

# A Novel 10-Gene Signature Predicts Poor Prognosis in Low Grade Glioma

Technology in Cancer Research & Treatment  
Volume 20: 1-11  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1533033821992084  
journals.sagepub.com/home/tct



Wentao Liu<sup>1</sup>, Jiaxuan Zou<sup>2</sup>, Rijun Ren<sup>1</sup>, Jingping Liu<sup>1</sup>,  
Gentang Zhang<sup>1</sup>, and Maokai Wang<sup>1</sup>

## Abstract

**Aim:** Low grade glioma (LGG) is a lethal brain cancer with relatively poor prognosis in young adults. Thus, this study was performed to develop novel molecular biomarkers to effectively predict the prognosis of LGG patients and finally guide treatment decisions. **Methods:** survival-related genes were determined by Kaplan-Meier survival analysis and multivariate Cox regression analysis using the expression and clinical data of 506 LGG patients from The Cancer Genome Atlas (TCGA) database and independently validated in a Chinese Glioma Genome Atlas (CGGA) dataset. A prognostic risk score was established based on a linear combination of 10 gene expression levels using the regression coefficients of the multivariate Cox regression models. GSEA was performed to analyze the altered signaling pathways between the high and low risk groups stratified by median risk score. **Results:** We identified a total of 1489 genes significantly correlated with patients' prognosis in LGG. The top 5 protective genes were *DISP2*, *CKMT1B*, *AQP7*, *GPR162* and *CHGB*, the top 5 risk genes were *SPI1*, *EYA3*, *ZSCAN20*, *ITPRIPL1* and *ZNF217* in LGG. The risk score was predictive of poor overall survival and relapse-free survival in LGG patients. Pathways of small cell lung cancer, pathways in cancer, chronic myeloid leukemia, colorectal cancer were the top 4 most enriched pathways in the high risk group. *SPI1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* and *GPR162* were significantly up-regulated, while *DISP2*, *CKMT1B*, *AQP7* were down-regulated in 523 LGG tissues as compared to 1141 normal brain controls. **Conclusions:** The 10-gene signature may become novel prognostic and diagnostic biomarkers to considerably improve the prognostic prediction in LGG.

## Keywords

low grade glioma, 10-gene signature, risk score, overall survival, relapse-free survival

## List of Abbreviations

LGG, Lower grade glioma; TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEG, Differentially expressed gene; ROC, Receiver operating curves; AUC, Area under curve; OS, Overall survival; RFS, Relapse-free survival; OR, Odd ratio; CI, Confidence interval; GSEA, Gene set enrichment analysis

Received: November 06, 2020; Revised: December 23, 2020; Accepted: January 13, 2021.

## Introduction

Gliomas are malignant tumors that arise from glial cells and the most prevalent type of adult brain tumors. The incidence rate of the disease is 6.03 per 100,000 in USA.<sup>1</sup> The incidence of the tumor in the People's Republic of China is 1-4/100,000.<sup>2</sup> Gliomas show high histological diversity, with astrocytoma the most prevalent histological subtype.<sup>1</sup> Gliomas are classified as Grades I to IV based on histology and clinical criteria.<sup>3</sup> Diffuse low-grade and intermediate-grade gliomas refer to World Health Organization grades II and III gliomas (hereafter

<sup>1</sup> Department of Neurosurgery, Qingdao Jiaozhou Central Hospital, Qingdao, Shandong Province, China

<sup>2</sup> Fuzhou Medical College of Nanchang University, Nanchang, Jiangxi Province, China

### Corresponding Author:

Maokai Wang, Department of Neurosurgery, Qingdao Jiaozhou Central Hospital, 29 Xuzhou Road, Qingdao 266300, Shandong Province, China.  
Email: csm791216@163.com



referred to as lower-grade gliomas [LGG]) and include astrocytomas, oligodendrogliomas, and oligoastrocytomas.<sup>1,3</sup>

Most patients initially receive resection at the time of diagnosis and then radiation therapy or treatment with temozolamide.<sup>4</sup> The mean survival time is approximately 7 years, the percentage of LGG patients who can survive for more than 2 decades is as low as 20%.<sup>4</sup> A number of studies have demonstrated the associations of genetic markers with overall survival of LGG patients. For instance, the 1p-19q deletion is a powerful predictor of chemotherapy response and survival and serves as a diagnostic marker for oligodendrogliomas which account for less than 5% of gliomas.<sup>5</sup> The *isocitrate dehydrogenase 1* and 2 (*IDH1* and *IDH2*) mutations were associated with prolonged overall survival and higher rate of response to temozolamide in LGG.<sup>6</sup> In addition to genetic markers, recent studies identified gene profiles which enabled to classify patients into high- and low-risk groups with different survival probabilities.<sup>7,8</sup> Thus, identifying key molecular biomarkers is critical to the improvement of survival of LGG patients.

In this study, we aimed to screen for survival-related genes with Kaplan-Meier survival analysis and multivariate Cox regression analysis using the expression and clinical data of 506 LGG patients from The Cancer Genome Atlas (TCGA).<sup>9</sup> We also established a prognostic risk score based on a linear combination of 10 gene expression levels to effectively predict the overall survival (OS) and relapse-free survival (RFS) of LGG patients. The prognosis-associated genes and risk score were validated in 2 independent datasets from the Chinese Glioma Genome Atlas (CGGA) database.<sup>10</sup> The completion of our study opens the avenue for developing molecular markers in prognostication and treatment decision making for LGG patients.

## Methods and Materials

### Data Acquisition

The gene expression and clinical data came from 2 different sources. The TCGA dataset consisted of normalized read counts of 20532 genes and clinical information of 506 LGG patients and was obtained from the TCGA database. The CGGA dataset of 444 LGG patients was downloaded from the CGGA database. In order to further validate the survival analysis and differentially expressed genes, we also obtained gene expression and clinical data of another independent 182 LGG patients and 20 normal brain controls from the CGGA database (hereinafter referred to as validation dataset). All patients provided written informed consent prior to enrollment in the study.

### Bioinformatics Analysis of Prognosis-Associated Genes

To study the biological functions and possible signaling pathways of prognosis-associated genes, the enrichments of gene ontology (GO) terms<sup>11</sup> and Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>12</sup> pathways were analyzed with the online bioinformatics tool of gprofiler (version 6.8).<sup>13</sup> The raw P values were corrected by the g:SCS algorithm which is a tailor-

made algorithm for computing multiple testing correction for p-values gained from GO and pathway enrichment analysis.

