



Gene Expression Profiling Stratifies IDH-Wildtype Glioblastoma With Distinct Prognoses

Yu-Qing Liu^{1,2}, Fan Wu^{1,2}, Jing-Jun Li^{1,2}, Yang-Fang Li^{1,2}, Xing Liu^{1,2}, Zheng Wang^{2,3} and Rui-Chao Chai^{1,2*}

¹ Department of Molecular Neuropathology, Beijing Neurosurgical Institute, Beijing, China, ² Chinese Glioma Genome Atlas Network, Beijing, China, ³ Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

Objectives: In the present study, we aimed to determine the candidate genes that may function as biomarkers to further distinguish patients with isocitrate dehydrogenase (IDH)-wildtype glioblastoma (GBM), which are heterogeneous with respect to clinical outcomes.

Materials and Methods: We selected 41 candidate genes associated with overall survival (OS) using univariate Cox regression from IDH-wildtype GBM patients based on RNA sequencing (RNAseq) expression data from the Chinese Glioma Genome Atlas (CGGA, $n = 105$) and The Cancer Genome Atlas (TCGA, $n = 139$) cohorts. Next, a seven-gene-based risk signature was formulated according to Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm in the CGGA RNAseq database as a training set, while another 525 IDH-wildtype GBM patient TCGA datasets, consisting of RNA sequencing and microarray data, were used for validation. Patient survival in the low- and high-risk groups was calculated using Kaplan-Meier survival curve analysis and the log-rank test. Uni- and multivariate Cox regression analysis was used to assess the prognosis value. Gene ontology (GO) and gene set enrichment analysis (GSEA) were performed for the functional analysis of the seven-gene-based risk signature.

Results: We developed a seven-gene-based signature, which allocated each patient to a risk group (low or high). Patients in the high-risk group had dramatically shorter overall survival than their low-risk counterparts in three independent cohorts. Univariate and multivariate analysis showed that the seven-gene signature remained an independent prognostic factor. Moreover, the seven-gene risk signature exhibited a striking prognostic validity, with AUC of 78.4 and 73.9%, which was higher than for traditional “age” (53.7%, 62.4%) and “GBM sub-type” (57.7%, 52.9%) in the CGGA- and TCGA-RNAseq databases, respectively. Subsequent bioinformatics analysis predicted that the seven-gene signature was involved in the inflammatory response, immune response, cell adhesion, and apoptotic process.

Conclusions: Our findings indicate that the seven-gene signature could be a potential prognostic biomarker. This study refined the current classification system of IDH-wildtype GBM and may provide a novel perspective for the research and individual therapy of IDH-wildtype GBM.

Keywords: glioma, GBM, IDH wildtype, risk signature, biomarker, prognosis

OPEN ACCESS

Edited by:

Liam Chen,
Johns Hopkins University,
United States

Reviewed by:

David D. Eisenstat,
University of Alberta, Canada
Ryuta Saito,
Tohoku University School of
Medicine, Japan

*Correspondence:

Rui-Chao Chai
chairuichao_glia@163.com;
chairuichao@bjtth.org

Specialty section:

This article was submitted to
Neuro-Oncology and Neurosurgical
Oncology,
a section of the journal
Frontiers in Oncology

Received: 25 April 2019

Accepted: 02 December 2019

Published: 17 December 2019

Citation:

Liu Y-Q, Wu F, Li J-J, Li Y-F, Liu X,
Wang Z and Chai R-C (2019) Gene
Expression Profiling Stratifies
IDH-Wildtype Glioblastoma With
Distinct Prognoses.
Front. Oncol. 9:1433.
doi: 10.3389/fonc.2019.01433

INTRODUCTION

Glioblastoma (GBM, WHO grade IV) is the most common and malignant primary intracranial tumor and is associated with a poor prognosis, with a median survival rate of 14–16 months, despite the use of intensive treatments, including surgery, radiotherapy, and chemotherapy (1–4). Isocitrate dehydrogenase (IDH) mutations are one of the most common and earliest detectable genetic alterations in diffuse gliomas, and evidence supports this mutation as a driver of glioma genesis (5). Based on the updated 2016 edition of World Health Organization (WHO) classification of central nervous system (CNS) tumors, GBM could be classified into IDH-wildtype, and IDH-mutant GBM, wherein the former (IDH-wildtype GBM) is associated with a worse prognosis (6, 7). IDH-mutant GBM accounts for about 12% of all GBM, with an occurrence rate in secondary GBM of 84.6%, while its counterpart in primary GBM is rare (5.0%) (8, 9).

Previous studies have indicated that a six-gene signature could further stratify the prognosis of IDH-mutant glioma using gene expression profiling (10). Given that IDH-wildtype and IDH-mutant GBM are regarded as distinct entities despite their similar histology (11), further stratification of patients with IDH-mutant or IDH-wildtype GBM could be a promising approach for the diagnosis and treatment of GBM. The methylation status of the methylguanine methyltransferase (MGMT) promoter has been reported to have predictive value for both IDH-mutant and IDH-wildtype GBM (12, 13). Our recent study presented a comprehensive somatic mutation landscape of secondary GBM and provided a protocol for MET-targeted therapy for precision neuro-oncology (14). A handful of recent studies have assessed the molecular spectrum of IDH-mutant GBM on the basis of genome-wide DNA methylation analysis, copy-number profiling, and gene expression profiling, respectively (11, 15). Nevertheless, the systematical investigation of IDH-wildtype GBM (88%), which is overwhelmingly more common than for IDH-mutant GBM, as well as being heterogeneous with respect to clinical outcomes, remains to be discussed completely.

In the present study, we aimed to further analyze and stratify IDH-wildtype GBM assessed by whole-genome expression profile analysis. We identified a seven-gene-based risk signature for IDH-wildtype GBM in the CGGA-RNAseq cohort, which was then validated in TCGA-RNAseq and TCGA-microarray cohorts. Furthermore, the prognostic value of our signature and underlying biological functions correlated with this signature were also systemically investigated. By improving our understanding of the molecular basis of IDH-wildtype GBM, we expect to develop a superior stratification of these tumors according to the risk signature and supply additional therapeutic targets for the treatment of for IDH-wildtype GBM.

MATERIALS AND METHODS

Samples and Data Collection

This study collected 630 GBM samples from three cohorts: the CGGA-RNA sequencing (RNAseq, $n = 105$), TCGA-RNAseq ($n = 139$), and TCGA-microarray ($n = 386$) cohorts. The

TABLE 1 | Clinicopathological characteristics of patients in the CGGA RNAseq, TCGA RNAseq, and TCGA microarray cohorts.

