



A Novel Six-mRNA Signature Predicts Survival of Patients With Glioblastoma Multiforme

Zhentaio Liu^{1,2†}, Hao Zhang^{3†}, Hongkang Hu^{1†}, Zheng Cai^{1,4}, Chengyin Lu¹, Qiang Liang¹, Jun Qian⁵, Chunhui Wang^{1*} and Lei Jiang^{1*}

¹ Department of Neurosurgery, Changzheng Hospital, Naval Medical University, Shanghai, China, ² Department of Neurosurgery, No. 988 Hospital of Joint Logistic Support Force, Zhengzhou, China, ³ Department of Orthopaedic Oncology, Changzheng Hospital, Naval Medical University, Shanghai, China, ⁴ Department of Pharmacy, Changhai Hospital, Naval Medical University, Shanghai, China, ⁵ Department of Neurosurgery, Tongji Hospital, Shanghai Tong Ji University School of Medicine, Shanghai, China

OPEN ACCESS

Edited by:

Dianwen Song,
Shanghai First People's Hospital,
China

Reviewed by:

Zhouxiao Li,
Ludwig Maximilian University
of Munich, Germany
Linhua Sun,
Peking University, China

*Correspondence:

Chunhui Wang
wanghi628@163.com
Lei Jiang
jl13jan@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 27 November 2020

Accepted: 08 February 2021

Published: 11 March 2021

Citation:

Liu Z, Zhang H, Hu H, Cai Z,
Lu C, Liang Q, Qian J, Wang C and
Jiang L (2021) A Novel Six-mRNA
Signature Predicts Survival of Patients
With Glioblastoma Multiforme.
Front. Genet. 12:634116.
doi: 10.3389/fgene.2021.634116

Glioblastoma multiforme (GBM) is a devastating brain tumor and displays divergent clinical outcomes due to its high degree of heterogeneity. Reliable prognostic biomarkers are urgently needed for improving risk stratification and survival prediction. In this study, we analyzed genome-wide mRNA profiles in GBM patients derived from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases to identify mRNA-based signatures for GBM prognosis with survival analysis. Univariate Cox regression model was used to evaluate the relationship between the expression of mRNA and the prognosis of patients with GBM. We established a risk score model that consisted of six mRNA (AACS, STEAP1, STEAP2, G6PC3, FKBP9, and LOXL1) by the LASSO regression method. The six-mRNA signature could divide patients into a high-risk and a low-risk group with significantly different survival rates in training and test sets. Multivariate Cox regression analysis confirmed that it was an independent prognostic factor in GBM patients, and it has a superior predictive power as compared with age, IDH mutation status, MGMT, and G-CIMP methylation status. By combining this signature and clinical risk factors, a nomogram can be established to predict 1-, 2-, and 3-year OS in GBM patients with relatively high accuracy.

Keywords: glioblastoma multiforme, six-mRNA signature, LASSO, prognosis, biomarkers

BACKGROUND

Glioblastoma, also known as glioblastoma multiforme (GBM), is the most common primary malignant cancer involving the central nervous system (CNS), characterized by aggressive invasiveness and an infiltrative growth pattern (Lapointe et al., 2018; Lee et al., 2018). According to the Central Brain Tumor Registry of the United States (CBTRUS), GBM accounts for 14.9% of all primary brain tumors and 55.4% of all gliomas (Alexander and Cloughesy, 2017). Current standard treatment for GBM patients comprises maximal safe surgical resection, concurrent

Abbreviations: GBM, Glioblastoma multiforme; OS, Overall survival; L1-LASSO, Least absolute shrinkage and selection operator; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; ROC, Receiver operating characteristic; AUC, Area Under Curve; KPS, Karnofsky; WT, wild type; MT, mutant type; IDH1, isocitrate dehydrogenase 1; MGMT, O6-methylguanine DNA-methyltransferase; CIMP, CpG Island Methylator Phenotype.

chemo-radiotherapy, and adjuvant chemotherapy with temozolomide (TMZ) (Laug et al., 2018). Despite recent advances in multi-modality strategies and individualized therapies, GBM patients usually have a dismal prognosis, with a median overall survival (OS) of less than 15 months (Lapointe et al., 2018).

The 2007 World Health Organization (WHO) classification system has been used for glioma classification over the past decade (Fuller and Scheithauer, 2007). This histological classification and grading system was largely based on visual criteria alone and fails to accurately evaluate the clinical outcomes of GBM patients. Actually, GBM patients with indistinguishable histopathological features can have different clinical outcomes. Accordingly, numerous efforts have been undertaken to characterize the underlying pathological molecular mechanisms and identify reliable prognostic markers for a precise prediction in GBM patients. More recently, some biomarkers including loss of 1p19q chromosome, methylation of the O6-methylguanine DNA-methyltransferase (MGMT), and iso-citrate dehydrogenases 1/2 (IDH1/2) have been found to be closely related to the therapy sensitivity and prognosis (Smits and van den Bent, 2017; Feyissa et al., 2019; Radke et al., 2019). MGMT promoter region was recognized to have a predictive value for the efficacy of TMZ-based chemotherapy, and mutations in the genes encoding for iso-citrate dehydrogenases 1/2 (IDH1/2) could predict a relatively long-term survival of GBM patients (Feyissa et al., 2019; Radke et al., 2019). In the 2016 updated version of the WHO classification system, molecular information has been integrated into pathological diagnosis for further subgroup stratification (Louis et al., 2016). However, due to the biological heterogeneity of GBM, individual biomarkers alone are unable to predict the therapy sensitivity and survival of GBM patients. In this regard, there is an urgent need to identify more effective tumor biomarkers for risk stratification and prognosis prediction.

RNA sequencing (RNA-Seq) is the most commonly used high-throughput technology for transcriptome profiling and offers more sensitive and accurate measurement of gene expression as compared with traditional gene expression arrays (Wang et al., 2009). The Cancer Genome Atlas (TCGA) has accrued RNA-Seq-based transcriptome data across various cancer types and thus provides a robust platform to systematically analyze the large-scale RNA-Seq data (Peng et al., 2015). A comprehensive characterization of gene expression changes in GBM will provide a large number of new potential predictive and prognostic molecular markers. In this study, we retrieved mRNA expression profiles from a large number of GBM patients and analyzed them by re-purposing the publicly available TCGA database to identify an effective signature of mRNAs for predicting survival of GBM patients.

METHODS

TCGA GBM and GSE108474 Datasets Download and Processing

The TCGA GBM data (Level 3 RNA-Seq) of 174 individuals with clinical information were extracted from the TCGA (GDC) database, including data from 169 GBM tissues and 5 normal

brain tissues. The exclusion criteria are as follows: (1) GBM was ruled out by histological diagnosis, (2) the existence of another malignancy with GBM, and (3) patient death due to unrelated causes. Only patients who were followed up for longer than a month were included in the subsequent analysis. Finally, 152 GBM patients were included in this study. Seventy percent of the TCGA patients were randomly selected as the training set and the remaining 30% of the patients were selected as the test set. As the data were downloaded from the public database, ethical approval was not applicable in this case. The data processing procedures met the policies of TCGA data access and human subject protection¹. GSE108474 from the GEO database was conducted through GPL (Affymetrix Human Genome U133 Plus 2.0 Array)². A total of 210 GBM samples were selected from the GSE108474 database. The GEO data were also downloaded from the public database platform. Count data from the TCGA set and GSE108474 set were integrated into expression matrix, which were converted to TPM and standardized to Z score data for subsequent model analysis.

Identification of Differentially Expressed mRNAs in GBM

Three bioconductor software analysis packages—Deseq2, limma + voom, and edgeR—are used to analyze the difference of mRNA expression read count data. The expression differences were characterized by logFC (log₂ fold change). mRNAs with $|\logFC| > 1$ ($P < 0.05$) were considered differentially expressed mRNAs and selected for further analysis.