### Survival Analyses

Student's t test was utilized to examine the associations between OS, RFS and quantitative factors in LGG patients. Fisher exact test was used to investigate the associations between OS, RFS and categorical variables. The associations of gene expression or risk score with OS were investigated with various statistical methods. In brief, LGG patients were split into the high and low expression groups according to the median expression values or risk score. Risk score = expression of gene 1  $\times$   $\beta_1$  + expression of gene 2  $\times$   $\beta_2$  + ... + expression of gene n  $\times$   $\beta_n$ .  $\beta$  values are the regression coefficients derived from the multivariate Cox regression analysis of the TCGA dataset. We used Kaplan-Meier survival analysis and log-rank methods to compare the difference of OS rates between the 2 groups using the survival package.<sup>14,15</sup> We performed multivariate Cox regression analysis to confirm whether gene expression or risk score was independent prognostic biomarkers after adjustment of the prognosis-related clinical factors using the survival package. ROC curve analysis was conducted by the R package of pROC to determine the predictive capabilities of prognosis-associated genes.<sup>16</sup> AUC values were computed accordingly.  $P < 0.05$  was defined as statistically significant. The prognosis-associated genes were then further classified into protective genes (odd ratio [OR]  $< 1$ ) and risk genes ( $0 < \text{OR} < 1$ ). We followed the same method to study the association between RFS and risk score. Linear regression model was used to study the association between risk score and each clinical factor in the TCGA and CGGA cohorts.

### Gene Set Enrichment Analysis and Differential Gene Expression Analysis

Gene set enrichment analysis (GSEA)<sup>17</sup> was implemented using a java software by comparing the high-risk (higher than median) group with low-risk group (lower than median). The default parameters were used. Transcripts per million (TPM) expression data of 523 LGG patients were obtained from the TCGA database. TPM expression data of 1141 normal brain tissues were obtained from The Genotype-Tissue Expression (GTEx) project.<sup>18</sup> Gene expression differences of top 10 prognosis-associated genes were compared by student t test between LGG patients and normal brain tissues. ROC curve analysis was conducted by the R package of pROC to determine the diagnostic values of prognosis-associated genes. AUC values were computed accordingly.

## Results

### The Associations Between OS, RFS and Clinicopathologic Factors in LGG

Patient's age, tumor weight, histological type, histologic grade, *IDH1* mutation and targeted therapy were significantly

**Table 1.** Association Between the Clinicopathologic Characteristics and Patients' OS as Well as RFS in the LGG Dataset.

Variables	Group	Alive	Dead	P value	Non-relapse	Relapse	P value	Statistical method
Age		41.68	49.06	≤0.001	42.09	43.62	0.29	Student t test
Tumor weight		338.38	285.01	0.02	331.86	328.05	0.89	Student t test
Gender	Female	179	45	0.24	123	60	0.83	Fisher's exact test
	Male	237	45		150	69		
History of cancer	No	183	27	0.34	122	49	0.3	Fisher's exact test
	Yes	107	22		76	40		
Histological type	Astrocytoma	151	41	0.17	94	54	0.26	Fisher's exact test
	Oligoastrocytoma	112	17		77	28		
	Oligodendroglioma	153	32		102	47		
Histologic grade	G2	217	26	≤0.001	140	57	0.20	Fisher's exact test
	G3	198	64		132	72		
<i>IDHI</i> mutation	Wild-type	79	34	≤0.001	44	41	≤0.001	Fisher's exact test
	Mutant	337	56		229	88		
<i>TP53</i> mutation	Wild-type	211	47	0.82	139	57	0.24	Fisher's exact test
	Mutant	205	43		134	72		
Radiation therapy	No	104	14	0.29	65	28	0.88	Fisher's exact test
	Yes	116	24		81	33		
Targeted therapy	No	167	22	0.03	124	46	0.08	Fisher's exact test
	Yes	199	48		145	80		

associated with OS in the TCGA cohort ( $P < 0.05$  for all cases, student's t test or Fisher exact test, Table 1). Moreover, histologic grade, *IDHI* mutation and 1p-19q codeletion exhibited significant associations with OS in the CGGA cohort ( $P < 0.05$ , Fisher exact test, supplementary Table 1). The remaining factors did not show significant associations with OS in LGG patients ( $P > 0.05$  for all cases, student's t test or Fisher exact test, Table 1).

With respect to the associations of clinicopathologic factors with RFS, LGG patients with *IDHI* mutation showed better RFS than those without *IDHI* mutation ( $P < 0.05$ , Fisher exact test, Table 1). While, histologic grade and chemotherapy were significantly associated with RFS in the LGG dataset ( $P > 0.05$  for all cases, Fisher exact test, student's t test, supplementary Table 1). No significant association was found between RFS and the remaining clinical factors ( $P > 0.05$  for all cases, student's t test or Fisher exact test, Table 1).