Characteristics		CGGA RNAseq cohort	TCGA RNAseq cohort	TCGA microarray cohort
		GBM-IDH wildtype ($n = 105$)	GBM-IDH wildtype ($n = 139$)	GBM-IDH wildtype ($n = 386$)
Age (years)	Median (range)	52 (8–81)	62 (24–89)	61 (19–89)
	Age \geq 45	35 (33.3%)	11 (7.9%)	43 (11.1%)
	Age $<$ 45	70 (66.7%)	128 (92.1%)	343 (88.9%)
Gender	Male	68 (64.8%)	89 (64.0%)	239 (61.9%)
	Female	37 (35.2%)	50 (36.0%)	147 (38.1%)
GBM sub-type	Proneural	5 (4.8%)	24 (17.3%)	72 (18.7%)
	Neural	10 (9.5%)	5 (3.6%)	70 (18.1%)
	Classical	41 (39.0%)	85 (61.1%)	108 (28.0%)
	Mesenchymal	49 (46.7%)	25 (18.0%)	115 (29.8%)
	NA	0 (0.00%)	0 (0.00%)	21 (5.4%)
MGMT promoter methylation status	Methylated	37 (35.2%)	43 (30.9%)	122 (31.6%)
	Unmethylated	65 (61.9%)	67 (48.2%)	151 (39.1%)
	NA	3 (2.9%)	29 (20.9%)	113 (29.3%)
TERT promoter status	Wildtype	49 (46.7%)	NA	NA
	Mutation	33 (31.4%)	NA	NA
	NA	23 (21.9%)	NA	NA
EGFR	Wildtype	NA	104 (74.8%)	NA
	Mutation	NA	34 (24.5%)	NA
	NA	NA	1 (0.7%)	NA
TP53	Wildtype	NA	106 (76.3%)	NA
	Mutation	NA	32 (23.0%)	NA
	NA	NA	1 (0.7%)	NA
Chr 7 gain/Chr 10 loss	Combined	NA	93 (66.9%)	NA
	No combined	NA	42 (30.2%)	NA
	NA	NA	4 (2.9%)	NA

CGGA, Chinese Glioma Genome Atlas; TCGA, The Cancer Genome Atlas; MGMT, methylguanine methyltransferase; TERT, telomerase reverse transcriptase; NA, not applicable; Chr, chromosome.

clinical and molecular information of the cases in the CGGA-RNAseq cohort were obtained from the CGGA database@uolink (<http://www.cgga.org.cn/>) and were used as the training set (16). Each case with newly diagnosed GBM was treated by the CGGA group. All tissues were diagnosed histologically by two or more neuropathologists, independently. The overall survival (OS) was calculated from the date of diagnosis until the death of the patient or the end of the clinical follow-up. The study protocol was approved by the Ethics Committee of the Beijing Tiantan Hospital. Another 525 GBM were included from TCGA-RNAseq and TCGA-microarray cohorts (<https://tcga-data.nci.nih.gov/tcga/>) (17, 18) as validation sets. The characteristics of the patient in the three cohorts are provided in **Table 1**.

Gene Oncology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA)

GO and KEGG pathway analyses were performed in DAVID (<http://david.abcc.ncifcrf.gov/home.jsp>) for the functional annotation of the genes correlated positively and negatively with the risk score in the two cohorts (19). GO was used to analyze the main function of the differential expression genes. KEGG was performed to analyze pathway enrichment. GSEA was performed to determine whether the gene sets were statistically different between the two groups (high-risk score vs. low-risk score) using GSEA v3 software (<http://www.broadinstitute.org/gsea/index.jsp>) (20).

Statistical Analysis

Univariate Cox regression analysis was performed to select genes associated with the OS in the CGGA-RNAseq and TCGA-RNAseq cohorts, respectively. Next, the risk associated genes ($HR > 1$) and protective genes ($HR < 1$) were both overlapped between the two cohorts. Ultimately, a total of 41 OS-correlated genes, consisting of 34 risk associated genes and 7 protective genes, were selected to perform further gene signature selection and risk-based classification in the training dataset. A risk signature was formulated according to Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm (21–25). The penalty parameter λ was chosen based on a 50-fold cross validation within the training dataset, which produced the minimum mean cross-validated error for the Cox model. Accordingly, seven genes and their regression coefficients were achieved. The risk score was then calculated in the training and validation datasets using the following equation:

$$\text{Risk score} = \sum_{i=1}^n \text{Coef}_i \times \text{Expr}_i$$

where Coef_i is the coefficient and Expr_i is the z-score-transformed relative expression value of each selected gene. Based on the median risk value (23, 26), patients in CGGA-RNAseq, TCGA-RNAseq, and TCGA microarray databases were divided into high- and low-risk groups. Kaplan-Meier survival curves were calculated and the log rank tests were conducted to assess the prognostic significance (13, 27). Differences in clinicopathologic features between the groups were determined using the Student's t- or Chi-square tests. Multivariate Cox regression analyses were used to confirm independent prognostic factors. All statistical analyses were performed using SPSS version 16.0 software (SPSS Inc., Chicago, IL). P -value < 0.05 was considered as statistically significant.

RESULTS

Screening for Critical Genes Stratified for IDH-Wildtype GBM Through Gene Expression

To stratify IDH-wildtype GBM based on the whole genome expression profiling, we firstly selected genes associated

with overall survival (OS) using univariate Cox regression from IDH-wildtype GBM patients in the CGGA ($n = 105$) and TCGA cohorts ($n = 139$), respectively. Next, the OS-related genes in the two cohorts were divided into two groups: protective genes ($HR < 1$) and risk associated genes ($HR > 1$). The protective genes and risk associated genes were then overlapped between the two cohorts, respectively. Finally, 41 candidate genes, including 34 risk associated genes and 7 protective genes, were selected (Figure 1A). The functional annotations of the 41 candidate genes were enriched in GO terms for biological processes including “Innate immune response,” “Glial cell-derived neurotrophic factor receptor signaling pathway,” and “Cellular response to lipopolysaccharide” (Figure 1B).