Identification of mRNAs Significantly Associated With OS and Prognostic Signature Construction in GBM

The semi-supervised method combining both the gene expression data and clinical data was used to identify candidate mRNAs with a prognostic value. Univariate Cox regression analysis was conducted to identify common mRNAs related to OS within each of the subgroups stratified by the IDH status. Within each group of clinical characteristics, the patient subclasses represented non-overlapping sets. Common mRNAs associated with OS in at least two independent subgroups were selected for the subsequent study, using an HR > 1 or HR < 1 with $P < 0.05$ as the cutoff.

Definition of the Prognostic Risk Model and ROC Curve Analysis

The TCGA dataset “caret” package was randomly divided into a training set and a testing set. An importance score was calculated by the supervised principal component method and assigned to each mRNA. Tenfold cross-validation was used to estimate the optimal feature threshold in supervised principal components and select significant mRNAs. The linear mRNA signature prognostic model was developed based on the supervised principal component method. The mRNA expression

¹<http://cancergenome.nih.gov/publications/publicationguidelines>

²<https://www.ncbi.nlm.nih.gov/geo/>

level was expressed as the *Z* score. The prognostic score was calculated as follows: Prognostic score = $(0.058 \times \text{Expression AACS}) + (0.015 \times \text{Expression STEAP1}) + (0.009 \times \text{Expression STEAP2}) + (0.039 \times \text{Expression G6PC3}) + (0.014 \times \text{Expression FKBP9}) + (0.067 \times \text{Expression LOXL1})$. Then, the prognostic scores were computed for the 152 patients using our mRNAs prognostic model. The optimal cutoff value of the prognostic score was decided in the ROC (receiver operating characteristic) curve analysis for predicting 1-, 2-, and 3-year survival of the training set. The OS curves were evaluated using the Kaplan–Meier (KM) and log-rank method. Time-dependent ROC curves were used to evaluate the predictive power of the mRNA signature model. All analyses were performed using the R (version 3.5.2) and Bioconductor (version 3.9).

Survival Analysis

The differences in clinical features including sex, age, IDH1, MGMT, and the survival status between the training set, internal testing set, and independent validation set were analyzed using the chi-square test. A univariate Cox model was performed to investigate the relationship between the continuous expression level of each DEmRNA and OS and for preliminary screening of clinical variables that were correlated with OS of the GBM patients. Hazard ratios (HRs) and *P*-values from univariate Cox regression analysis were used to identify candidate survival-related DEmRNAs. DEmRNAs with HR for death > 1 were defined as high-risk RNAs, and those with HR < 1 were defined as protective RNAs. The common DEmRNAs meeting the criterion of *P*-value < 0.05 were selected as survival-related DEmRNAs and further analyzed in LASSO regression to identify the most powerful prognostic markers. Finally, a multi-mRNA-based classifier was constructed for predicting OS. To quantify the risk of OS, a standard form of risk score (RS) for each GBM patient was calculated in combination with the relative expression levels of the mRNAs.

Statistical Analysis

Mann–Whitney *U*-test and χ^2 -test were used to compare the correlation between continuous variables and classified variables between the training data and test data, respectively. The independent prognostic variables of OS were screened by univariate and multivariate regression analyses. L1-LASSO penalization was carried out using “glmnet” and “pliable” software packages. KM method was used to draw the survival curve based on survminer software package, and log-rank test was used to compare it. Using “rms” software package, Cox’s regression coefficient was calculated to establish the risk assessment formula and mRNA nomogram. Knowing that time-dependent ROC curve analysis is widely used in biomedical reports to evaluate the prediction accuracy of six mRNA signatures, we used the “timeROC” software package to analyze the time-dependent ROC curve. The volcanoes and heat maps were drawn by using the “ggplot2” software package of the R software. All other statistical tests were carried out with R software version 3.5.2 and the corresponding basic software packages. The value of *P* < 0.05 was determined to be statistically significant.

RESULTS

Baseline Characteristics of the Patients and Analysis Flow

Included in this study were 152 GBM patients from TCGA database and 210 GBM patients from the GEO database (GSE108474). The detailed baseline characteristics of the training set and test set are listed in **Table 1** from the TCGA database. The results showed no significant difference in the baseline characteristics between the two sets (all *P* > 0.05). A total of 210 GBM patients from the GEO database make up an independent validation set. Our research process is shown in **Figure 1**.

Candidate OS-Related mRNAs of GBM Patients in the TCGA Cohort

In the TCGA database, 5754 differentially expressed mRNAs [false detection rate < 0.05 and log2 fold change (Log2fc) ≥ 1] were identified by using the expression profiles of 19,781 mRNAs in 152 cases of GBM and 5 cases of normal brain tissues. These 5754 mRNA volcanoes were visualized through the “ggplot2” package of R software (**Figures 2A–C**). These 5754 differentially expressed mRNAs were considered to be the potential prognostic markers of GBM patients, in which the expression of 3054 mRNAs was up-regulated and that of 2700 mRNAs was down-regulated. To filtrate out the mRNAs associated with OS, the Cox’s regression and LASSO regression analyses (lambda parameter selection 1SE) were performed in sequence (**Figures 2D,E**). Finally, six mRNAs related to OS were selected, including AACS, STEAP1, STEAP2, G6PC3, FKBP9, and LOXL1 (**Table 2**).

Development of the Risk Score Formula and Six-mRNA-Based Prognostic Model

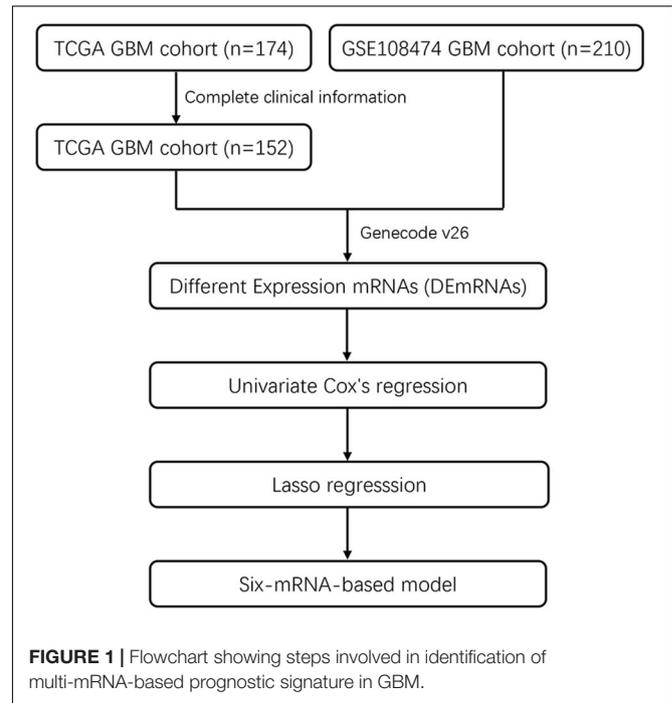
To facilitate the application of identified OS-related mRNAs in clinical practice, the risk score of each patient was calculated using the following equation: Risk score = $(0.058 \times \text{Expression AACS}) + (0.015 \times \text{Expression STEAP1}) + (0.009 \times \text{Expression STEAP2}) + (0.039 \times \text{Expression G6PC3}) + (0.014 \times \text{Expression FKBP9}) + (0.067 \times \text{Expression LOXL1})$. According to the risk score, Youden index (Hughes, 2015) was set to the cutoff value, based on which GBM patients were categorized into a low-risk group and a high-risk group (**Figure 3A**). The survival status of GBM patients in the low- and high-risk groups was obvious in both cohorts (**Figure 3B**). The heat map showed the expression of six mRNAs (AACS, STEAP1, STEAP2, G6PC3, FKBP9, and LOXL1) in the samples (**Figure 3C**). In addition, compared with the low-risk group, KM survival analysis showed that the prognosis of GBM patients in the high-risk group was significantly worse than that in the low-risk group (training and test set in TCGA database *P* < 0.00011 , validation set of GEO database *P* < 0.0001) (**Figures 4A,B**). To evaluate the prognostic prediction performance of GBM patients based on the six-mRNA signature model, we compared the AUC values at different time points to draw the survival ROC curve. The AUC value of the primary set and the validation set at 1, 2, and 3 years was

TABLE 1 | Baseline characteristic of study patients.