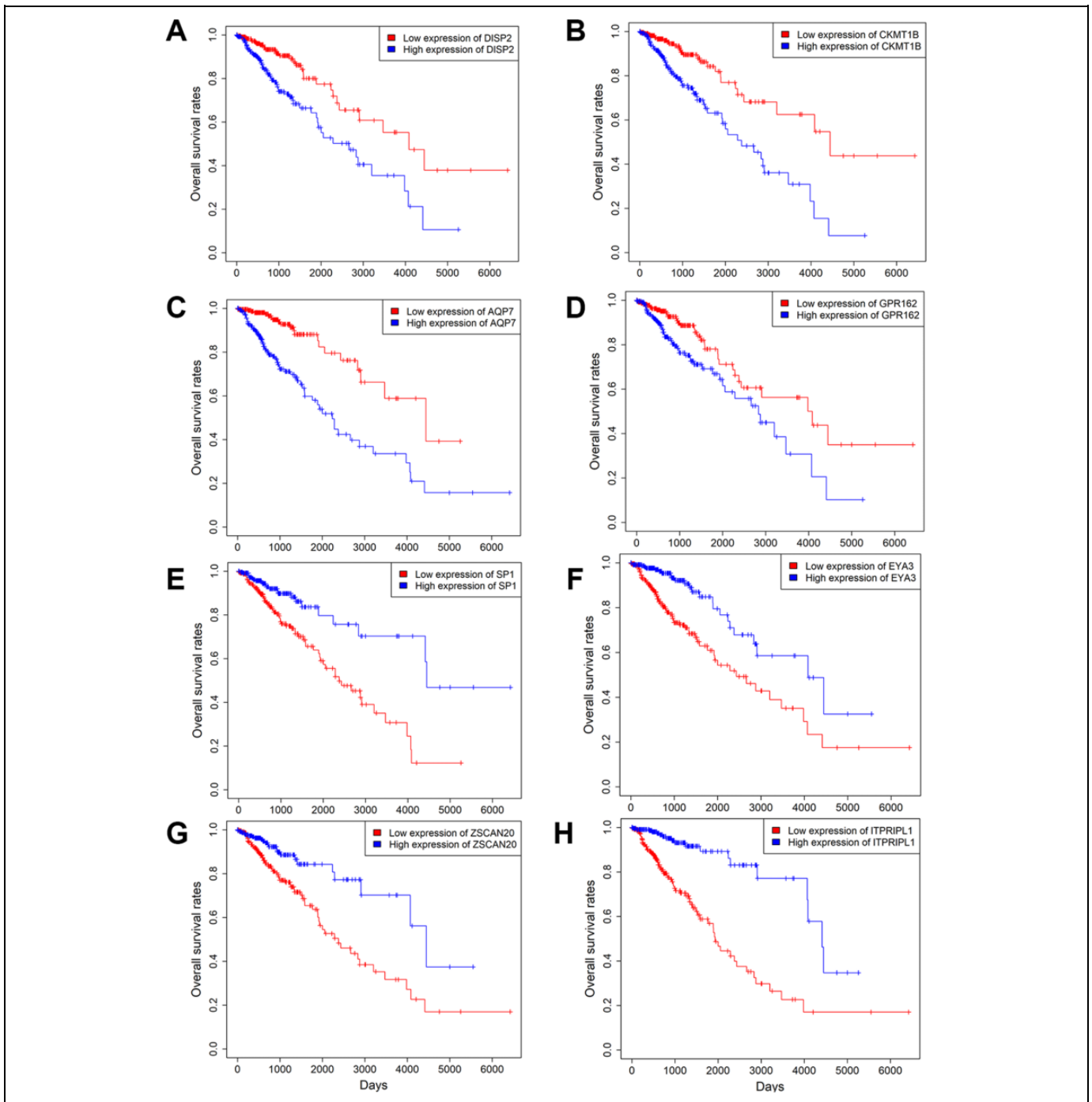
### Overall Survival Analysis

To evaluate the predictive capability of gene expression for patients' OS, the 506 LGG patients in the TCGA dataset were divided into low and high expression groups according to median expression values. Kaplan-Meier survival analysis showed that high expression levels of 3494 genes and 4181 genes were associated with favorable or poor prognosis respectively, such as dispatched RND transporter family member 2 (*DISP2*), creatine kinase, mitochondrial 1B (*CKMT1B*), aquaporin 7 (*AQP7*) and G protein-coupled receptor 162 (*GPR162*), Sp1 transcription factor (*SPI*), EYA transcriptional coactivator and phosphatase 3 (*EYA3*), zinc finger and SCAN domain containing 20 (*ZSCAN20*), ITPRIP like 1 (*ITPRIPL1*) ( $P < 0.05$  for all cases, log rank test, Figure 1, supplementary Figure 1). Then multivariate Cox regression analysis was performed between OS and

the mortality-associated features, including patients' age, tumor weight, histological type, histologic grade and *IDHI* mutation and 7675 gene expression levels. Multivariate Cox regression analysis confirmed that high expression levels of 1561 genes were significantly associated with decreased mortality, such as *DISP2*, *CKMT1B*, *AQP7*, *GPR162*. The hazard ratios of the 4 genes ranged from 0.22 to 0.27, with a mean of 0.24 ( $P < 0.05$  for all cases, supplementary Table 2). While high expression levels of 1812 genes were associated with increased mortality, such as *SPI*, *EYA3*, *ZSCAN20*, *ITPRIPL1*. The hazard ratio ranged from 4.77 to 5.73 for the 4 genes, with a mean of 5.12 ( $P < 0.05$  for all cases, supplementary Table 2, supplementary Figure 1).

### Validation of Survival Analyses in the CGGA Dataset

In order to validate the findings above, the associations between 3373 gene expression levels and mortality were evaluated in 444 LGG samples of the CGGA dataset. Of 3373 prognosis-associated genes, Kaplan-Meier survival analysis confirmed that high expression levels of 666 genes were associated with favorable prognosis in LGG. In contrast, high expression levels of 1433 genes were associated with poor prognosis ( $P < 0.05$  for all cases, log rank test, supplementary Figure 1). Then multivariate Cox regression analysis was performed between OS and the mortality-associated features, including histologic grade, *IDHI* mutation and 1p19q codeletion and 2099 gene expression levels. Multivariate Cox regression analysis confirmed that high expression levels of 421 genes were associated with decreased mortality, while high expression levels of 1068 genes were associated with increased mortality. The top 4 protective genes *DISP2*, *CKMT1B*, *AQP7*, *GPR162* and top 4 risk genes *SPI*, *EYA3*, *ZSCAN20*, *ITPRIPL1* were confirmed to be significantly associated with OS in the



**Figure 1.** Kaplan-Meier survival analysis of patients' OS with DISP2, CKMT1B, AQP7, GPR162, SP1, EYA3, ZSCAN20, *ITPRIPL1* (A-G) expression levels in the TCGA cohort. The difference in overall survival was significantly different between high expression and low expression groups ( $P < 0.001$  for all cases, log rank test).

CGGA dataset ( $P < 0.05$  for all cases, supplementary Figure 1, supplementary Table 3).

### Bioinformatics Analysis of Prognosis-Associated Genes

The survival-related genes were further analyzed for their involvement in pathways and possible biological roles. The

GO term enrichment analysis indicated that the 1068 risk genes were mainly enriched in 236 GO terms, such as DNA replication, cell cycle, mitotic spindle checkpoint, extrinsic apoptotic signaling pathway via death domain receptors, regulation of cyclin-dependent protein serine/threonine kinase activity, integrin-mediated signaling pathway (Supplementary Table 4, adjusted P values  $< 0.05$  for all cases). The 421

**Table 2.** Multivariate Analyses Between OS and the Risk Score in the TCGA and CGGA Datasets.

TCGA dataset				CGGA dataset			
Clinical feature	HR	95%CI	P value	Clinical feature	HR	95%CI	P value
Age	1.07	1.04-1.09	<0.001	Histologic grade	3.19	2.23-4.58	<0.001
Histologic grade	2.51	1.3-4.85	0.01	IDH1	0.56	0.39-0.79	<0.001
Tumor weight	1	0.99-1	0.13	1p-19q codeletion	0.49	0.33-0.74	<0.001
IDH1	0.51	0.28-0.91	0.02	Risk score	1.89	1.36-2.62	<0.001
Targeted therapy	0.61	0.33-1.13	0.12				
Risk score	8.3	3.68-18.7	<0.001				

Notably, HR and CI refer to hazard ratio and confidence interval respectively.

protective genes were significantly enriched in 11 GO terms, such as trans-synaptic signaling, chemical synaptic transmission, nervous system development, neurotransmitter transport. The KEGG pathway enrichment analysis suggested that 1068 risk genes were significantly enriched in the DNA replication, cell cycle and homologous recombination, while the 421 protective genes were significantly enriched in nicotine addiction, synaptic vesicle cycle, neuroactive ligand-receptor interaction (Supplementary Table 5, adjusted P values < 0.05 for all cases).