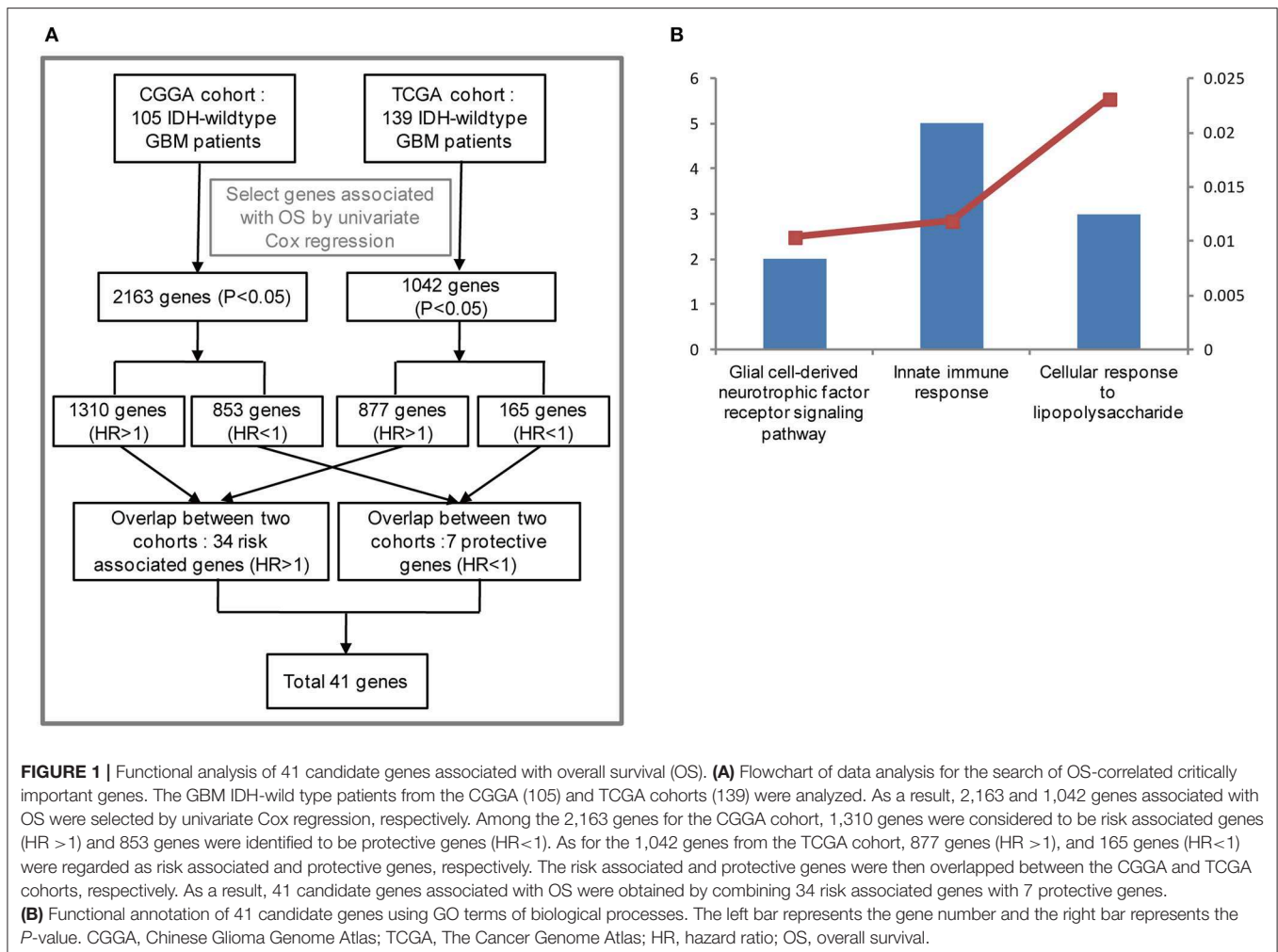
Identification of a Seven-Genes Risk Signature for IDH-Wildtype GBM

Given the prognostic importance of these 41 candidate genes, we attempted to develop a gene expression-based signature that could further stratify IDH-wild type GBM derived from these genes. To this end, using LASSO regression algorithm, seven genes, including 4 protective (*ZNF419*, *FOXG1*, *STARD7*, and *ZBTB16*) genes and 3 risk associated genes (*CD180*, *SDK1*, and *CYP21A2*), were selected as active covariates to assess their prognostic value, thereby obtaining the risk scores for the patients in the training cohort (Figures 2A–C).

According to their median risk score, patients were assigned to either a low- or high-risk group. In GBM with IDH-wildtype, Kaplan-Meier analysis showed that patients in the high-risk group ($n = 53$) had a lower OS than patients in the low-risk group ($n = 52$) in the training cohort (median OS = 8.47 vs. 17.13 months; $P < 0.0001$; Figure 3A). Furthermore, we validated the prognostic value of the risk score in the TCGA-RNAseq cohort. Consequently, we found that OS differed significantly between the high-risk ($n = 70$) and low-risk groups ($n = 69$) in IDH-wildtype GBM patients in the TCGA cohort (median OS = 9.27 vs. 15.57 months; $P = 0.0003$; Figure 3F).

Moreover, given that glioma sub-types are stratified according to the MGMT promoter methylation status, wherein telomerase reverse transcriptase (TERT) status showed distinct tumor characteristics and OS outcomes, we investigated the prognostic value of risk score in these populations. A similar trend was observed in these patients, although no significant difference was found in the unmethylated MGMT promoter or TERT mutation patients (most likely due to the small sample size) (Figures 3B–E,G,H). In summary, our results indicated that the high-risk group was markedly correlated with an unfavorable prognosis in patients with IDH-wildtype GBM.

To further validate the prognostic value of the seven-gene-based risk signature in other cohorts, we computed the risk scores for each patient in the TCGA microarray databases with the same formula. Patients were divided into low- and high-risk groups according to their median risk value. The survival analysis suggested that patients in the high-risk group ($n = 193$) had a lower OS than patients in the low-risk group ($n = 193$; median OS = 12.4 vs. 15.57 months; $P = 0.0097$; Figure 3I). For the



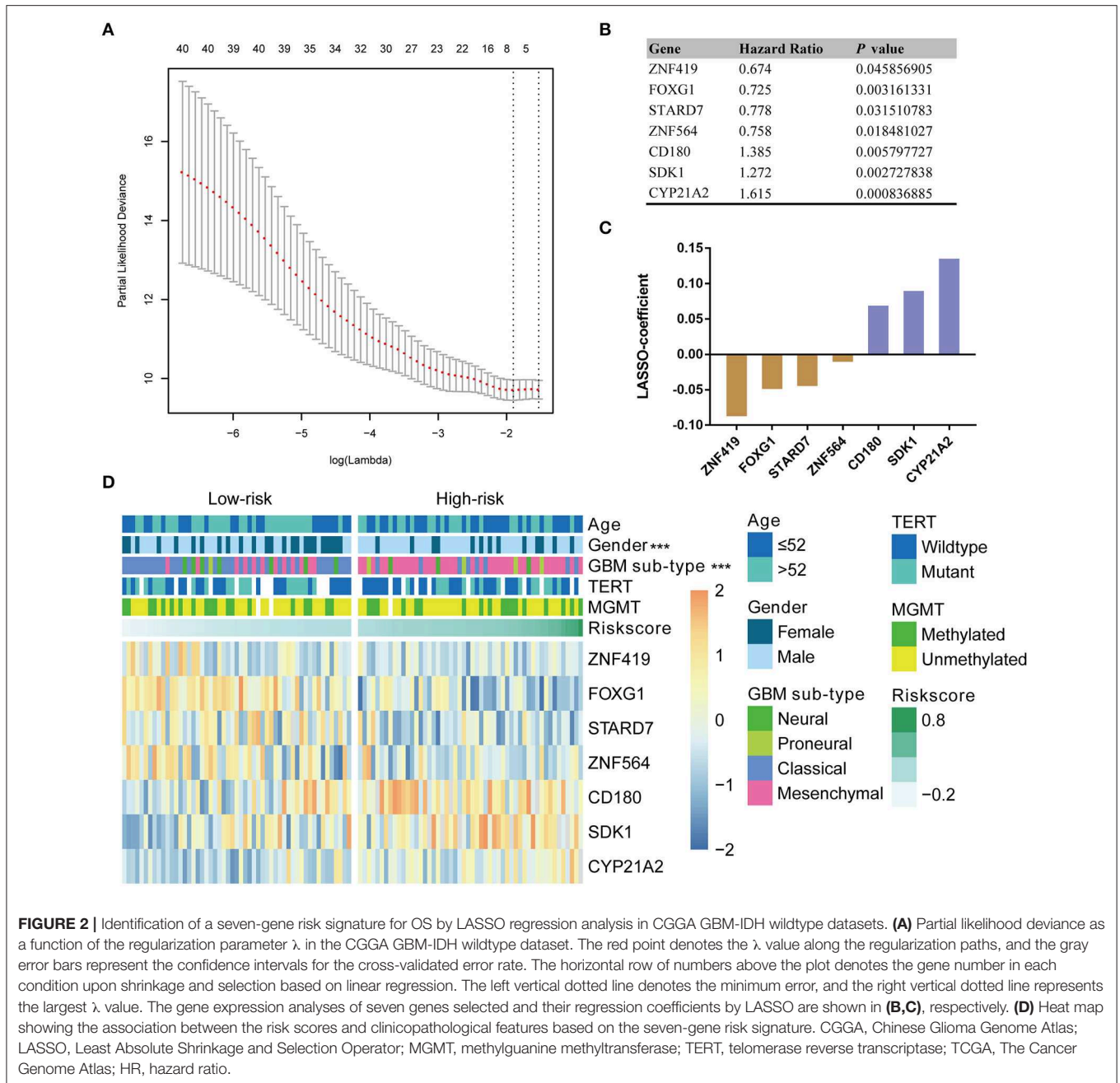
group with a methylated MGMT promoter, compared to the low-risk patients, the OS of high-risk patients was also significantly lower ($P = 0.0219$; **Figure 3J**). On the other hand, there were no significant differences between patients with a methylated MGMT promoter ($P > 0.05$; **Figure 3K**).