Variables	Training date	Test date	Overall	P-value
No. of patients	107	45	152	
Age (years), <i>N</i> (%)				1
Mean (SD)	59.2 (13.4)	59.8 (13.4)	59.6 (13.3)	
Median [Min, Max]	59.0 [21.0, 79.0]	62.0 [21.0, 85.0]	60.0 [21.0, 85.0]	
Age (years), <i>N</i> (%)				0.668
< 50	11 (24.4%)	22 (20.6%)	33 (21.7%)	
≥ 50	34 (75.6%)	85 (79.4%)	119 (78.3%)	
Sex, <i>N</i> (%)				0.143
Female	20 (44.4%)	34 (31.8%)	54 (35.5%)	
Male	25 (55.6%)	73 (68.2%)	98 (64.5%)	
KPS type, <i>N</i> (%)				0.921
<70	9 (20.0%)	23 (21.5%)	32 (21.1%)	
≥70	26 (57.8%)	58 (54.2%)	84 (55.3%)	
Unknown	10 (22.2%)	26 (24.3%)	36 (23.7%)	
IDH1 type, <i>N</i> (%)				0.486
MT	2 (4.4%)	7 (6.5%)	9 (5.9%)	
WT	40 (88.9%)	97 (90.7%)	137 (90.1%)	
Unknown	3 (6.7%)	3 (2.8%)	6 (3.9%)	
CIMP type, <i>N</i> (%)				0.787
G-CIMP	3 (6.7%)	6 (5.6%)	9 (5.9%)	
Non-G-CIMP	41 (91.1%)	100 (93.5%)	141 (92.8%)	
Unknown	1 (2.2%)	1 (0.9%)	2 (1.3%)	
MGMT type, <i>N</i> (%)				0.864
Methylated	18 (40.0%)	38 (35.5%)	56 (36.8%)	
Unmethylated	18 (40.0%)	45 (42.1%)	63 (41.4%)	
Unknown	9 (20.0%)	24 (22.4%)	33 (21.7%)	
Race type, <i>N</i> (%)				0.448
Black or African American	3 (6.7%)	6 (5.6%)	9 (5.9%)	
White	42 (93.3%)	95 (88.8%)	137 (90.1%)	
Asian	0 (0%)	5 (4.7%)	5 (3.3%)	
Unknown	0 (0%)	1 (0.9%)	1 (0.7%)	
Status, <i>N</i> (%)				0.825
Alive	9 (20.0%)	20 (18.7%)	29 (19.1%)	
Dead	36 (80.0%)	87 (81.3%)	123 (80.9%)	
OS time				0.883
Mean (SD)	461 (427)	439 (364)	445 (382)	
Median [Min, Max]	394 [64.0, 2680]	382 [33.0, 2130]	383 [33.0, 2680]	

0.660, 0.668, and 0.674, and 0.680, 0.741, and 0.743, respectively (Figures 4C,D).

To analyze risk factors affecting the prognosis of GBM patients, univariate Cox's analysis was performed and the result showed that the prognosis of GBM patients was associated with age, IDH1 mutation, MGMT methylation, CIMP methylation,

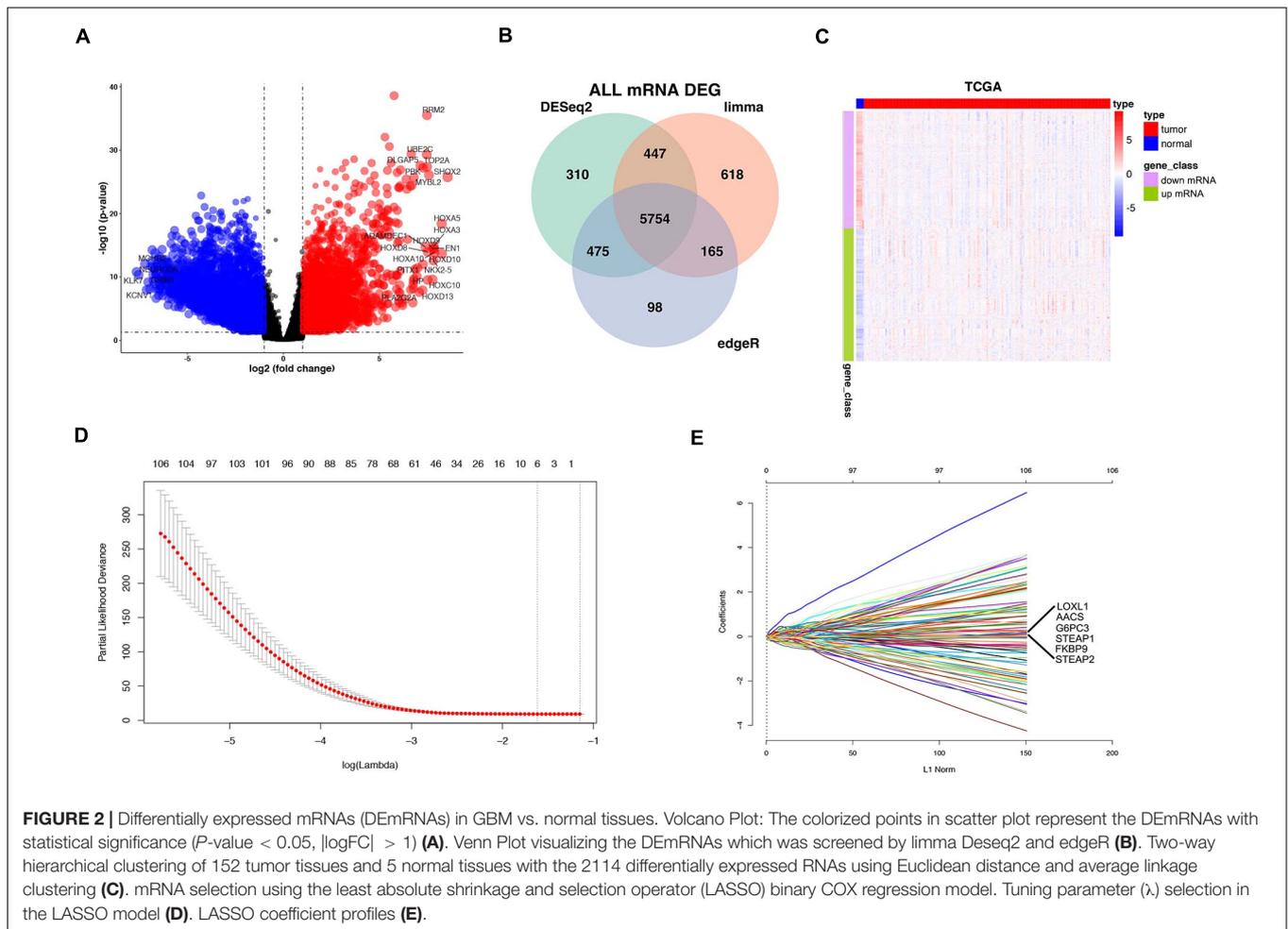


and the six-mRNA signature, but not with gender and KPS (Karnofsky) score. Subsequent multivariate analysis showed that the six-mRNA signature was an independent prognostic survival factor in GBM patients (Table 3).

Predictive Ability of the Six-mRNA-Based Model in Different Risk Stratification

To show whether the six-mRNA-based model could predict the OS time of patients, clinical risk factors including age, KPS score, IDH1 mutation status, MGMT methylation status, and CIMP methylation status in GBM patients were stratified. In patients with age ≥ 50 years ($P = 0.006$), age < 50 years ($P = 0.013$), MGMT methylated ($P < 0.005$), KPS ≥ 70 ($P < 0.005$), CIMP unmethylated ($P < 0.005$), and IDH1 WT ($P = 0.015$), OS of GBM patients in the low-risk group was significantly better than that in the high-risk group. However, in patients with KPS < 70 ($P = 0.088$) and MGMT unmethylated ($P = 0.139$), there was no significant difference in OS between the low-risk and high-risk groups (Figures 5A–H).