### Risk Score is a Negative Predictor for Overall Survival in LGG

We developed a risk score formula by linear combination of expression values of top 5 protective and top 5 risk genes using the coefficients of the multivariate Cox regression models of the TCGA dataset. Risk score =  $0.25 \times \text{expression of } DISP2 + 0.22 \times \text{expression of } CKMT1B + 0.23 \times \text{expression of } AQP7 + 0.27 \times \text{expression of } GPR162 + 0.26 \times \text{expression of } CHGB + 4.99 \times \text{expression of } SPI1 + 5.73 \times \text{expression of } EYA3 + 4.77 \times \text{expression of } ZSCAN20 + 5.13 \times \text{expression of } ITPRIPL1 + 4.58 \times \text{expression of } ZNF217$ . Deceased LGG patients showed significantly higher risk scores than living LGG patients in both cohorts ( $P < 0.05$  for all cases, student t test, Figure 2A). LGG patients could be divided into the low-risk group and high-risk group according to the median risk score. Kaplan-Meier survival analysis suggested that LGG patients with high risk scores showed lower survival probabilities than those with low risk scores in the TCGA dataset ( $P < 0.05$  for all cases, Figure 2B). The multivariate Cox regression analysis further confirmed that risk score was a negative factor for predicting overall survival independently of prognosis-related clinical features in LGG. Using the same methods, the negative association between risk score and overall survival was also validated in the CGGA dataset ( $P < 0.05$  for all cases, Table 2 and Figure 2C). ROC curves were plotted to further explore the predictive capabilities of the risk scores, the AUC values were 73.67% and 66% in the TCGA and CGGA cohorts respectively, suggesting the risk score is good predictor of OS in LGG (Figure 2D).

### Risk Score is a Negative Predictor for Relapse-Free Survival in LGG

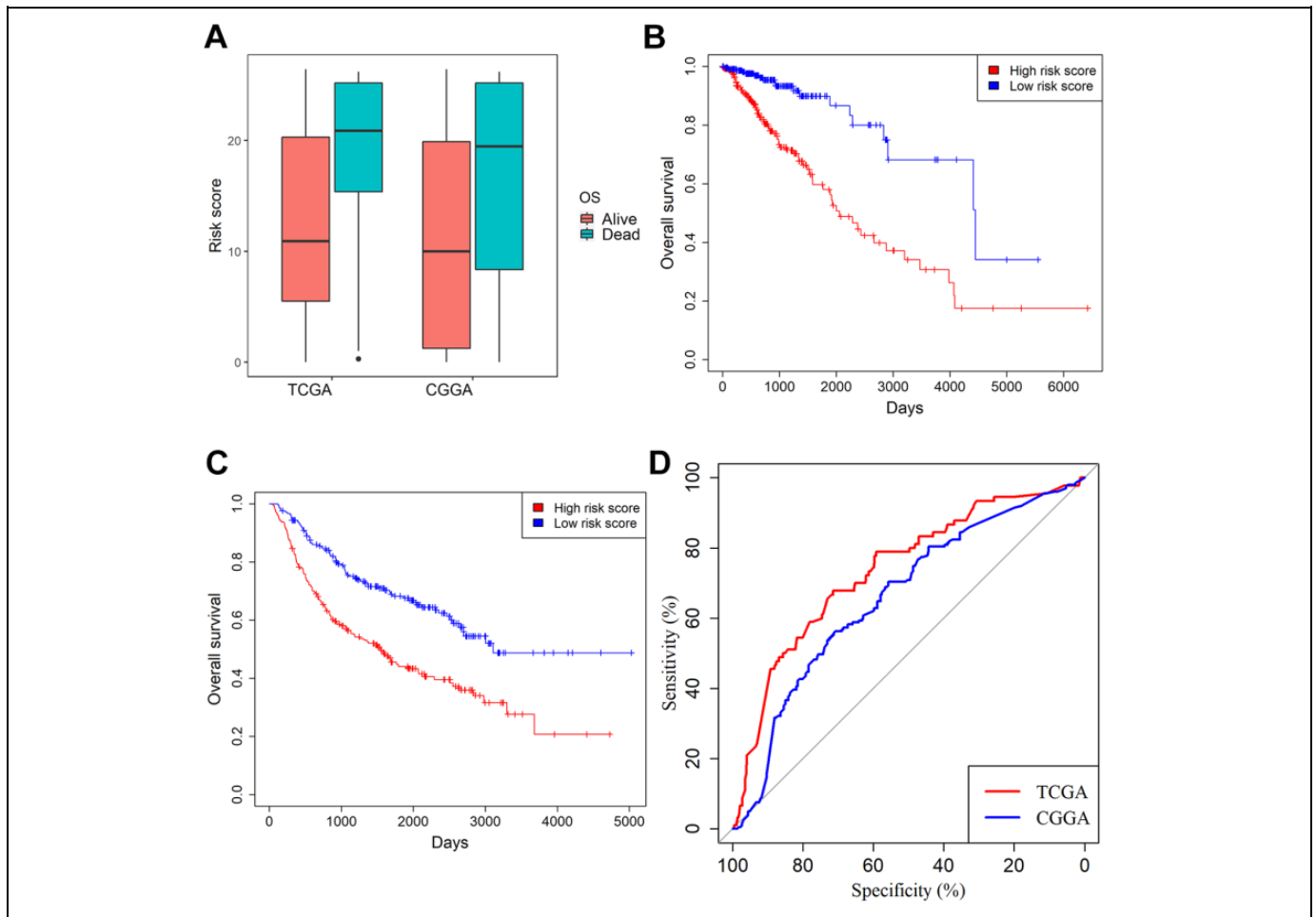
In order to further characterize the associations of risk scores with RFS, we followed the overall survival analysis methods. Recurrent tumors showed significantly higher risk scores than non-recurrent tumors in both cohorts ( $P < 0.05$  for all cases, student t test, Figure 3A). Kaplan-Meier RFS analysis showed that high risk score was associated with an inferior RFS ( $P < 0.05$ , log rank test, Figure 3B). Multivariate Cox regression analysis exhibited that increased risk score was associated with increased RFS after the adjustment of IDH mutation ( $P < 0.05$ , supplementary Table 6). In order to validate the findings above, we analyzed the associations between RFS and risk score in the CGGA dataset. The Kaplan-Meier analysis together with multivariate analysis confirmed that increased risk score was associated with increased RFS ( $P < 0.05$ , Figure 3C, supplementary Table 6). ROC curves were plotted to further explore the predictive capabilities of the risk scores, the AUC values were 61.52% and 53.9% in the TCGA and CGGA cohorts respectively, suggesting the risk score is an effective predictor for RFS in LGG (Figure 3D).