Subsequently, we explored whether the prognostic value of the seven-gene signature could be extended to IDH-mutant GBM and lower grade glioma (LGG, WHO grade II-III), by calculating a risk score using the same formula in the CGGA and TCGA cohorts. A Kaplan-Meier analysis showed that there was no significant difference between the low- and high-risk groups in IDH-mutant GBM from the CGGA-RNAseq and TCGA-microarray datasets (**Figures S1A,B**). According to the WHO 2016 update to the classification strategy, LGG were categorized into three subtypes (IDH-wildtype LGG, LGG-IDHmut- non-codel and LGG-IDHmut-codel) based on the status of IDH mutation and 1p/19q codeletion (7). The results showed that the survival time of the high-risk group was remarkably shorter than that of the low-risk group in LGG-IDHmut-non-codel (**Figure S1D**), whereas there was no significant difference in IDH-wildtype LGG and LGG-IDHmut-codel between the two risk groups in the CGGA-RNAseq cohort (**Figures S1C,E**).

Furthermore, we found there was no significant difference in the survival of the high- and low-risk groups in all three subtypes of LGG from the TCGA-RNAseq cohort (**Figures S1F-H**). In summary, these results indicated that the seven-gene signature could not predict the prognosis of patients with IDH-mutant GBM and LGG, identified as an exclusive prognostic marker for IDH-wildtype GBM.

Seven-Gene Signature Is an Independent Prognostic Factor for IDH-Wildtype GBM

We further evaluated the prognostic value of the seven-gene signature for IDH-wildtype GBM patients. Uni- and multivariate Cox regression analyses of the clinical features and seven-gene-based risk score for OS were performed to determine the prognostic significance of the seven-gene signature in IDH-wildtype GBM patients from the CGGA datasets. The results showed that the seven-gene signature was independently associated with OS by adjusting for clinicopathological factors (age, gender, GBM sub-type, radiotherapy, chemotherapy, MGMT promoter methylation status, and TERT status; $P < 0.001$; **Figure 4A**). Consistently, the seven-gene risk signature was validated as an independent indicator after



multivariate Cox regression analyses in the TCGA cohort ($P < 0.001$; **Figure 4B**).

Prognostic Validity of the Seven-Gene Signature for IDH-Wildtype GBM

Subsequently, we determined the specificity and sensitivity of the risk score in the prediction of 1-year survival by calculating the area under the curve (AUC) of the risk score and the pathologic features using the receiving operator characteristic (ROC) curve. As shown in **Figures 4C,D**, the seven-gene signature showed a striking prognostic validity, with an AUC of 78.4 and 73.9%, which were higher those found for the traditional “age” (53.7%,

62.4%) and “GBM sub-type” (57.7%, 52.9%) in the CGGA- and TCGA-RNaseq databases, respectively. These data indicate that the seven-gene signature could be used as a potential prognostic marker of IDH-wildtype GBM.

Association of the Seven-Gene Signature With Other Clinicopathological Features of IDH-Wildtype GBM

To evaluate the performance of the identified signature as a classifier, we classified the CGGA dataset into low- and high-risk groups using the median risk score as a cutoff point, and found a significant difference in several clinical characteristics