To build a convenient and sensitive predictive tool for clinical use, a new prognostic model based on the six-mRNA model was established by combining signature and four clinical risk factors (age, IDH1 status, MGMT status, and CIMP status) to predict 1-, 2-, and 3-years OS in GBM patients. The results showed that the six-mRNA signature and MGMT methylated status contributed most to OS in 1, 2, and 3 years, followed by the CIMP methylated status, patient age, and the IDH1 mutation status in six-mRNA-based nomograms. Each variable obtained a nomogram score on the point scale. By calculating the total score of the nomogram, we easily obtained the nomogram prediction probability of each patient for 1-, 2-, and 3-years OS (Figure 6).



Comparison With Other Prognostic Factors

To compare the predictive accuracy of different prognostic factors, ROC analysis showed that the six-miRNA signature had higher prognostic accuracy than mRNA alone (Figures 7A–C). More importantly, the accuracy of six-mRNA signature prediction was also significantly better than that of clinical risk factors such as age, IDH1 mutation status, MGMT methylation, and G-CIMP methylation status (Figure 7D).

TABLE 2 | Six prognostic mRNAs associated with OS in the training set.

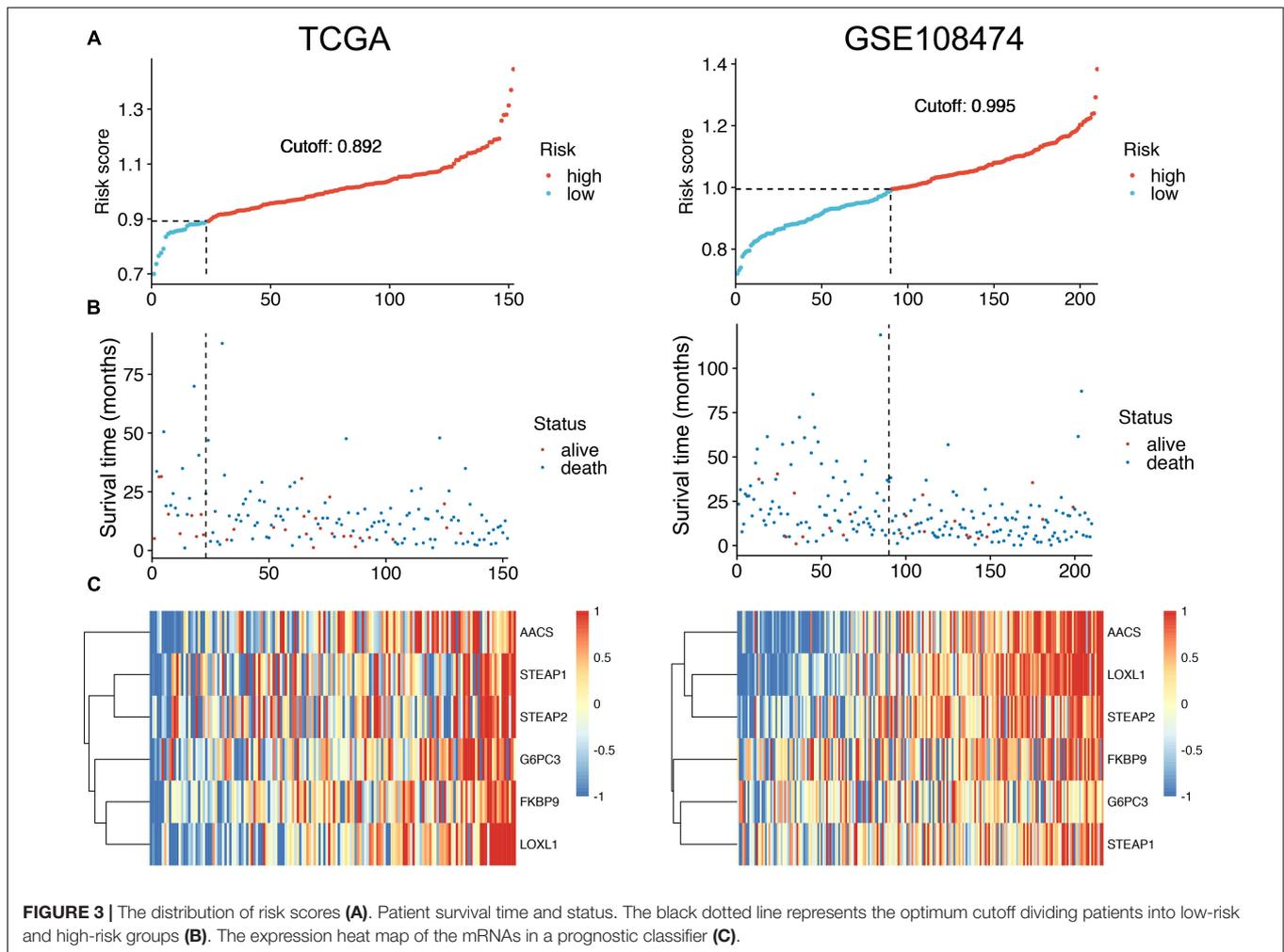
Name	Coefficient	Type	Down/up-regulated	HR	95% CI	P-value
AACS	0.058	Risky	Down	1.132	1.028–1.126	0.011*
STEAP1	0.015	Risky	Up	1.012	1.003–1.021	0.012*
STEAP2	0.009	Risky	Down	1.08	1.035–1.126	<0.001*
G6PC3	0.039	Risky	Up	1.012	1.005–1.019	<0.001*
FKBP9	0.014	Risky	Up	1.003	1.001–1.005	0.004*
LOXL1	0.067	Risky	Up	1.009	1.004–1.013	<0.001*

* $P < 0.05$ was considered statistically significant.

DISCUSSION

Transcriptional aberrations play a crucial role in the complex regulatory network of glioblastoma and are of great importance for the prognosis of this deadly cancer (Bian et al., 2018). Previous studies mainly focused on the function of individual differently expressed mRNAs, which may not be sufficient to clarify the underlying mechanism of tumorigenesis and predict the prognosis of GBM patients (Stoltz et al., 2015; Hu et al., 2017; Yu et al., 2017). Given the high molecular heterogeneity of GBM, integrating multiple biomarkers into a prediction model could improve the prognostic value as compared with a single biomarker. Such multiple-gene signature can be used in prognosis prediction and personalized therapy due to its great power and superiority of risk stratification in many cancers including GBM (Lai et al., 2019; Zuo et al., 2019). Based on this gene signature, patients can be divided into different risk groups so that more sophisticated or intensive targeted therapies could be selected accordingly.

Currently, many prognostic studies have sought to find GBM-specific genomic signatures based on massive genomic data generated by the large-scale cancer genomics projects such as the TCGA database. Yao et al. (2012) have introduced



an algorithm using Cox proportional hazard regression and random re-sampling to identify functional gene sets linked to patient survival based on microarray gene expression datasets (TCGA and REMBRANDT). Another study conducted by Jing et al. (Han and Puri, 2018) found that the expression of interleukin-13 receptor $\alpha 1$ and $\alpha 2$ genes was associated with poor prognosis in GBM patients by using TCGA data. In addition, Wang et al. (2016) developed a signature with three genes that had prognostic values for patients with MGMT promoter-methylated GBM. However, it should be noted that the predictive power might be limited by the small number of individual mRNAs or specific tumor entities in these studies. Recently, Zuo et al. (2019) built an RNA-Seq-based six-gene signature to predict survival in patients with GBM by analyzing RNA-Seq-based gene expression profiles in the CGGA and TCGA databases. They found that this signature could divide patients into a high-risk or low-risk group with different OS and was independent of other clinicopathological features. However, the mRNAs in Zuo's six-gene signature were not all screened from differentially expressed genes in glioma and normal brain tissues, which may limit its significance in revealing tumor heterogeneity. For this reason, we

screened the differentially expressed genes by Cox proportional hazard regression in the present study and only chose GBM survival-related PCGs for further analysis in LASSO regression. Consequently, a six-gene signature was generated using gene expression data from two public databases and was validated in two cohorts of patients.