### Risk Score Is Associated With Clinical Factors in LGG

Linear regression model was used to analyze the association between risk score and each clinical factor in the TCGA and CGGA cohorts. In the TCGA cohort, risk score was significantly positively correlated with histologic grade, *TP53* mutation, radiation therapy, targeted molecular therapy and negatively correlated with *IDH1* mutation and histological type ( $p < 0.05$  for all cases, supplementary Figure 2A). Furthermore, the risk score exhibited significantly negative correlation with *IDH1* mutation and 1p19q codeletion ( $p < 0.05$  for all cases, supplementary Figure 2B).

### Risk Score Related Pathway Analysis

In order to understand why risk score is predictive of LGG patients' survival, we partitioned the LGG samples into high and low risk groups according to the median risk score. GSEA was performed to analyze the altered signaling pathways between the 2 different risk groups. Nine signaling pathways



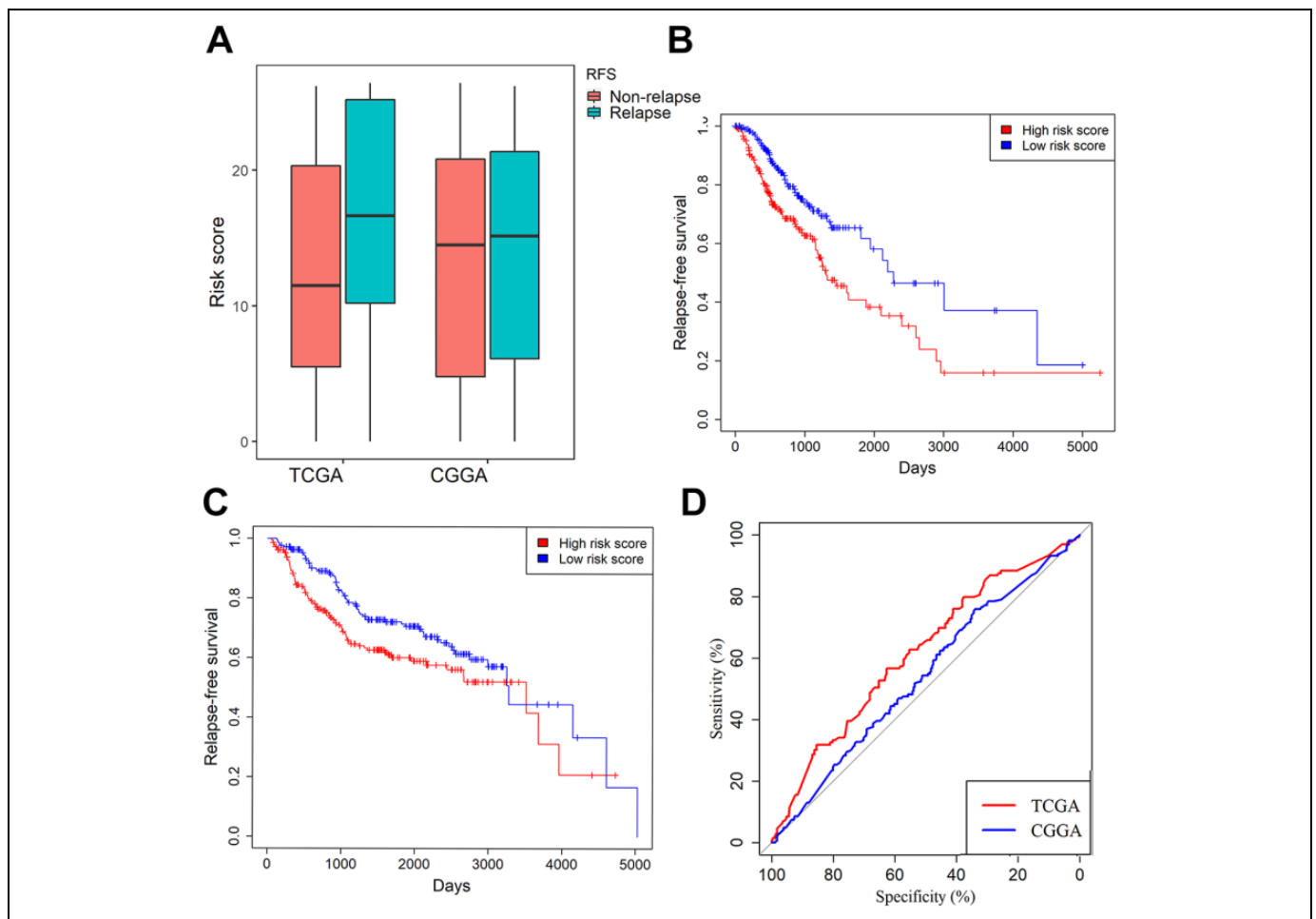
**Figure 2.** Risk score is negatively associated with OS in LGG. A. The difference of risk scores between the living and deceased LGG patients in the 2 cohorts ( $P < 0.05$  for all cases, student t test). B. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the TCGA cohort. C. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the CGGA cohort. The difference in overall survival was significantly different between high and low risk score groups ( $P < 0.001$  for all cases, log rank test). D. The ROC curves for the risk scores in the TCGA and CGGA cohorts.

were significantly enriched in the high risk group of both the TCGA and CGGA cohorts, with small cell lung cancer, pathways in cancer, chronic myeloid leukemia, colorectal cancer the top 4 pathways most significantly enriched in the high risk score group (Figure 4 and supplementary Figure 3,  $p < 0.05$  for all cases). In contrast, genes in the pathways of cardiac muscle contraction, oxidative phosphorylation were significantly enriched in the low risk group of the 2 cohorts (Figure 4 and supplementary Figure 4,  $p < 0.05$  for all cases). These results suggest that the survival of glioma patients can be accurately predicted by the risk score, perhaps the above-mentioned pathways might play a critical role in the association of risk score with survival.

