorter survival time than those in the mesenchymal subgroup (or in non-mesenchymal subgroup) with a low-risk score (Figure 3C). The prognostic significance of the risk score was further validated through univariate and multivariate analysis (Supplementary Table S6). Furthermore, high-risk score was also correlated with poor outcome in GBM patients treated with temozolomide chemotherapy or radiotherapy (Figure 3D and 3E), suggesting that the risk score could predict the treatment response. Taken together, these data demonstrate that the risk score model can predict the prognosis as well as the response to therapy in GBM patients.

Risk score model predicts the prognosis and treatment response among independent GEO cohorts of GBM

To further validate the prognostic and predictive role of risk score, the value of the risk score in GBM compared to the normal brain tissue was evaluated and Kaplan–Meier survival analysis was conducted in independent GEO datasets of GBM. As shown in Figure 4A and 4B, the risk score in GBM was higher than that in normal brain tissues (Figure 4A), and GBM patients with a high-risk score had a shorter survival time (Figure 4B). High-risk score was also correlated with poor outcome in GBM patients who underwent radiotherapy

(Figure 4C). Combination analysis of the risk score and expression subtypes revealed that the risk score was the highest in mesenchymal GBM, and that mesenchymal GBM patients with a high-risk score had a shorter survival time compared to those with a low-risk score (Figure 4D).

Combination analysis of the risk score and IDH status showed that GBM patients with IDH-mutation and a high-risk score had a poor outcome compared to those with IDH-mutation and a low-risk score, whereas there was no statistically significant difference in survival between GBM