Various factors may affect the survival of GBM patients, including age, KPS score, extent of surgical removal, MGMT methylated, and the IDH1 mutation status. Our univariate analysis of the TCGA cohort showed that age, IDH1 mutation, MGMT methylated, and CIMP methylated were significantly associated with OS of GBM patients, which is consistent with numerous previous studies (Hegi et al., 2008; Cohen et al., 2013; de Souza et al., 2018). In our study, in GBM patients with age ≥ 50 years ($P = 0.006$), age < 50 years ($P = 0.013$), MGMT methylated ($P < 0.005$), KPS ≥ 70 ($P < 0.005$), CIMP unmethylated ($P < 0.005$), and IDH1 WT ($P = 0.015$), OS of patients in the low-risk group was significantly better than that in the high-risk group. Furthermore, our subsequent multivariate analysis confirmed that this six-gene signature was an independent prognostic predictor of survival for GBM patients.

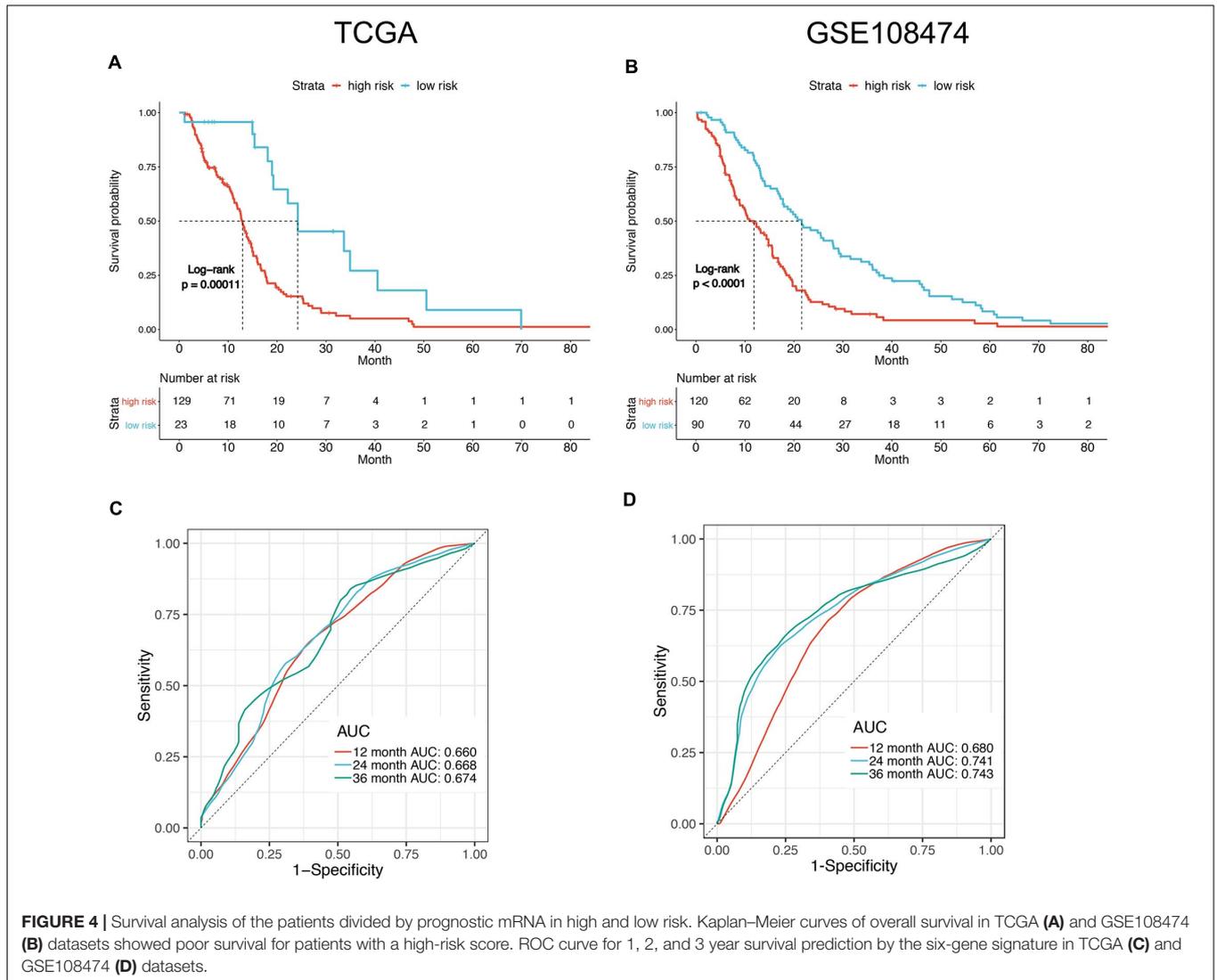


FIGURE 4 | Survival analysis of the patients divided by prognostic mRNA in high and low risk. Kaplan–Meier curves of overall survival in TCGA (A) and GSE108474 (B) datasets showed poor survival for patients with a high-risk score. ROC curve for 1, 2, and 3 year survival prediction by the six-gene signature in TCGA (C) and GSE108474 (D) datasets.

TABLE 3 | Univariate and multivariate Cox’s regression analyses in the training set.

Variables	Patients (N)	Coefficient	Univariate analysis			Multivariate analysis			
			HR	95% CI	P-value	HR	95% CI	P-value	
Age	≥50/ <50	119/33	0.579	1.784	1.13–2.818	0.013*	1.143	0.692–5.563	0.601
Gender	Female/Man	54/98	0.062	1.064	0.732–1.547	0.745			
KPS	≥70/ <70	108/44	−0.084	0.915	0.611–1.368	0.663			
IDH1	Unknown/Mut	6/9	0.217	1.242	0.176–8.758	0.828			
MGMT	WT/Mut	137/9	1.176	3.242	1.317–7.979	0.011*	1.324	0.234–7.493	0.751
	Methylated/Unknown	33/63	−0.306	0.736	0.458–1.182	0.205			
CIMP	Methylated/Unmethylated	56/63	−0.474	0.622	0.408–0.949	0.028*	0.681	0.444–1.045	0.079
	Unknown/non-G-CIMP	2/141	−17.360	0.000	0–inf	0.995			
Six-mRNA signature	G-CIMP/non-G-CIMP	9/141	−1.196	0.303	0.111–0.822	0.019*	0.835	0.138–5.031	0.844
	High/Low	129/23	1.048	2.850	1.621–5.02	0.00028*	2.901	1.513–5.563	0.001*

*P < 0.05 was considered statistically significant.