#### Independent Validation of Survival Analysis in the Validation Dataset

We also validated the prognostic importance of the 10 genes and risk score in the validation dataset from the CGGA

database. The Kaplan-Meier analysis validated *DISP2*, *CKMT1B*, *AQP7*, *GPR162* and *CHGB* expression levels were positively correlated with overall survival, while *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1* and *ZNF217* expression levels were negatively correlated with OS in the validation dataset ( $p < 0.05$  for all cases, log rank test, supplementary Table 7). Then multivariate analysis was carried out between OS and the mortality-associated features, including histologic grade, *IDH1* mutation and 1p19q codeletion and 10 gene expression levels. Multivariate analysis confirmed *DISP2*, *CKMT1B*, *AQP7* and *CHGB* expression levels were significantly associated with favorable OS after adjustment of survival-related features, while *SP1*, *EYA3* expression levels were significantly associated with inferior OS ( $p < 0.05$  for all cases, supplementary Table 7). We also calculated the risk score for each LGG patient using the risk score formula as mentioned above and evaluated its relation with OS and RFS in the validation cohort. As expected, the risk score was significantly adversely associated with OS and RFS ( $P < 0.05$  for all cases, log rank test,



**Figure 3.** Risk score is negatively associated with RFS in LGG. A. The difference of risk scores between the LGG patients with relapse and non-relapse in the 2 cohorts ( $P < 0.05$  for all cases, student t test). B. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the TCGA cohort. C. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the CGGA cohort. The difference in overall survival was significantly different between high and low risk score groups ( $P < 0.05$  for all cases, log rank test). D. The ROC curves for the risk scores in the TCGA and CGGA cohorts.

supplementary Figure 5). Moreover, the negative association was further verified in the multivariate analysis in the validation dataset ( $P < 0.05$  for all cases, supplementary Table 8).

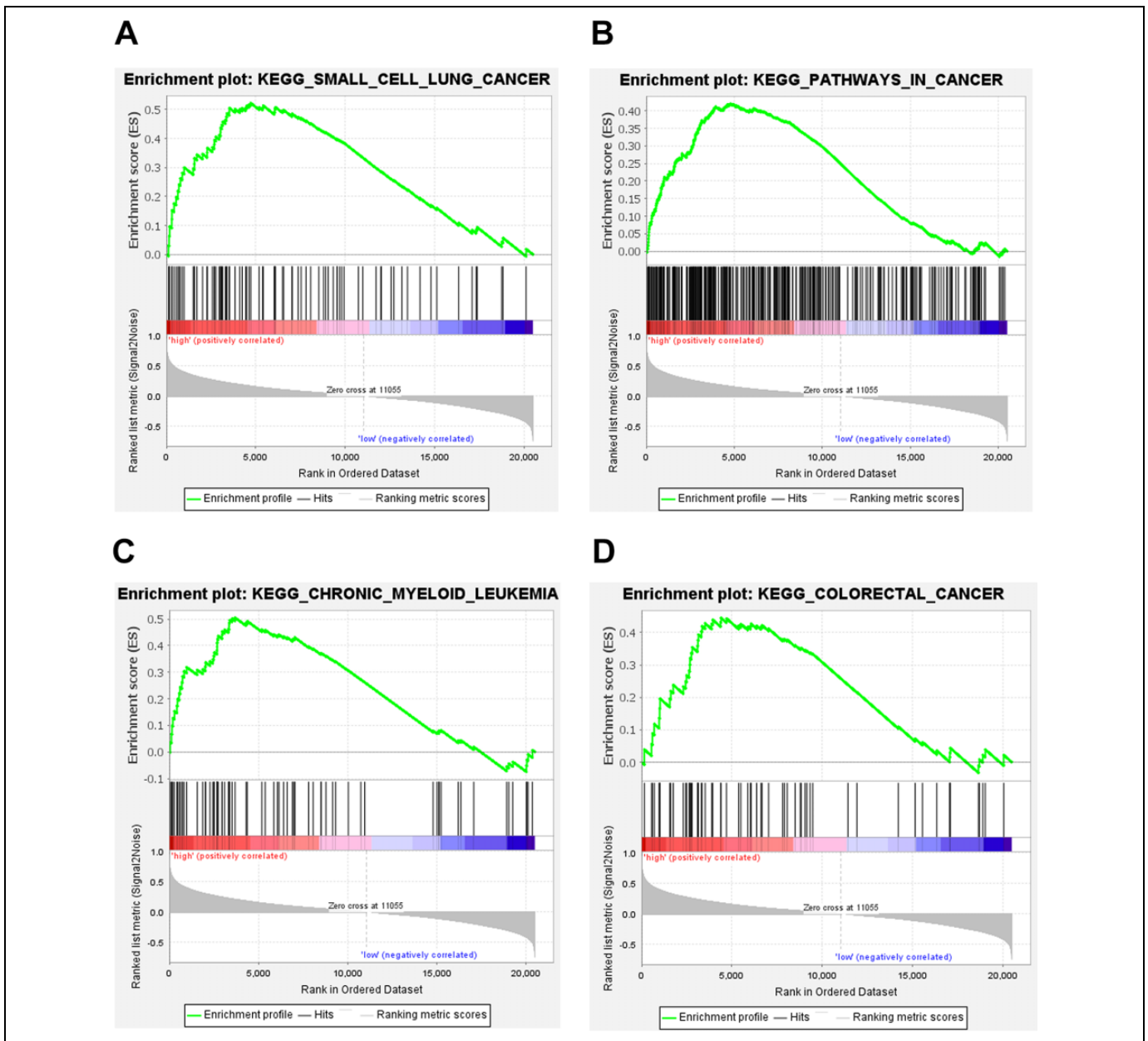
### Assessment of Diagnostic Value

Of the top 10 prognosis-associated genes, *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* and *GPR162* were significantly up-regulated, while *DISP2*, *CKMT1B*, *AQP7* were down-regulated in 523 LGG tissues as compared to 1141 normal brain controls ( $P$  values  $< 0.001$  for all cases, student t test, Figure 5A). ROC curves were plotted to further explore the diagnostic values of the 10 genes. Three genes, *SP1*, *ZSCAN20*, *ZNF217*, exhibited high accuracy in differentiating LGG tissues from brain normal tissues ( $P$  values  $< 0.05$ ,  $AUC > 0.8$  for all cases, Figure 5B). We also validated the gene expression difference of the 10 genes between 182 LGG patients and 20 normal brain controls from the CGGA database. *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* were confirmed to be

significantly up-regulated, while *CKMT1B* was down-regulated in LGG ( $P$  values  $< 0.001$  for all cases, student t test, Figure 5C).