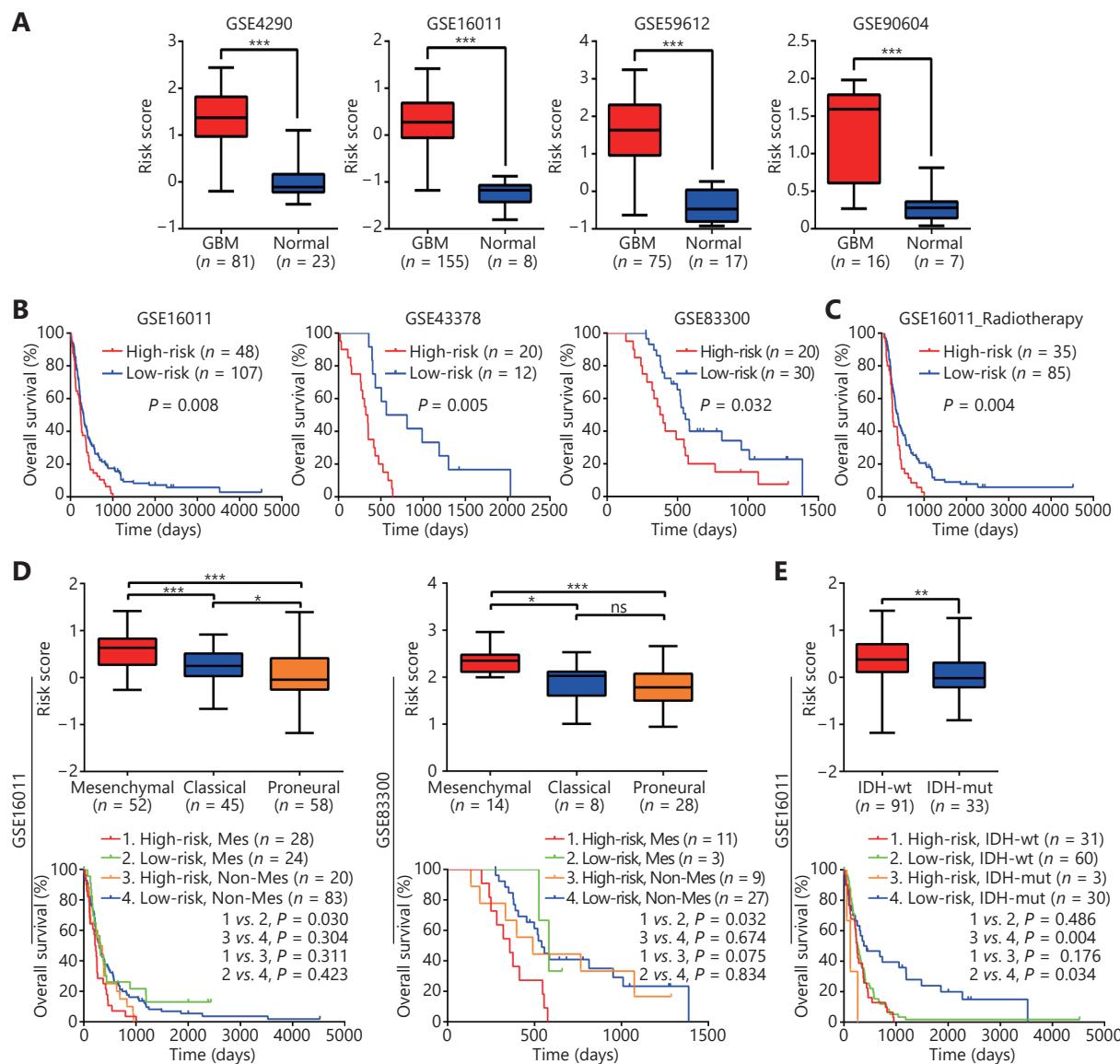


Figure 4 Performance of risk score in predicting the survival and treatment response in independent GEO GBM cohorts. (A) Comparison of risk scores between GBM and normal brain tissues in GSE4290, GSE16011, GSE59612 and GSE90604. (B) Kaplan-Meier overall survival analysis of high-risk and low-risk GBM patients in GSE16011, GSE43378 and GSE83300. (C) Kaplan-Meier overall survival analysis of GBM patients with radiotherapy in GSE16011 according to the risk score. (D) Risk scores in mesenchymal GBM and non-mesenchymal (proneural and classical) GBM (up), and Kaplan-Meier overall survival analysis of GBM patients stratified by risk scores combined with expression subtypes (down). Mes, mesenchymal; Non-Mes, non-mesenchymal, including proneural and classical subtypes. (E) Risk scores in IDH-wt GBM and IDH-mut GBM (up), and Kaplan-Meier overall survival analysis of GBM patients stratified by risk scores combined with IDH status (down). IDH, isocitrate dehydrogenase; IDH-wt, IDH-wild type; IDH-mut, IDH-mutation. Data are shown as mean \pm SEM, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns, not significant.

patients in the IDH-wild type subgroup with a high-risk score and those in the IDH-wild type subgroup with a low-risk score (**Figure 4E**), possibly due to the small sample size and tumor heterogeneity. Taken together, these data confirm that the risk score model can be applied to predict the prognosis and treatment response of GBM patients.

The high-risk score is enriched with multiple tumor-associated pathways

The four-gene signature-derived risk score in prognosis and treatment response prediction may be closely correlated with some important signaling pathways. Thus, we conducted GSEA analysis among high-risk and low-risk patients to identify the pathways associated with the four-gene signature. As shown in **Figure 5**, multiple tumor-associated pathways such as vasculature development, cell adhesion, response to oxidative stress, intracellular signal transduction, and so on, were enriched in GBM patients with a high-risk score (**Figure 5**).

Validation of the differential expression of the four genes in GBM cells

To further validate the differential expression of the four genes (OSMR, HOXC10, SCARA3 and SLC39A10) in GBM, we conducted Western blot analysis to detect the

corresponding proteins of these four genes in a GBM cell line, primary GBM cells, and in a normal glial cell line. As shown in **Figure 6**, OSMR, HOXC10, SCARA3 were highly expressed by GBM cells compared to the normal glial cell line, whereas SLC39A10 was lowly expressed (**Figure 6**). Taken together, these data confirm that the four genes are differentially expressed by GBM cells.

Discussion

GBM is characterized with a complex molecular background^{5,20}, and abundant high-throughput data regarding GBM facilitate exploration of prognostic and predictive molecular biomarkers. In the current study, by combination analysis of multiple gene expression data of GBM from TCGA and GEO cohorts, we uncovered four key genes (OSMR, HOXC10, SCARA3, and SLC39A10) that are aberrantly expressed in GBM and are also associated with survival. The four-gene signature-derived risk score model functions well in predicting the prognosis and treatment response of GBM patients. Moreover, the high-risk score is enriched with multiple tumor-associated pathways.