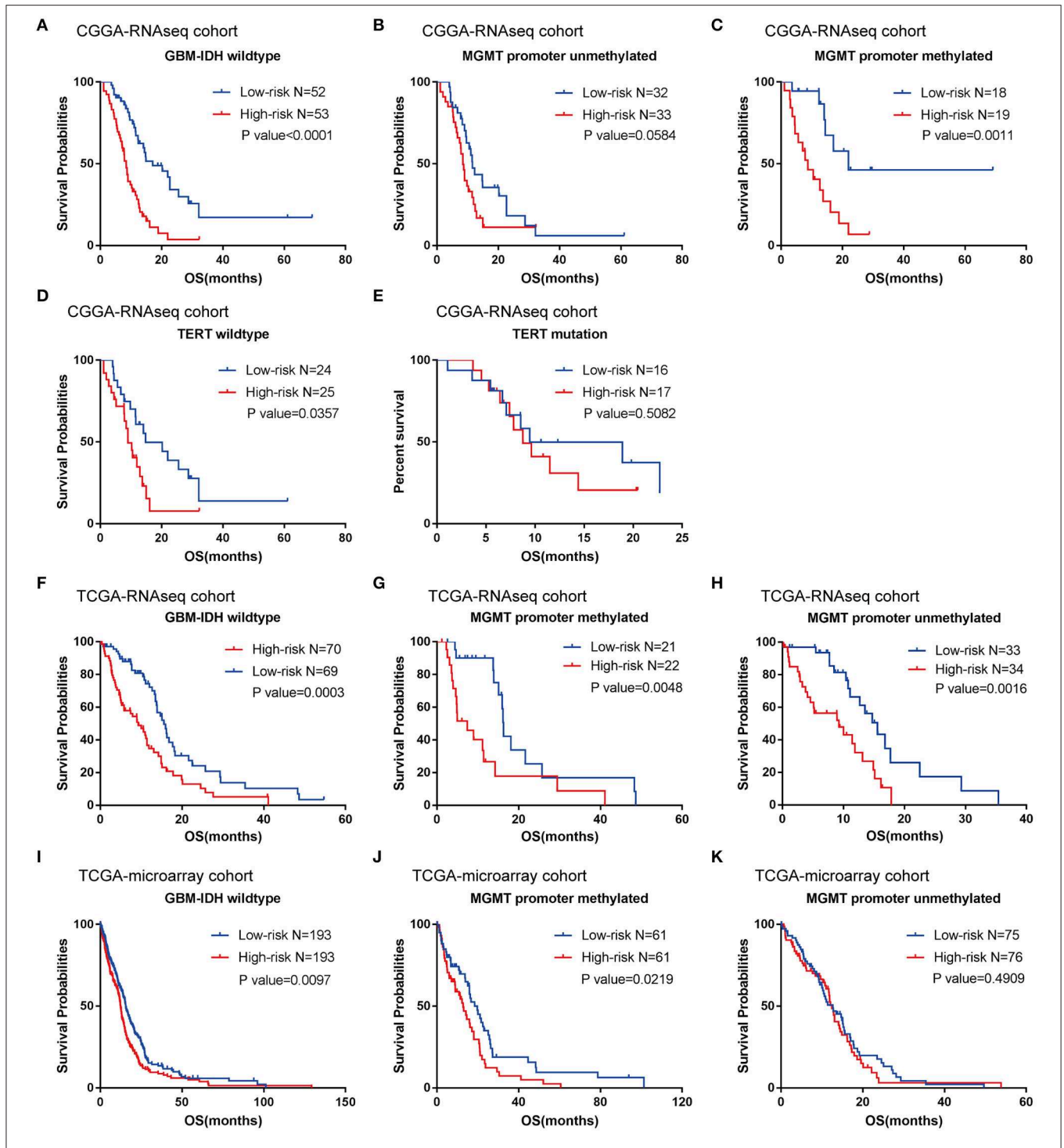


FIGURE 3 | Prognostic significance of the seven-gene signature-based risk scores in GBM-IDH wildtype samples from CGGA and TCGA-RNAseq datasets. **(A–E)** Prognosis efficiency of the seven-gene risk signature in the total GBM-IDH wildtype **(A)**, and MGMT promoter unmethylated **(B)** and methylated **(C)**, TERT promoter wildtype **(D)**, and mutation samples **(E)** from the CGGA-RNAseq datasets, respectively. **(F–H)** Prognosis efficiency of the seven-gene risk signature in TCGA GBM-IDH wildtype datasets **(F)** and MGMT promoter unmethylated **(G)** and methylated **(H)** of GBM-IDH wildtype from the TCGA-RNAseq datasets, respectively. **(I–K)** Prognosis efficiency of the seven-gene risk signature in IDH wildtype GBM from TCGA-microarray datasets **(I)** and MGMT promoter unmethylated **(J)** and methylated **(K)** from the CGGA GBM-IDH wildtype datasets, respectively. The *P*-value shown in each panel were determined using a log-rank test between the two groups. *P* < 0.05 was considered as statistically significant. CGGA, Chinese Glioma Genome Atlas; GBM, glioblastoma; OS, overall survival; TCGA, The Cancer Genome Atlas; IDH, isocitrate dehydrogenases.

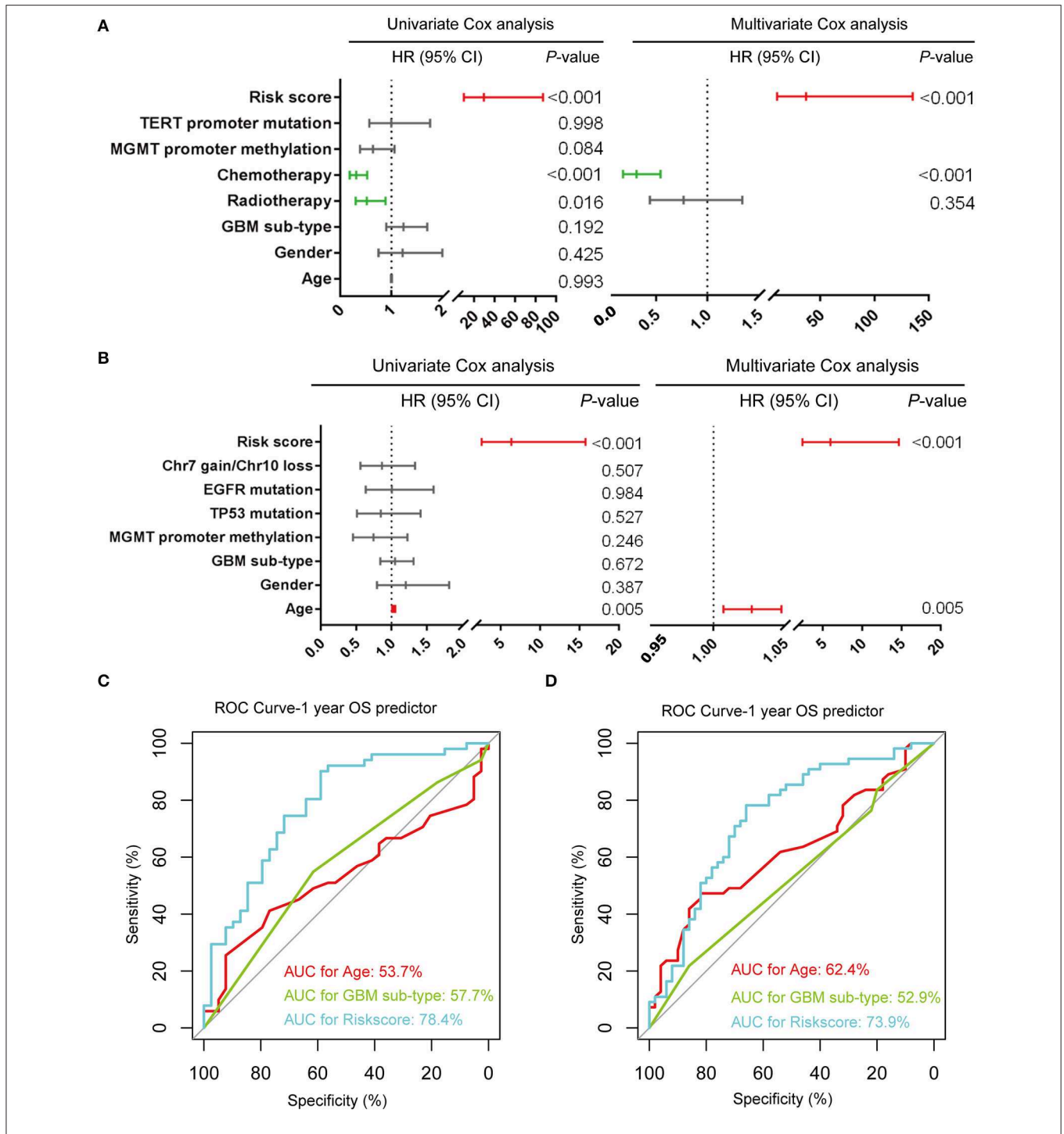


FIGURE 4 | Prognostic validity of the seven-gene signature-based risk scores in GBM-IDH wildtype samples from CGGA and TCGA-RNAseq datasets. **(A,B)** Uni- and multivariate Cox regression analysis of the clinical features and seven-gene-based risk score for OS in GBM-IDH wildtype from CGGA **(A)** and TCGA **(B)** datasets. Variables with prognostic significance in univariate Cox regression analysis were included in further multivariate Cox analysis. Gender (female and male); GBM sub-type (neural, proneural, mesenchymal, and classical); MGMT promoter methylation status (methylated and unmethylated); TERT promoter status (mutant and wildtype); TP53 status (mutant and wildtype); EGFR status (mutant and wildtype); Chr7 gain/Chr10 loss status (combined and not combined); radiotherapy (yes and no); chemotherapy (yes and no); risk score (low and high). **(C,D)** Comparison between the seven-gene signature and traditional risk factors such as age and GBM sub-type in terms of sensitivity and specificity for predicting 1-year survival in the CGGA **(C)** and TCGA **(D)** datasets, respectively. CI, confidence interval; CGGA, Chinese Glioma Genome Atlas; HR, hazard ratio; OS, overall survival; ROC, receiver operating characteristic.

between the two groups (Figure 2D). We found that male patients accounted for a large proportion, 75.5% of the total, of the high-risk group, compared to a proportion of male patients of 53.8% in the low-risk group ($P < 0.001$). As shown in Table S1, the classical and mesenchymal subtypes were found in 61.5 and 23.1%, 17.0 and 69.8% of low-risk and high-risk groups, respectively ($P < 0.001$) (Table S1).