### Discussion

Over the past decade, despite significant advances in the treatment of LGG, LGG remains a serious threat to public health around the world. A number of studies have demonstrated the associations of molecular tumor markers with overall survival for LGG patients. For instance, the 1p-19q deletion is a powerful predictor of chemotherapy response and survival and serves as a diagnostic marker for oligodendrogliomas which account for less than 5% of gliomas.<sup>5</sup> The *IDH1* and *IDH2* mutations are associated with prolonged overall survival and higher rate of response to temozolomide in low-grade gliomas.<sup>6</sup> It remains critical and urgent to identify new molecular biomarkers and develop prognostic prediction models with high accuracy. In recent years, gene-signatures based on aberrant



**Figure 4.** KEGG pathways associated with risk score. GSEA based on the expression of TCGA dataset revealed significant pathways associated with risk score, including small cell lung cancer (A), pathways in cancer(B), chronic myeloid leukemia (C), colorectal cancer (D) ( $P < 0.05$  for all cases, GSEA). For each gene set, vertical bars along the x-axis of the GSEA plot represent the positions of genes within the ranked list. Negative GSEA enrichment score curve indicates anti-enrichment (down-regulation), and positive curve denotes enrichment (up-regulation).

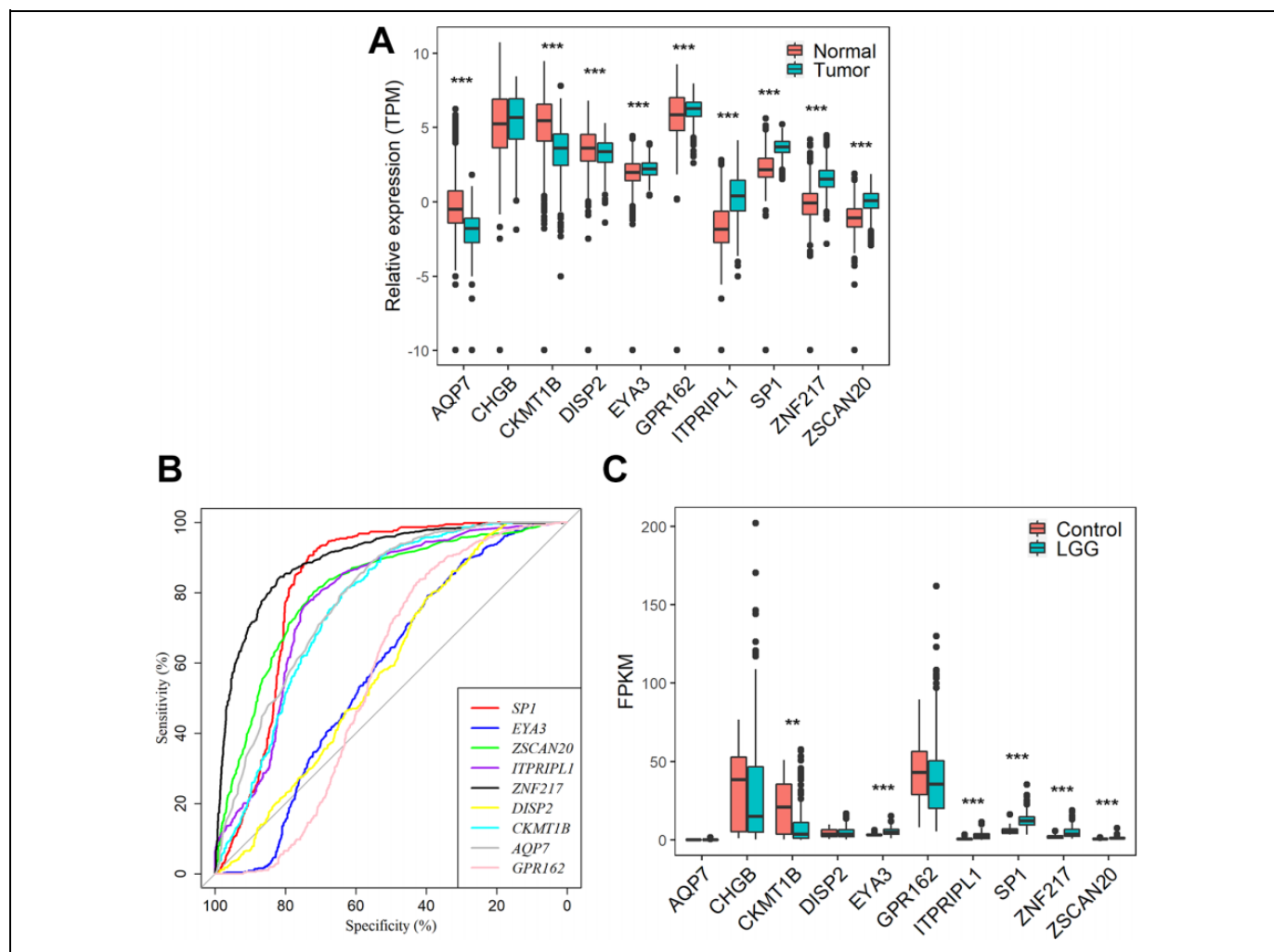
mRNA have drawn much attention and displayed great potential in prognostic prediction of LGG patients.<sup>19-21</sup>

In this study, we found 1489 genes expression was significantly associated with OS of LGG patients. The GO term and KEGG pathway enrichment analyses indicated that the 1068 risk genes were mainly enriched in the regulation of DNA replication, cell cycle, DNA repair signaling pathways. While, the 421 protective genes were significantly enriched in the nicotine addiction, synaptic vesicle cycle, neuroactive ligand-receptor interaction. The difference in the GO terms and KEGG pathway of prognosis-associated genes is of great importance

to the identification of prognostic biomarkers and eventually novel therapeutic targets in LGG. For instance, given the significant enrichment of risk genes in the cell cycle signaling pathway, cell cycle genes may become potential candidates for developing prognostic biomarkers or druggable targets in LGG.

We also identified a 10-gene panel comprising *DISP2*, *CKMT1B*, *AQP7*, *GPR162*, *CHGB*, *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* expression levels that could predict the OS of LGG patients. Furthermore, we established a risk score model using a linear combination of 10 gene expression levels





**Figure 5.** Differential expression gene analysis A. The expression difference of *DISP2*, *CKMT1B*, *AQP7*, *GPR162*, *CHGB*, *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* between 523 LGG tissues and 1141 normal brain controls in the TCGA cohort. B. The ROC curves for the 9 genes in the TCGA cohort. C. The expression difference of *DISP2*, *CKMT1B*, *AQP7*, *GPR162*, *CHGB*, *SP1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* between 325 LGG tissues and 20 normal brain controls in the CGGA cohort. \*\* and \*\*\* represent P value < 0.01 and 0.001 respectively, student t test was used for the comparison.

weighted by  $\beta$ -values of multivariate Cox regression models. The risk score was negatively correlated with OS and RFS after adjustment of known prognosticators. The clinical significance of risk score was evaluated and KEGG pathways were identified. Nine signaling pathways were significantly enriched in the high risk group of both the TCGA and CGGA cohorts, including adherens junction, basal transcription factors, chronic myeloid leukemia, colorectal cancer, neurotrophin signaling pathway, pathways in cancer, small cell lung cancer, TGF-beta signaling pathway and Ubiquitin mediated proteolysis. Neurotrophins are a family of trophic factors involved in differentiation and survival of neural cells.<sup>12</sup> We believe neurotrophin signaling pathway and other pathways are implicated in the molecular mechanism by which risk score is associated with survival of LGG patients.