The initial step of our study was to screen out the DEGs of GBM by combination analysis of gene expression data from both TCGA and GEO datasets, rather than based on just one GBM dataset. In this way, the discovered genes (483 up-regulated and 765 down-regulated) might be stably and

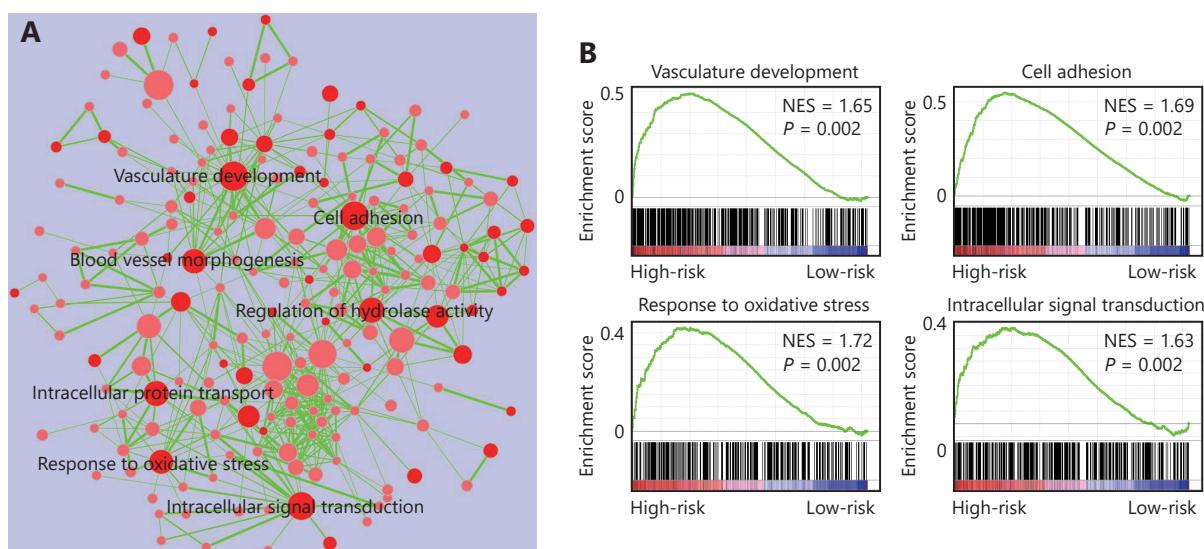


Figure 5 Four-gene signature associated biological pathways in high-risk GBM. (A) Enrichment map showing the pathways enriched in high-risk GBM through GSEA analysis. Nodes represent enriched gene sets with P -value below 0.05. Node with deep red color correlates with small P -value. Node size corresponds to the number of genes within gene set. Edge thickness corresponds to the number of shared genes between gene sets. (B) Representative enriched pathways in high-risk GBM through GSEA analysis. NES, normalized enrichment score.