In the TCGA-RNAseq cohort, patients were divided into low- and high-risk groups based on their median risk value. A marked difference was found in several molecular features between the two groups (Figure S2). The combination of “gain of chromosome 7” and “loss of chromosome 10” was found in 78.3 and 55.7% of patients in the low-risk and high-risk groups, respectively ($P < 0.001$). Moreover, we found that 30.4 and 18.6% of samples in the low- and high-risk groups, respectively, were found to harbor mutations in the epidermal growth factor receptor (EGFR) gene ($P = 0.029$) (Table S2). In conclusion, for the CGGA and TCGA cohorts, in comparison with the low-risk group, the high-risk group tended to consist of patients with a worse prognosis.

A previous study identified four clinically relevant subtypes (neural, proneural, classical, mesenchymal) of GBM using integrated genomic analysis (28). Here, we investigated the association between the seven-gene signature and subtype, and found that patients with mesenchymal GBM had a higher risk score than those with classical GBM in the CGGA ($P < 0.0001$; Figure S3A) and TCGA cohorts ($P < 0.05$; Figure S3E). On the other hand, there was no significant correlation between the risk signature and other features, such as age, MGMT status, TERT status, and P53 status in the CGGA (Figures S3B–D) and TCGA cohorts (Figures S3F–H).

Functional Annotation of the Seven-Gene Signature

To investigate the potentially altered functional features correlated with the seven-gene signature, GO and KEGG analyses were conducted based on 589 high-risk score positively-related genes ($P < 0.05$) and 152 negatively-related genes ($P < 0.05$) using Pearson correlation analysis. GO enrichment showed that the top five involved biological processes, that is upregulated gene- in the high-risk group, were “inflammatory response,” “immune response,” “cell adhesion,” “innate immune response,” and “apoptotic process.” In contrast, downregulated genes in the high-risk group were closely associated with neurogenesis functions, such as “brain development,” “nervous system development,” “axon guidance,” and “ion transmembrane transport” (Figure 5A).

Moreover, KEGG pathway analysis showed that positively-related genes in the high-risk group were primarily enriched in biological processes for “TNF signaling pathway,” “cytokine-cytokine receptor interaction,” “leukocyte transendothelial migration,” “toll-like receptor signaling pathway,” and “chemokine signaling pathway,” whereas the negatively correlated genes were enriched in biological terms including “GABAergic synapse,” “Rap1 signaling pathway,” and “glioma” (Figure 5B).

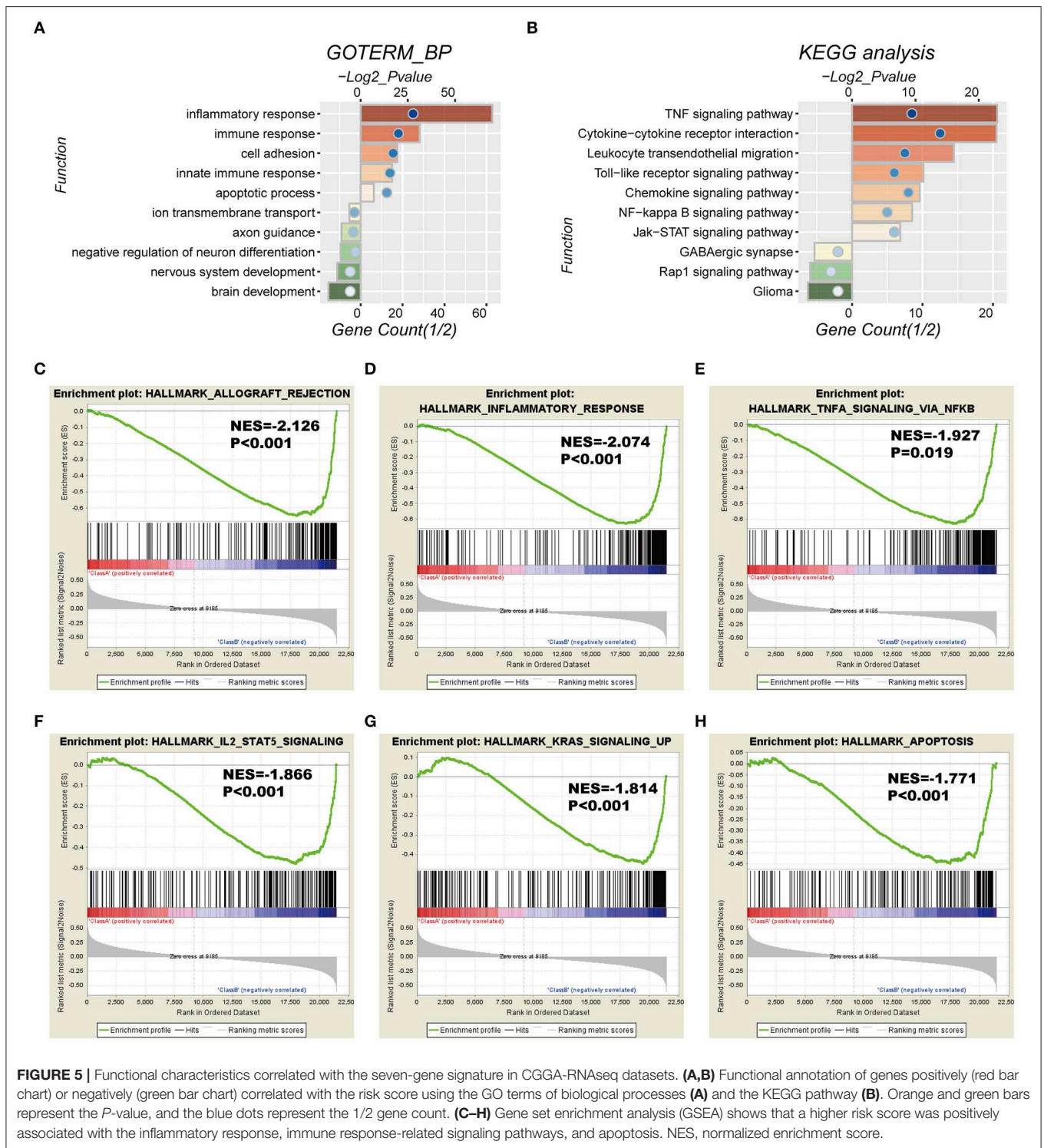
In addition, GSEA analyses were performed for validation, showing that the high-risk groups were positively associated with inflammatory response ($P < 0.001$) and TNF α signaling via NF κ B ($P = 0.019$), IL2-STAT5 signaling ($P < 0.001$), K-ras signaling ($P < 0.001$), and apoptosis ($P < 0.001$; Figures 5C–H). Consistently, these results were validated in the TCGA cohort (Figure S4).