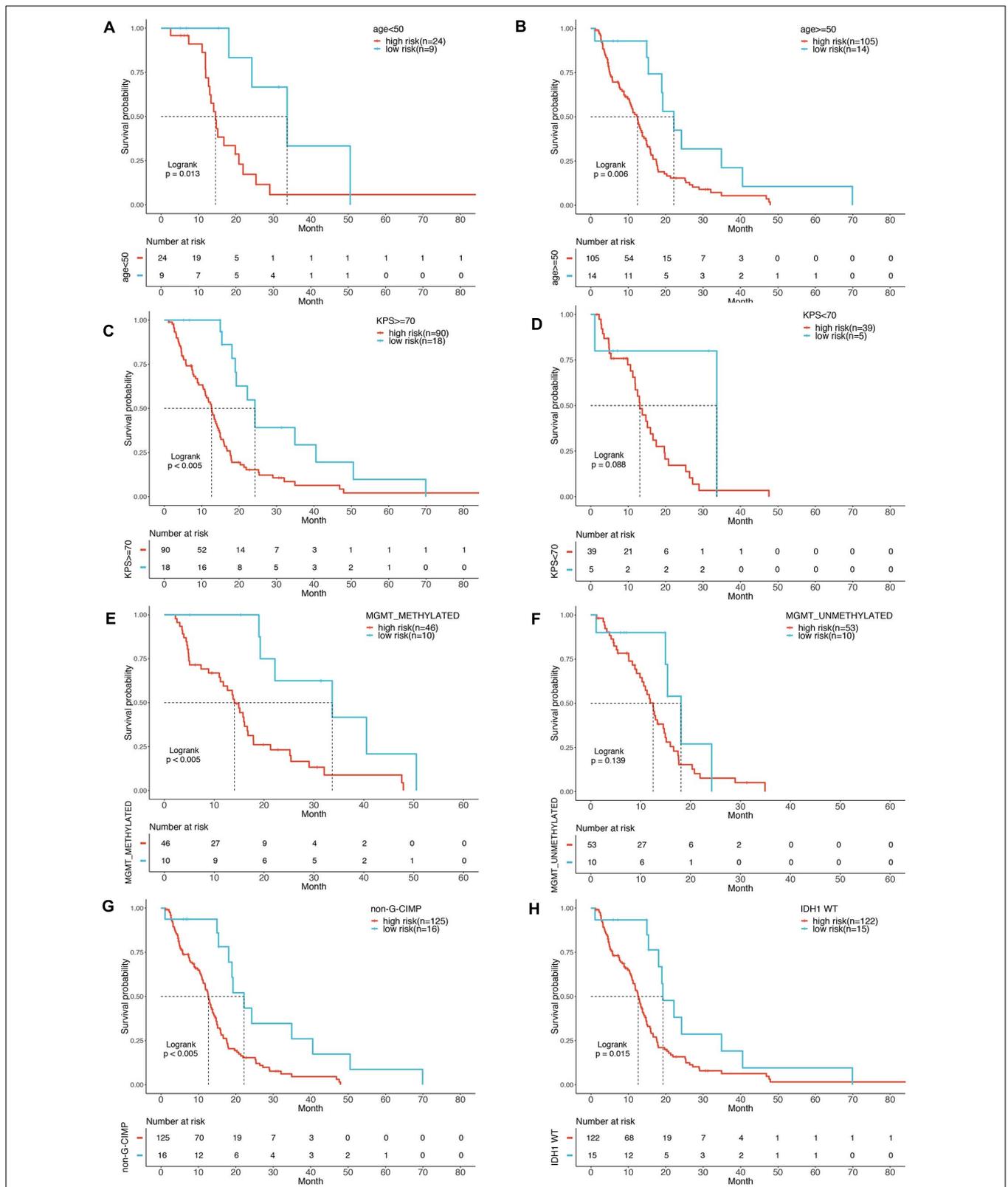
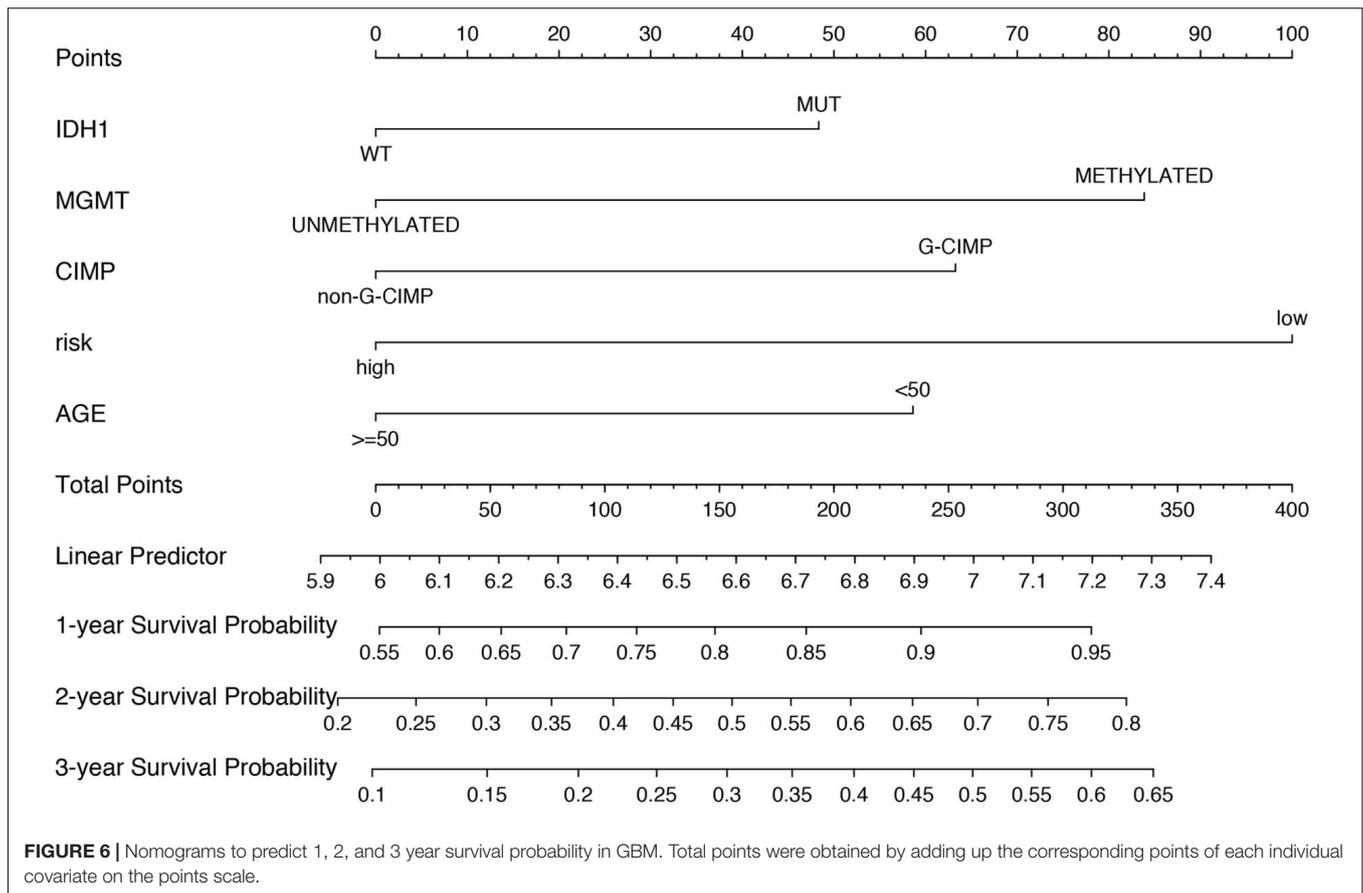


FIGURE 5 | Stratification analysis by different clinical variables. Meier curve analysis of overall survival in high- and low-risk group for younger (age < 50) (A) and older patients (age ≥ 50) (B). KPS ≥ 70 (C), and < 70 (D). MGMT methylated status as methylated (E) and unmethylated (F). G-CIMP methylated (G). IDH1 mutation status as WT (H).



The six-mRNA signature identified in this study includes four up-regulated mRNAs (STEAP1, G6PC3, FKBP9, and LOXL1) and two down-regulated mRNAs (STEAP2 and AACS) with respect to their expression levels in GBM tissue samples. Among these genes, a few genes have established roles in cancers, and other genes might be potential new biomarkers for cancers. STEAP1 and STEAP2, the members of the human six-transmembrane epithelial antigen of the prostate (STEAP) protein family, are highly over-expressed in many different cancer entities including prostate, bladder, breast, colon, and lung carcinoma, Ewing's sarcoma, anaplastic thyroid carcinoma, and malignant melanoma (Hubert et al., 1999; Valenti et al., 2009; Gomes et al., 2012). The high rate of co-expression of STEAP1 and 2 has been observed in cancer cell lines. However, the functional role of STEAP1 and STEAP2 in glioma has not yet been established. Acetoacetyl-CoA synthetase (AACS, acetoacetate-CoA ligase) plays an important role in cholesterol homeostasis and normal neuronal development and was found abundant in normal brain tissues (Ohgami et al., 2003). Knockdown of AACS in primary neurons decreased the expression of MAP-2 and NeuN, two known markers of neuronal differentiation (Hasegawa et al., 2012). G6PC3 is known to encode enzymes that have glucose-6-phosphatase activity, which is ubiquitously expressed in various tissues (Hayee et al., 2011). TP53 can reduce gluconeogenesis by down-regulating

the expression of G6PC gene in colon and liver cancer cells and *in vivo*, thus implying an important regulatory relationship between TP53 and G6PC gene (Zhang et al., 2014). FKBP9 belongs to the FK506-binding protein (FKBP) family, which has peptidyl-prolyl cis-trans isomerases with the enzymatic function attributable to the FKBP domain (Jo et al., 2001). LOXL1, like other Lysyl oxidase (LOX) family members, has an established role in modifying the tumor microenvironment by crosslinking collagens and elastin in the extracellular matrix (Behmoaras et al., 2008). Increased LOXL1 was found in pancreatic ductal adenocarcinoma (Le Calve et al., 2016). In addition, the expression of LOXL1 was found to be correlated with T invasion, lymph node metastasis, and lymphatic and venous invasion in gastric cancer (Kasashima et al., 2014).