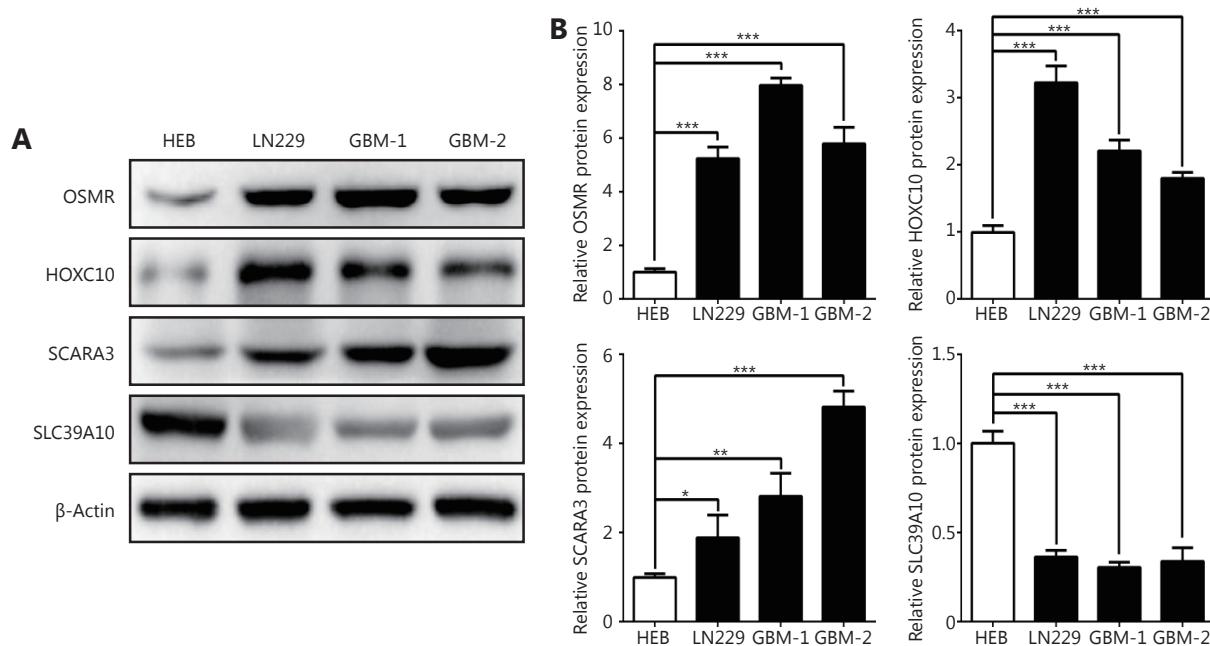


Figure 6 Validation of the differential expression of the four genes in GBM cells. (A and B) Western blot images (A) and the relevant quantification (B) of OSMR, HOXC10, SCARA3 and SLC39A10 in GBM cell line LN229, primary GBM cells (GBM-1 and GBM-2), and normal glial cell line HEB. The relative expression of target proteins is quantified in comparison with β -Actin and normalized to the corresponding expression in HEB cells. Data are shown as mean \pm SEM from three independent experiments, * P < 0.05, ** P < 0.01, *** P < 0.001.

specifically expressed in GBM compared to normal brain tissue. We then conducted univariate and multivariate Cox regression analysis of the DEGs and found four important genes associated with survival, *i.e.*, OSMR, HOXC10, SCARA3, and SLC39A10. Genes that are upregulated and inversely correlated with survival (OSMR, HOXC10, and SCARA3) may function as oncogenes, whereas SLC39A10, the gene that is downregulated and positively correlated with survival, may function as a GBM suppressor gene. Moreover, the proteins of these four genes were also differentially expressed in GBM cells as validated by western blot. Thus, the risk score model was established based on the signature of these four DEGs. According to previous studies, gene signatures comprising the expression of multiple genes are useful in reducing the complexity of tumor data, and are more static, robust, and reproducible in prognosis prediction, whereas the expression level of an individual gene may change in a very short period²¹. In our study, the four-gene signature-derived risk score could predict the survival outcome of GBM patients from TCGA cohort, and its prognostic role was reproducible in independent GEO datasets of GBM. Additionally, the risk score model could also predict the response of GBM patients to radiotherapy or chemotherapy. Thus, we established a promising multi-gene

signature-derived risk score model that can be used as a prognostic and predictive biomarker in GBM.

Among the four genes, OSMR is a member of the interleukin 6 receptor family, which transduces oncostatin M (OSM)-induced signaling mainly through the JAK/STAT pathway and plays an important role in hematopoiesis, development, and inflammation as well as in tumor progression²². HOXC10 belongs to the homeobox gene family that encodes a family of highly conserved transcription factors. It is associated with cell proliferation, differentiation, and morphogenesis, and also has protumoral functions in tumor growth, migration, invasion, metastasis, and drug resistance^{23,24}. SCARA3 encodes a macrophage scavenger receptor-like protein, and its expression can be induced by oxidative stress and enables depletion of reactive oxygen species, thus, protecting cells from oxidative stress-induced cell killing²⁵. SLC39A10 is a zinc transporter involved in zinc uptake²⁶, whereas its role in GBM progression is poorly understood. The biological functions of these four genes have partially provided clues for understanding the prognostic and predictive significance of risk score, but the role of these genes in GBM progression and the underlying mechanisms need further investigation.

The prognostic and predictive value of the four-gene

signature-derived risk score may be useful for guiding therapeutic strategies to improve the clinical outcome of GBM patients. Patients with high-risk scores should receive aggressive treatment, whereas those with low-risk scores should avoid additional treatment that could lead to unnecessary toxicities. Therefore, it is important to translate the risk score into clinical practice, though there is a long way to go. For clinical application, prospective clinical studies are warranted to further validate the prognostic and predictive significance of the risk score.

Conclusions

In conclusion, the four-gene signature-derived risk score model functions well in predicting the prognosis and treatment response in GBM patients, with the potential to optimize treatment decisions. Future studies are needed to explore the biological roles of these four genes in GBM progression and to further validate the prognostic and predictive value of the risk score for clinical practice.