DISCUSSION

In the current study, we studied various candidate genes with potential functions as biomarkers for the stratification of IDH-wildtype GBM with distinct prognoses using whole-genome expression data. We first screened 41 candidate genes, closely associated with OS of IDH-wildtype GBM, by combining the CGGA-RNAseq and TCGA-RNAseq datasets. We then created a seven-gene-based risk signature for IDH-wildtype GBM in the CGGA-RNAseq cohort, which was subsequently validated in the TCGA-RNAseq and TCGA-microarray cohorts. Moreover, the seven-gene risk signature, identified as an independent prognostic significance for IDH1-wildtype GBM, exhibited a greater prognostic value than other factors, underscoring the superiority of a gene expression profile-based signature (29, 30). Finally, bioinformatics analysis was used to predict that the seven-gene signature was involved in the inflammatory response, immune response, cell adhesion, and apoptotic process. To summarize, our seven-gene-based signature refined the current classification system of IDH-wildtype GBM and contributed to improving our understanding of the carcinogenesis and development of IDH-wildtype GBM.

In this study, we established a seven-gene-based signature based on the diversity of genes, including protective (*ZNF419*, *FOXG1*, *STARD7*, and *ZBTB16*) and risk associated (*CD180*, *SDK1*, and *CYP21A2*) genes, which could classify IDH-wildtype GBM into low- and high-risk groups to distinguish between the clinical outcomes. Among these genes, several had been previously studied in various tumors. Some studies have suggested that FoxG1 functions as an oncogene by promoting proliferation, as well as inhibiting differential responses in glioblastoma, by downregulating FoxO/Smad signaling (31). Moreover, low FoxG1 and high Olig-2 labeling indices define a prognostically favorable subset in IDH-mutant gliomas (32). ZBTB16 (Zinc Finger and BTB Domain Containing 16), is a transcription factor involved in the regulation of diverse biological processes, including cell proliferation, differentiation, organ development, stem cell maintenance, and innate immune cell development. A number of recent studies have now implicated PLZF in cancer progression as a tumor suppressor (33). CYP21A2 (21-hydroxylase) is a steroidogenic enzyme crucial for the synthesis of mineralocorticoids and glucocorticoids, and was identified as the most frequently mutated gene in esophageal squamous cell carcinoma by whole exome sequencing (34, 35).

The 2016 WHO classification made a clear difference between GBM that were IDH-mutant and those that were IDH-wildtype, and IDH-wildtype GBM carried a worse prognosis. Consistently, in the cohorts used in this study, the median OS of the



IDH-wildtype GBM are 11.5, 13.3, and 13.83 months in the CGGA-RNAseq cohort, TCGA-RNAseq cohort and TCGA-microarray cohort, respectively, while, those of the IDH-mutant GBM are 18.5 months (CGGA-RNAseq cohort), 34.13 months (TCGA-RNAseq cohort), and 35.9 months (TCGA microarray cohort). However, we observed that the survival of patients with

IDH-wildtype GBM varies from <3 months to more than 3 years, and the similar findings have also been reported in the previous study (3). Therefore, to further stratify IDH-wildtype GBM becomes important and meaningful.

In this study, we determined a seven-gene-based risk signature, a useful tool for risk stratification, to distinguish

between prognoses for IDH-wildtype GBM in three independent cohorts. The median OS of patients with low- and high-risk are significantly different in the CGGA-RNAseq (17.13 vs. 8.47 months), TCGA-RNAseq (15.57 vs. 9.27 months) and TCGA-microarray cohorts (15.57 vs. 12.4 months). This is very meaningful to the post-operative management of these patients with IDH-wildtype GBM. Our previous study presented a gene signature based on GBM stem-like cell relevant genes for primary GBM (36). In addition, several previous studies have found a local immune signature for GBM, indicating the relationship between prognosis and the local immune response (19). Meanwhile, the strength of our study was based on the systematical expression profiling, the robust nature of risk score method (37), and validation across multi-platforms and multi-populations. Although the predictive value of the seven-gene signature was confirmed in distinct datasets, a prospective study with a larger sample size will be needed to assess its clinical relevance.

In conclusion, our results indicate that the seven-gene signature could be a potential prognostic biomarker, providing a novel perspective for research and treatment of IDH-wildtype GBM.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.cgga.org.cn/>.

REFERENCES

- Ricard D, Idbaih A, Ducray F, Lahutte M, Hoang-Xuan K, Delattre JY. Primary brain tumours in adults. *Lancet*. (2012) 379:1984–96. doi: 10.1016/S0140-6736(11)61346-9
- Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med*. (2015) 372:2481–98. doi: 10.1056/NEJMoa1402121
- Jiang T, Mao Y, Ma W, Mao Q, You Y, Yang X, et al. CGCG clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett*. (2016) 375:263–73. doi: 10.1016/j.canlet.2016.01.024
- Chai RC, Wu F, Wang QX, Zhang S, Zhang KN, Liu YQ, et al. m(6)A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas. *Aging*. (2019) 11:1204–25. doi: 10.18632/aging.101829
- Agnihotri S, Aldape KD, Zadeh G. Isocitrate dehydrogenase status and molecular subclasses of glioma and glioblastoma. *Neurosurg Focus*. (2014) 37:E13. doi: 10.3171/2014.9.FOCUS14505
- Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. (2009) 360:765–73. doi: 10.1056/NEJMoa0808710
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
- Watanabe T, Nobusawa S, Kleihues P, Ohgaki H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol*. (2009) 174:1149–53. doi: 10.2353/ajpath.2009.080958
- Kloosterhof NK, Bralten LB, Dubbink HJ, French PJ, van den Bent MJ. Isocitrate dehydrogenase-1 mutations: a fundamentally new understanding of diffuse glioma? *Lancet Oncol*. (2011) 12:83–91. doi: 10.1016/S1470-2045(10)70053-X
- Cheng W, Ren X, Zhang C, Cai J, Han S, Wu A. Gene expression profiling stratifies IDH1-mutant glioma with distinct prognoses. *Mol Neurobiol*. (2017) 54:5996–6005. doi: 10.1007/s12035-016-0150-6
- Korshunov A, Casalini B, Chavez L, Hielscher T, Sill M, Ryzhova M, et al. Integrated molecular characterization of IDH-mutant glioblastomas. *Neuropathol Appl Neurobiol*. (2019) 45:108–18. doi: 10.1111/nan.12523
- Arita H, Yamasaki K, Matsushita Y, Nakamura T, Shimokawa A, Takami H, et al. A combination of TERT promoter mutation and MGMT methylation status predicts clinically relevant subgroups of newly diagnosed glioblastomas. *Acta Neuropathol Commun*. (2016) 4:79. doi: 10.1186/s40478-016-0351-2
- Chai RC, Liu YQ, Zhang KN, Wu F, Zhao Z, Wang KY, et al. A novel analytical model of MGMT methylation pyrosequencing offers improved predictive performance in patients with gliomas. *Mod Pathol*. (2018) 32:4–15. doi: 10.1038/s41379-018-0143-2
- Hu H, Mu Q, Bao Z, Chen Y, Liu Y, Chen J, et al. Mutational landscape of secondary glioblastoma guides MET-targeted trial in brain tumor. *Cell*. (2018) 175:1665–78 e1618. doi: 10.1016/j.cell.2018.09.038
- Wu F, Chai RC, Wang Z, Liu YQ, Zhao Z, Li GZ, et al. Molecular classification of IDH-mutant glioblastomas based on gene expression profiles. *Carcinogenesis*. (2019) 40:853–60. doi: 10.1093/carcin/bgz032
- Wu F, Zhao Z, Chai R, Liu Y, Wang K, Wang Z, et al. Expression profile analysis of antisense long non-coding RNA identifies WDFY3-AS2 as a prognostic biomarker in diffuse glioma. *Cancer Cell Int*. (2018) 18:107. doi: 10.1186/s12935-018-0603-2
- Brennan CW, Verhaak RG, McKenna A, Campos B, Nounshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell*. (2013) 155:462–77. doi: 10.1016/j.cell.2013.09.034