In adopting the prediction model of this study, the following limitations need to be considered. First of all, the biological functions and molecular mechanisms of the six mRNAs in the prediction model in glioma remain unclear and further researches are needed. Second, clinical data of postoperative intervention measures for GBM patients in TCGA and GEO databases are incomplete, such as radiotherapy and chemotherapy, so we cannot conduct a comprehensive analysis of OS. Third, the prediction model cannot effectively predict the patient OS with KPS <70 and MGMT demethylation between the low- and high-risk

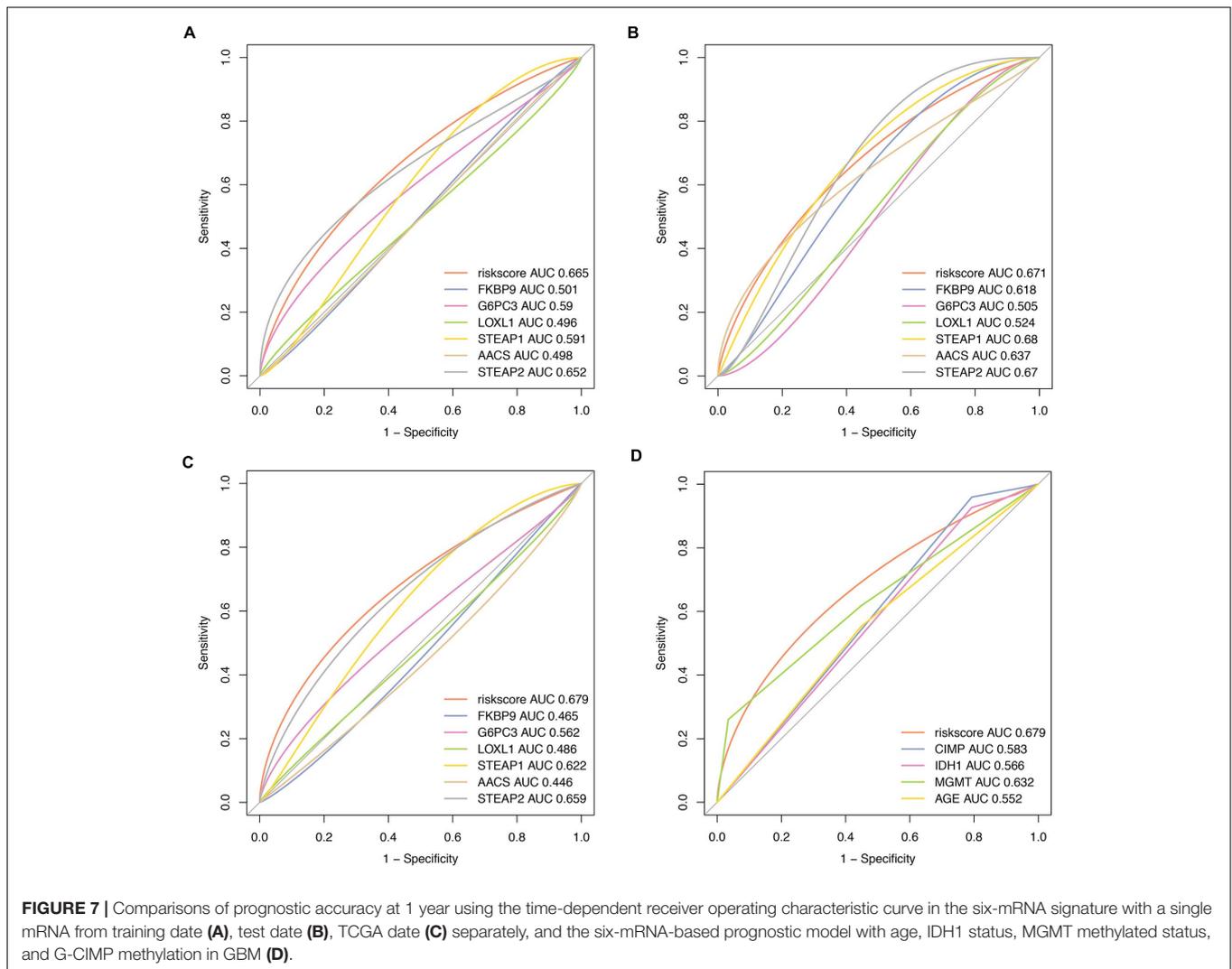


FIGURE 7 | Comparisons of prognostic accuracy at 1 year using the time-dependent receiver operating characteristic curve in the six-mRNA signature with a single mRNA from training date (A), test date (B), TCGA date (C) separately, and the six-mRNA-based prognostic model with age, IDH1 status, MGMT methylated status, and G-CIMP methylation in GBM (D).

group. Therefore, multi-center, large-scale, prospective studies are needed to validate the predictive model in clinical practice.

CONCLUSION

In this study, we have identified a six-gene signature for predicting survival of GBM patients by analyzing RNA-Seq gene expression profiles in the TCGA and GEO databases. This multiple-RNA-based signature could effectively stratify GBM patients into low- and high-risk groups with separate survival curves. Our multivariate analysis demonstrated that this six-gene signature is an independent prognostic predictor of survival of GBM patients. To the best of our knowledge, this is the first report about the prognostic value of the six RNAs (STEAP1, G6PC3, FKBP9, STEAP2, AACS, and LOXL1) in GBM, which may serve as new genetic clues for a better understanding about the development, progression, and recurrence of GBM.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://gepia.cancer-pku.cn/GSE108474>.

AUTHOR CONTRIBUTIONS

ZL, HZ, and HH analyzed the data and wrote the manuscript. ZC and CL designed the study and performed data. QL and JQ prepared the figures and tables. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 81270038 and 81802480), the Shanghai Natural Science Foundation (Nos. 19ZR144890 and 20ZR1457400), and the Scientific Research Initial Funding of Shanghai Tongji Hospital (No. RCQD1704).