Acknowledgements

This study was supported by the National Key R&D Program of China (Grant No. 2016YFA0101203 to XB and 2016YFC1201801 to XZ), the National Natural Science Foundation of China (Grant No. 81372273 and 81773145 to XZ), the funding from Key Laboratory of Tumor Immunology and Pathology (Army Medical University), Ministry of Education of China (Grant No. 2017jszl09 to MC) and the Basic and Applied Fund of First Affiliated Hospital of Army Military Medical University (Grant No. SWH2016BZGFSBJ-04 and SWH2016JCZD-04 to XZ).

Conflict of interest statement

No potential conflicts of interest are disclosed.

References

- Wen PY, Reardon DA. Neuro-oncology in 2015: progress in glioma diagnosis, classification and treatment. *Nat Rev Neurol.* 2016; 12: 69-70.
- Tanaka S, Louis DN, Curry WT, Batchelor TT, Dietrich J. Diagnostic and therapeutic avenues for glioblastoma: no longer a dead end? *Nat Rev Clin Oncol.* 2013; 10: 14-26.
- Malkki H. Neuro-oncology: novel biomarkers for high-grade astrocytoma. *Nat Rev Neurol.* 2014; 10: 121.
- Xu YY, Gao P, Sun Y, Duan YR. Development of targeted therapies in treatment of glioblastoma. *Cancer Biol Med.* 2015; 12: 223-37.
- Filbin MG, Suva ML. Gliomas genomics and epigenomics: arriving at the start and knowing it for the first time. *Annu Rev Pathol.* 2016; 11: 497-521.
- Kim J, Lee IH, Cho HJ, Park CK, Jung YS, Kim Y, et al. Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell.* 2015; 28: 318-28.
- Wang QH, Hu BL, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell.* 2017; 32: 42-56.
- Anwar SL, Wulaningsih W, Watkins J. Profile of the breast cancer susceptibility marker rs4245739 identifies a role for miRNAs. *Cancer Biol Med.* 2017; 14: 387-95.
- Shi Y, Ping YF, Zhou WC, He ZC, Chen C, Bian BSJ, et al. Tumour-associated macrophages secrete pleiotrophin to promote PTPRZ1 signalling in glioblastoma stem cells for Tumour growth. *Nat Commun.* 2017; 8: 15080.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550.
- Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003; 19: 185-93.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4: 249-64.
- Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43: e4.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102: 15545-50.
- Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One.* 2010; 5: e13984.
- Tan J, Liu M, Zhang JY, Yao YL, Wang YX, Lin Y, et al. Capillary morphogenesis protein 2 is a novel prognostic biomarker and plays oncogenic roles in glioma. *J Pathol.* 2018; 245: 160-71.
- Li Y, He ZC, Zhang XN, Liu Q, Chen C, Zhu Z, et al. Stanniocalcin-1 augments stem-like traits of glioblastoma cells through binding and activating NOTCH1. *Cancer Lett.* 2018; 416: 66-74.
- Wang Z, Wang B, Shi Y, Xu C, Xiao HL, Ma LN, et al. Oncogenic miR-20a and miR-106a enhance the invasiveness of human glioma stem cells by directly targeting TIMP-2. *Oncogene.* 2015; 34: 1407-19.
- Yan H, Parsons DW, Jin GL, McLendon R, Rasheed BA, Yuan WS,

- et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med.* 2009; 360: 765-73.
20. Reifenberger G, Wirsching HG, Knobbe-Thomsen CB, Weller M. Advances in the molecular genetics of gliomas - implications for classification and therapy. *Nat Rev Clin Oncol.* 2017; 14: 434-52.
21. Subramanian J, Simon R. What should physicians look for in evaluating prognostic gene-expression signatures? *Nat Rev Clin Oncol.* 2010; 7: 327-34.
22. Caffarel MM, Coleman N. Oncostatin M receptor is a novel therapeutic target in cervical squamous cell carcinoma. *J Pathol.* 2014; 232: 386-90.
23. Pathiraja TN, Nayak SR, Xi YX, Jiang SM, Garee JP, Edwards DP, et al. Epigenetic reprogramming of *HOXC10* in endocrine-resistant breast cancer. *Sci Transl Med.* 2014; 6: 229ra41.
24. Tang XL, Ding BX, Hua Y, Chen H, Wu T, Chen ZQ, et al. *HOXC10* promotes the metastasis of human lung adenocarcinoma and indicates poor survival outcome. *Front Physiol.* 2017; 8: 557.
25. Brown CO, Schibler J, Fitzgerald MP, Singh N, Salem K, Zhan FH, et al. Scavenger receptor class A member 3 (SCARA3) in disease progression and therapy resistance in multiple myeloma. *Leuk Res.* 2013; 37: 963-9.
26. Kaler P, Prasad R. Molecular cloning and functional characterization of novel zinc transporter rZip10 (*Slc39a10*) involved in zinc uptake across rat renal brush-border membrane. *Am J Physiol Renal Physiol.* 2007; 292: F217-29.