ETHICS STATEMENT

The study protocol was approved by the Ethics Committees of participating hospitals. All patients provided written and informed consent.

AUTHOR CONTRIBUTIONS

Y-QL and R-CC designed the study. Y-QL and FW reviewed the literature, created the figures and tables, and were responsible for the writing and critical editing of the manuscript. J-JL, Y-FL, XL, and ZW contributed to the data collection and analysis, as well as the critical revision of the manuscript. R-CC supervised the critical revision of the manuscript.

FUNDING

The authors represent the Chinese Glioma Cooperative Group (CGCG). This work was supported by grants from the National Natural Science Foundation of China (grant numbers 81903078, 81672479, 81702460, and 81773208).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2019.01433/full#supplementary-material>

18. Zhang C, Cheng W, Ren X, Wang Z, Liu X, Li G, et al. Tumor purity as an underlying key factor in Glioma. *Clin Cancer Res.* (2017) 23:6279–91. doi: 10.1158/1078-0432.CCR-16-2598
19. Cheng W, Ren X, Zhang C, Cai J, Liu Y, Han S, et al. Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology.* (2016) 86:2226–34. doi: 10.1212/WNL.0000000000002770
20. 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. doi: 10.1073/pnas.0506580102
21. Gao J, Kwan PW, Shi D. Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural Netw.* (2010) 23:257–64. doi: 10.1016/j.neunet.2009.07.001
22. Qian Z, Li Y, Fan X, Zhang C, Wang Y, Jiang T, et al. Molecular and clinical characterization of IDH associated immune signature in lower-grade gliomas. *Oncoimmunology.* (2018) 7:e1434466. doi: 10.1080/2162402X.2018.1434466
23. Chai RC, Li YM, Zhang KN, Chang YZ, Liu YQ, Zhao Z, et al. RNA processing genes characterize RNA splicing and further stratify lower-grade glioma. *JCI Insight.* (2019) 5:130591. doi: 10.1172/jci.insight.130591
24. Chai RC, Zhang KN, Chang YZ, Wu F, Liu YQ, Zhao Z, et al. Systematically characterize the clinical and biological significances of 1p19q genes in 1p/19q non-codeletion glioma. *Carcinogenesis.* (2019) 40:1229–39. doi: 10.1093/carcin/bgz102
25. Liu YQ, Chai RC, Wang YZ, Wang Z, Liu X, Wu F, et al. Amino acid metabolism-related gene expression-based risk signature can better predict overall survival for glioma. *Cancer Sci.* (2019) 110:321–33. doi: 10.1111/cas.13878
26. Chai RC, Chang YZ, Wang QW, Zhang KN, Li JJ, Huang H, et al. A novel DNA methylation-based signature can predict the responses of MGMT promoter unmethylated glioblastomas to temozolomide. *Front Genet.* (2019) 10:910. doi: 10.3389/fgene.2019.00910
27. Chai RC, Zhang KN, Liu YQ, Wu F, Zhao Z, Wang KY, et al. Combinations of four or more CpGs methylation present equivalent predictive value for MGMT expression and temozolomide therapeutic prognosis in gliomas. *CNS Neurosci Ther.* (2018) 25:314–22. doi: 10.1111/cns.13040
28. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* (2010) 17:98–110. doi: 10.1016/j.ccr.2009.12.020
29. Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooi JJ, Stubbs AP, Duijm JE, et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* (2009) 69:9065–72. doi: 10.1158/0008-5472.CAN-09-2307
30. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* (2004) 64:6503–10. doi: 10.1158/0008-5472.CAN-04-0452
31. Wang L, Wang J, Jin T, Zhou Y, Chen Q. FoxG1 facilitates proliferation and inhibits differentiation by downregulating FoxO/Smad signaling in glioblastoma. *Biochem Biophys Res Commun.* (2018) 504:46–53. doi: 10.1016/j.bbrc.2018.08.118
32. Schafer S, Behling F, Skardelly M, Koch M, Ott I, Paulsen F, et al. Low FoxG1 and high Olig-2 labelling indices define a prognostically favourable subset in isocitrate dehydrogenase (IDH)-mutant gliomas. *Neuropathol Appl Neurobiol.* (2018) 44:207–23. doi: 10.1111/nan.12447
33. Stopsack KH, Gerke T, Tyekucheva S, Mazzu YZ, Lee GM, Chakraborty G, et al. Low expression of the androgen-induced tumor suppressor gene PLZF and lethal prostate cancer. *Cancer Epidemiol Biomarkers Prev.* (2019) 28:707–14. doi: 10.1158/1055-9965.EPI-18-1014
34. Neunzig J, Milhim M, Schiffer L, Khatri Y, Zapp J, Sanchez-Guijo A, et al. The steroid metabolite 16(beta)-OH-androstenedione generated by CYP21A2 serves as a substrate for CYP19A1. *J Steroid Biochem Mol Biol.* (2017) 167:182–91. doi: 10.1016/j.jsbmb.2017.01.002
35. Deng J, Weng X, Ye J, Zhou D, Liu Y, Zhao K. Identification of the germline mutation profile in esophageal squamous cell carcinoma by whole exome sequencing. *Front Genet.* (2019) 10:47. doi: 10.3389/fgene.2019.00047
36. Chai R, Zhang K, Wang K, Li G, Huang R, Zhao Z, et al. A novel gene signature based on five glioblastoma stem-like cell relevant genes predicts the survival of primary glioblastoma. *J Cancer Res Clin Oncol.* (2018) 144:439–47. doi: 10.1007/s00432-017-2572-6
37. Yan W, Zhang W, You G, Zhang J, Han L, Bao Z, et al. Molecular classification of gliomas based on whole genome gene expression: a systematic report of 225 samples from the Chinese Glioma Cooperative Group. *Neuro Oncol.* (2012) 14:1432–40. doi: 10.1093/neuonc/nos263

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 © 2019 Liu, Wu, Li, Li, Liu, Wang and Chai. 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.