REFERENCES

- Alexander, B. M., and Cloughesy, T. F. (2017). Adult glioblastoma. *J. Clin. Oncol.* 35, 2402–2409.
- Behmoaras, J., Slove, S., Seve, S., Vranckx, R., Sommer, P., and Jacob, M. P. (2008). Differential expression of lysyl oxidases LOXL1 and LOX during growth and aging suggests specific roles in elastin and collagen fiber remodeling in rat aorta. *Rejuvenation Res.* 11, 883–889. doi: 10.1089/rej.2008.0760
- Bian, S., Repic, M., Guo, Z., Kavirayani, A., Burkard, T., Bagley, J. A., et al. (2018). Genetically engineered cerebral organoids model brain tumor formation. *Nat. Methods* 15, 631–639. doi: 10.1038/s41592-018-0070-7
- Cohen, A. L., Holmen, S. L., and Colman, H. (2013). IDH1 and IDH2 mutations in gliomas. *Curr. Neurol. Neurosci. Rep.* 13:345.
- de Souza, C. F., Sabedot, T. S., Malta, T. M., Stetson, L., Morozova, O., Sokolov, A., et al. (2018). A distinct DNA methylation shift in a subset of glioma CpG island methylator phenotypes during tumor recurrence. *Cell Rep.* 23, 637–651. doi: 10.1016/j.celrep.2018.03.107
- Feyissa, A. M., Worrell, G. A., Tatum, W. O., Chaichana, K. L., Jentoft, M. E., Guerrero Cazares, H., et al. (2019). Potential influence of IDH1 mutation and MGMT gene promoter methylation on glioma-related preoperative seizures and postoperative seizure control. *Seizure* 69, 283–289. doi: 10.1016/j.seizure.2019.05.018
- Fuller, G. N., and Scheithauer, B. W. (2007). The 2007 revised World Health Organization (WHO) classification of tumours of the central nervous system: newly codified entities. *Brain Pathol.* 17, 304–307. doi: 10.1111/j.1750-3639.2007.00084.x
- Gomes, I. M., Maia, C. J., and Santos, C. R. (2012). STEAP proteins: from structure to applications in cancer therapy. *Mol. Cancer Res.* 10, 573–587. doi: 10.1158/1541-7786.mcr-11-0281
- Han, J., and Puri, R. K. (2018). Analysis of the cancer genome atlas (TCGA) database identifies an inverse relationship between interleukin-13 receptor alpha1 and alpha2 gene expression and poor prognosis and drug resistance in subjects with glioblastoma multiforme. *J. Neurooncol.* 136, 463–474. doi: 10.1007/s11060-017-2680-9
- Hasegawa, S., Kume, H., Iinuma, S., Yamasaki, M., Takahashi, N., and Fukui, T. (2012). Acetoacetyl-CoA synthetase is essential for normal neuronal development. *Biochem. Biophys. Res. Commun.* 427, 398–403. doi: 10.1016/j.bbrc.2012.09.076
- Hayee, B., Antonopoulos, A., Murphy, E. J., Rahman, F. Z., Sewell, G., Smith, B. N., et al. (2011). G6PC3 mutations are associated with a major defect of glycosylation: a novel mechanism for neutrophil dysfunction. *Glycobiology* 21, 914–924. doi: 10.1093/glycob/cwr023
- Hegi, M. E., Liu, L., Herman, J. G., Stupp, R., Wick, W., Weller, M., et al. (2008). Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. *J. Clin. Oncol.* 26, 4189–4199. doi: 10.1200/jco.2007.11.5964
- Hu, X., Martinez-Ledesma, E., Zheng, S., Kim, H., Barthel, F., Jiang, T., et al. (2017). Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. *Neuro Oncol.* 19, 786–795. doi: 10.1093/neuonc/now285
- Hubert, R. S., Vivanco, I., Chen, E., Rastegar, S., Leong, K., Mitchell, S. C., et al. (1999). STEAP: a prostate-specific cell-surface antigen highly expressed in human prostate tumors. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14523–14528. doi: 10.1073/pnas.96.25.14523
- Hughes, G. (2015). Youden's index and the weight of evidence. *Methods Inf. Med.* 54, 198–199.
- Jo, D., Lyu, M. S., Cho, E. G., Park, D., Kozak, C. A., and Kim, M. G. (2001). Identification and genetic mapping of the mouse Fkbp9 gene encoding a new member of FK506-binding protein family. *Mol. Cells* 12, 272–275.
- Kasahima, H., Yashiro, M., Kinoshita, H., Fukuoka, T., Morisaki, T., Masuda, G., et al. (2014). Lysyl oxidase-like 2 (LOXL2) from stromal fibroblasts stimulates the progression of gastric cancer. *Cancer Lett.* 354, 438–446. doi: 10.1016/j.canlet.2014.08.014
- Lai, J., Wang, H., Pan, Z., and Su, F. (2019). A novel six-microRNA-based model to improve prognosis prediction of breast cancer. *Aging (Albany NY)* 11, 649–662. doi: 10.18632/aging.101767
- Lapointe, S., Perry, A., and Butowski, N. A. (2018). Primary brain tumours in adults. *Lancet* 392, 432–446. doi: 10.1016/s0140-6736(18)30990-5
- Laug, D., Glasgow, S. M., and Deneen, B. (2018). A glial blueprint for gliomagenesis. *Nat. Rev. Neurosci.* 19, 393–403. doi: 10.1038/s41583-018-0014-3
- Le Calve, B., Griveau, A., Vindrieux, D., Marechal, R., Wiel, C., Svrcek, M., et al. (2016). Lysyl oxidase family activity promotes resistance of pancreatic ductal adenocarcinoma to chemotherapy by limiting the intratumoral anticancer drug distribution. *Oncotarget* 7, 32100–32112. doi: 10.18632/oncotarget.8527
- Lee, E., Yong, R. L., Paddison, P., and Zhu, J. (2018). Comparison of glioblastoma (GBM) molecular classification methods. *Semin. Cancer Biol.* 53, 201–211. doi: 10.1016/j.semcancer.2018.07.006
- Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820. doi: 10.1007/s00401-016-1545-1
- Ohgami, M., Takahashi, N., Yamasaki, M., and Fukui, T. (2003). Expression of acetoacetyl-CoA synthetase, a novel cytosolic ketone body-utilizing enzyme, in human brain. *Biochem. Pharmacol.* 65, 989–994. doi: 10.1016/s0006-2952(02)01656-8
- Peng, L., Bian, X. W., Li, D. K., Xu, C., Wang, G. M., Xia, Q. Y., et al. (2015). Large-scale RNA-Seq transcriptome analysis of 4043 cancers and 548 normal tissue controls across 12 TCGA cancer types. *Sci. Rep.* 5:13413.
- Radke, J., Koch, A., Pritsch, F., Schumann, E., Misch, M., Hempt, C., et al. (2019). Predictive MGMT status in a homogeneous cohort of IDH wildtype glioblastoma patients. *Acta Neuropathol. Commun.* 7:89.
- Smits, M., and van den Bent, M. J. (2017). Imaging correlates of adult glioma genotypes. *Radiology* 284, 316–331. doi: 10.1148/radiol.2017151930
- Stoltz, K., Sinyuk, M., Hale, J. S., Wu, Q., Otvos, B., Walker, K., et al. (2015). Development of a Sox2 reporter system modeling cellular heterogeneity in glioma. *Neuro Oncol.* 17, 361–371. doi: 10.1093/neuonc/nou320
- Valenti, M. T., Dalle Carbonare, L., Donatelli, L., Bertoldo, F., Giovanazzi, B., Caliar, F., et al. (2009). STEAP mRNA detection in serum of patients with solid tumours. *Cancer Lett.* 273, 122–126.
- Wang, W., Zhang, L., Wang, Z., Yang, F., Wang, H., Liang, T., et al. (2016). A three-gene signature for prognosis in patients with MGMT promoter-methylated glioblastoma. *Oncotarget* 7, 69991–69999.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63.
- Yao, J., Zhao, Q., Yuan, Y., Zhang, L., Liu, X., Yung, W. K., et al. (2012). Identification of common prognostic gene expression signatures with biological meanings from microarray gene expression datasets. *PLoS One* 7:e45894. doi: 10.1371/journal.pone.0045894
- Yu, Z., Sun, Y., She, X., Wang, Z., Chen, S., Deng, Z., et al. (2017). SIX3, a tumor suppressor, inhibits astrocytoma tumorigenesis by transcriptional repression of AURKA/B. *J. Hematol. Oncol.* 10:115.
- Zhang, P., Tu, B., Wang, H., Cao, Z., Tang, M., Zhang, C., et al. (2014). Tumor suppressor p53 cooperates with SIRT6 to regulate gluconeogenesis by promoting FoxO1 nuclear exclusion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10684–10689.
- Zuo, S., Zhang, X., and Wang, L. (2019). A RNA sequencing-based six-gene signature for survival prediction in patients with glioblastoma. *Sci. Rep.* 9:2615.

Conflict of Interest: 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 © 2021 Liu, Zhang, Hu, Cai, Lu, Liang, Qian, Wang 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.