Cite this article as: Cao M, Cai J, Yuan Y, Shi Y, Wu H, Liu Q, et al. A four-gene signature-derived risk score for glioblastoma: prospects for prognostic and response predictive analyses. *Cancer Biol Med.* 2019; 16: 595-605. doi: 10.20892/j.issn.2095-3941.2018.0277

Supplementary materials

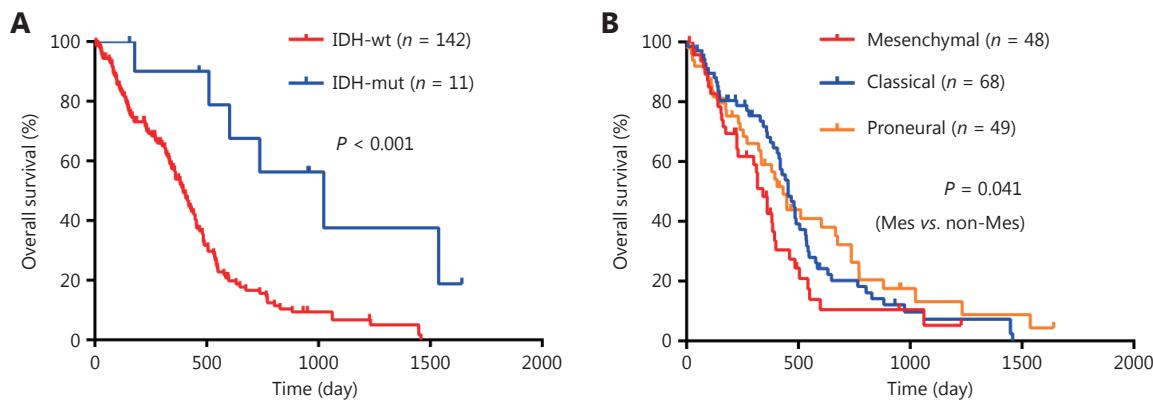


Figure S1 Survival analysis of GBM patients. (A-B) Kaplan-Meier overall survival analysis of GBM patients stratified by IDH status (A), and expression subtypes (B). IDH, isocitrate dehydrogenase; IDH-wt, IDH-wild type; IDH-mut, IDH-mutation. Mes, mesenchymal; Non-Mes, non-mesenchymal, including proneural and classical subtypes.

Table S1 TCGA samples used in the study

Sample ID (normal)	Sample ID (GBM)				
TCGA-06-0675-11	TCGA-02-0047-01	TCGA-06-0646-01	TCGA-06-5856-01	TCGA-19-1389-02	TCGA-28-2513-01
TCGA-06-0678-11	TCGA-02-0055-01	TCGA-06-0649-01	TCGA-06-5858-01	TCGA-19-1390-01	TCGA-28-2514-01
TCGA-06-0680-11	TCGA-02-2483-01	TCGA-06-0686-01	TCGA-06-5859-01	TCGA-19-1787-01	TCGA-28-5204-01
TCGA-06-0681-11	TCGA-02-2485-01	TCGA-06-0743-01	TCGA-08-0386-01	TCGA-19-2619-01	TCGA-28-5207-01
TCGA-06-AABW-11	TCGA-02-2486-01	TCGA-06-0744-01	TCGA-12-0616-01	TCGA-19-2620-01	TCGA-28-5208-01
	TCGA-06-0125-01	TCGA-06-0745-01	TCGA-12-0618-01	TCGA-19-2624-01	TCGA-28-5209-01
	TCGA-06-0125-02	TCGA-06-0747-01	TCGA-12-0619-01	TCGA-19-2625-01	TCGA-28-5213-01
	TCGA-06-0129-01	TCGA-06-0749-01	TCGA-12-0821-01	TCGA-19-2629-01	TCGA-28-5215-01
	TCGA-06-0130-01	TCGA-06-0750-01	TCGA-12-1597-01	TCGA-19-4065-01	TCGA-28-5216-01
	TCGA-06-0132-01	TCGA-06-0878-01	TCGA-12-3650-01	TCGA-19-4065-02	TCGA-28-5218-01
	TCGA-06-0138-01	TCGA-06-0882-01	TCGA-12-3652-01	TCGA-19-5960-01	TCGA-28-5220-01
	TCGA-06-0141-01	TCGA-06-1804-01	TCGA-12-3653-01	TCGA-26-1442-01	TCGA-32-1970-01
	TCGA-06-0152-02	TCGA-06-2557-01	TCGA-12-5295-01	TCGA-26-5132-01	TCGA-32-1980-01
	TCGA-06-0156-01	TCGA-06-2558-01	TCGA-12-5299-01	TCGA-26-5133-01	TCGA-32-1982-01
TCGA-06-0157-01	TCGA-06-2559-01	TCGA-14-0736-02	TCGA-26-5134-01	TCGA-32-2615-01	
TCGA-06-0158-01	TCGA-06-2561-01	TCGA-14-0781-01	TCGA-26-5135-01	TCGA-32-2616-01	
TCGA-06-0168-01	TCGA-06-2562-01	TCGA-14-0787-01	TCGA-26-5136-01	TCGA-32-2632-01	
TCGA-06-0171-02	TCGA-06-2563-01	TCGA-14-0789-01	TCGA-26-5139-01	TCGA-32-2634-01	
TCGA-06-0174-01	TCGA-06-2564-01	TCGA-14-0790-01	TCGA-27-1830-01	TCGA-32-2638-01	
TCGA-06-0178-01	TCGA-06-2565-01	TCGA-14-0817-01	TCGA-27-1831-01	TCGA-32-4213-01	
TCGA-06-0184-01	TCGA-06-2567-01	TCGA-14-0871-01	TCGA-27-1832-01	TCGA-32-5222-01	
TCGA-06-0187-01	TCGA-06-2569-01	TCGA-14-1034-01	TCGA-27-1834-01	TCGA-41-2571-01	
TCGA-06-0190-01	TCGA-06-2570-01	TCGA-14-1034-02	TCGA-27-1835-01	TCGA-41-2572-01	
TCGA-06-0190-02	TCGA-06-5408-01	TCGA-14-1402-02	TCGA-27-1837-01	TCGA-41-3915-01	
TCGA-06-0210-01	TCGA-06-5410-01	TCGA-14-1823-01	TCGA-27-2519-01	TCGA-41-4097-01	
TCGA-06-0210-02	TCGA-06-5411-01	TCGA-14-1825-01	TCGA-27-2521-01	TCGA-41-5651-01	
TCGA-06-0211-01	TCGA-06-5412-01	TCGA-14-1829-01	TCGA-27-2523-01	TCGA-76-4925-01	
TCGA-06-0211-02	TCGA-06-5413-01	TCGA-14-2554-01	TCGA-27-2524-01	TCGA-76-4926-01	
TCGA-06-0219-01	TCGA-06-5414-01	TCGA-15-0742-01	TCGA-27-2526-01	TCGA-76-4927-01	
TCGA-06-0221-02	TCGA-06-5415-01	TCGA-15-1444-01	TCGA-27-2528-01	TCGA-76-4928-01	
TCGA-06-0238-01	TCGA-06-5416-01	TCGA-16-0846-01	TCGA-28-1747-01	TCGA-76-4929-01	
TCGA-06-0644-01	TCGA-06-5417-01	TCGA-16-1045-01	TCGA-28-1753-01	TCGA-76-4931-01	
TCGA-06-0645-01	TCGA-06-5418-01	TCGA-19-0957-02	TCGA-28-2509-01	TCGA-76-4932-01	

Table S2 The clinical characteristics of GBM patients in TCGA cohort

Characteristic	GBM patients (<i>n</i> = 165)
Gender	
Male	106
Female	59
Age	
≥50	128
<50	37
Grade	
I-III	0
IV	165
KPS score	
≥80	88
<80	36
Missing	41
IDH status	
Mutation	11
Wild type	142
Missing	12
MGMT status	
Methylated	58
Unmethylated	67
Missing	40
Expression subtypes	
Mesenchymal	48
Classical	68
Proneural	49
Radiotherapy	
Yes	127
No	4
Missing	34
Chemotherapy	
Yes	112
No	19
Missing	34

Table S3 The clinical information of GBM patients from GEO database

Characteristic	GSE4290	GSE16011	GSE43378	GSE59612	GSE83300	GSE90604
Normal tissues	23	8	0	17	0	7
GBM cases	81	159	32	75	50	16
GBM_gender						
Male		108	20	44	24	10
Female		51	12	31	25	6
Missing	81	0	0	0	1	0
GBM_age						
≥50		NA	107	22	67	19
<50		NA	52	10	8	31
GBM_grade						
I-III		0	0	0	0	0
IV	81	155	32	75	50	16
GBM_cases with survival	NA	155	32	NA	50	NA

NA, not available.

Notes:

- (1) All the GBM cases in the above GEO datasets were included in our study except for 4 of 159 GBM cases in GSE16011 without survival information.
- (2) The glioma samples but not GBM samples in GSE4290 (26 astrocytoma and 50 oligodendrogloma), GSE16011 (29 astrocytoma, 8 pilocytic astrocytoma, 52 oligodendrogloma and 28 mixed glioma) and GSE43378 (5 astrocytoma and 13 anaplastic oligodendrogloma) were not included in our study.
- (3) In GSE90604, the data from 7 normal brain tissues were included while the data from 2 normal astrocyte cell lines were excluded in our study.

Table S4 The clinical information of GBM specimens used for establishment of primary GBM cells

Cases	Primary cells	Gender	Age	Tumor location	WHO grade	Histology	IDH status
Case 1	GBM-1	Female	22	Right temporal, parietal and occipital lobe	IV	Glioblastoma	Wild type
Case 2	GBM-2	Female	37	Left temporal and parietal lobe	IV	Glioblastoma	Wild type

Table S5 Correlation between the risk score and clinicopathological parameters in TCGA GBM cohort

Factor	Risk score		<i>P</i> value
	Low (<i>n</i> = 77)	High (<i>n</i> = 88)	
Gender			0.633
Male	48 (62.3%)	58 (65.9%)	
Female	29 (37.7%)	30 (34.1%)	
Age			0.517
≥50	58 (75.3%)	70 (79.5%)	
<50	19 (24.7%)	18 (20.5%)	
KPS score			0.124
≥80	39 (50.6%)	49 (55.7%)	
<80	22 (28.6%)	14 (15.9%)	
Missing	16 (20.8%)	25 (28.4%)	
IDH status			0.004**
Mutation	10 (13.0%)	1 (1.1%)	
Wild type	64 (83.1%)	78 (88.6%)	
Missing	3 (3.9%)	9 (10.2%)	
MGMT status			0.214
Methylated	28 (36.4%)	30 (34.1%)	
Unmethylated	35 (45.5%)	32 (36.4%)	
Missing	14 (18.2%)	26 (29.5%)	
Expression subtypes			0.028*
Mesenchymal	16 (20.8%)	32 (36.4%)	
Non-mesenchymal	61 (79.2%)	56 (63.6%)	
Radiotherapy			0.180
Yes	62 (80.5%)	65 (73.9%)	
No	0 (0.0%)	4 (4.5%)	
Missing	15 (19.5%)	19 (21.6%)	
Chemotherapy			0.946
Yes	53 (68.8%)	59 (67.0%)	
No	9 (11.7%)	10 (11.4%)	
Missing	15 (19.5%)	19 (21.6%)	

KPS, Karnofsky performance status; IDH, isocitrate dehydrogenase; MGMT, O(6)-methylguanine-DNA methyltransferase. *, *P* < 0.05; **, *P* < 0.01.

Table S6 Univariate and multivariate Cox regression analysis of clinicopathological parameters and risk score for overall survival in TCGA GBM cohort

Factor	Univariate analysis		Multivariate analysis	
	HR (95%CI)	P	HR (95%CI)	P
Gender	0.995 (0.691, 1.432)	0.977		
Age	1.029 (1.013, 1.044)	< 0.001	1.034 (1.013, 1.055)	0.001
KPS score	0.893 (0.568, 1.405)	0.625		
IDH status	0.229 (0.092, 0.568)	0.001		
MGMT status	0.568 (0.368, 0.875)	0.010	0.600 (0.382, 0.941)	0.026
Expression subtypes	1.509 (1.015, 2.245)	0.042		
Radiotherapy	0.467 (0.171, 1.279)	0.138		
Chemotherapy	0.672 (0.396, 1.141)	0.141		
Risk score	2.714 (1.924, 3.829)	< 0.001	2.046 (1.321, 3.168)	0.001

KPS, Karnofsky performance status; IDH, isocitrate dehydrogenase; MGMT, O(6)-methylguanine-DNA methyltransferase; HR, hazard ratio; CI, confidence interval.