The 10 genes play diverse roles in the tumorigenesis of cancers. For instance, the Specificity protein 1 (*Sp1*)

transcription factor is known to regulate tumorigenesis in many cancers. *Sp1* is over-expressed and associated with poor prognosis in many cancer types. Inhibiting *Sp1* expression suppresses the proliferation, metastasis and induces apoptosis in colon cancer,<sup>22</sup> glioma,<sup>23,24</sup> pharyngeal carcinoma<sup>25</sup> and lung cancer.<sup>26</sup> E2F suppression and *Sp1* overexpression induce the differentiation-specific marker, transglutaminase type 1, in a squamous cell carcinoma cell line KJD-1/SV40.<sup>27</sup> *ZNF217* is a member of the Kruppel-like family of transcriptional factors, it was firstly amplified in breast cancer and associated with aggressive tumor behavior and poor clinical prognosis.<sup>28,29</sup> *ZNF217* expression was increased in prostate cancer, Elevated expression of *ZNF217* promotes prostate cancer growth by restraining ferroportin-conducted iron egress.<sup>30</sup> *ZNF217* is a crucial mediator and indicator of bone metastasis. The *ZNF217* promotes breast cancer metastasis to the bone.<sup>31</sup> *ZNF217* gene amplification was significantly correlated with lymph node

metastasis in ovarian clear cell carcinoma. Inhibition of *ZNF217* profoundly suppressed cell migration and invasion in ovarian clear cell carcinoma.<sup>32</sup> *ZNF217* expression was associated with inferior relapse-free survival and overall survival in gastric cancer<sup>33</sup> and breast cancer.<sup>34</sup> *CKMT1B* expression has been shown to be associated with short disease-free survival and relapse in breast cancer, which is discordant with our study.<sup>35</sup> Low expression of *AQP7* is correlated with tumor grade and the aggressive features of hepatocellular carcinoma.<sup>36</sup> Moreover, *AQP7* is prognostic of overall survival in patients with breast cancer. In mouse breast cancer models, reduced expression of *Aqp7* caused reduced primary tumor burden and lung metastasis.<sup>37</sup> *Eya3* promotes triple-negative breast cancer-associated immune suppression. *Eya3* loss decreases tumor growth in immune-competent mice and is associated with increased numbers of infiltrated CD8+ T cells.<sup>38</sup> Silence of *EYA3* increased H2A.XY<sup>142</sup>ph and inhibited cell viability, migration and percent cells in S stage in gastric cancer<sup>39</sup> and decreased survival of Ewing sarcoma cells.<sup>40</sup> These studies and the results in our study suggest that the 5 genes may have oncogenic or tumor suppressor function in cancers.

In addition to prognostic value, the 10 genes also showed diagnostic value for LGG patients. Our study reported *DISP2*, *CKMT1B*, *AQP7* were significantly down-regulated and *SPI1*, *EYA3*, *ZSCAN20*, *ITPRIPL1*, *ZNF217* and *GPR162* were significantly up-regulated in the LGG samples. They exhibited high accuracy in differentiating LGG tissues from normal brain tissues. Lastly, the 10 genes may also pave the way for developing targeted therapies for LGG patients. For instance, silencing the expression of *Sp1* resulted in marked decrease in cell proliferation, metastasis in various cancer types.<sup>22,26,32</sup>

## Conclusion

In summary, this study revealed a novel 10-gene signature which has prognostic and diagnostic values and successfully classifies LGG patients with different prognostic probabilities. A higher risk score indicates a poorer prognosis. These findings will help researchers identify new treatments for LGG and to provide more therapeutic targets to cure LGG patients in the future.


## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Maokai Wang  <https://orcid.org/0000-0003-3691-4616>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Ostrom QT, Gittleman H, Farah P, et al. CBTRUS Statistical Report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol.* 2013; 12(1):28-36. <http://neuro-oncology.oxfordjournals.org/content/12/5/500.abstract>
- Chang L, Su J, Jia X, Ren H. Treating malignant glioma in Chinese patients: update on temozolomide. *Onco Targets Ther.* 2014; 7:235-244. doi:10.2147/OTT.S41336
- Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuro-pathol.* 2007;114(2):97-109. doi:10.1007/s00401-007-0243-4
- Claus EB, Walsh KM, Wiencke JK, et al. Survival and low-grade glioma: the emergence of genetic information. *Neurosurg Focus.* 2015;38(1):1-10. doi:10.3171/2014.10.FOCUS12367. DISCLOSURE
- Gadji M, Fortin D, Tsanaclis AM, Drouin R. Is the 1p/19q deletion a diagnostic marker of oligodendrogliomas? *Cancer Genet Cytogenet.* 2009;194(1):12-22. doi:10.1016/j.cancergencyto.2009.05.004
- Houillier C, Wang X, Kaloshi G, et al. IDH1 or IDH2 mutations predict longer survival and response to temozolomide in low-grade gliomas. *Neurology.* 2010;75(17):1560-1566. doi:10.1212/WNL.0b013e3181f96282
- Hsu JBK, Chang TH, Lee GA, Lee TY, Chen CY. Identification of potential biomarkers related to glioma survival by gene expression profile analysis. *BMC Med Genomics.* 2019;11(Suppl 7): 1-18. doi:10.1186/s12920-019-0479-6
- Wang Y, Liu X, Guan G, Xiao Z, Zhao W, Zhuang M. Identification of a five-pseudogene signature for predicting survival and its ceRNA network in Glioma. *Front Oncol.* 2019;9(October): 1059. doi:10.3389/fonc.2019.01059
- Chang K, Creighton CJ, Davis C, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet.* 2013;45(11):1113-1120. doi:10.1038/ng.2764
- Wang Y, Qian T, You G, et al. Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. *Neuro Oncol.* 2015;17(2): 282-288. doi:10.1093/neuonc/nou130
- Consortium TGO. Gene ontology: tool for the identification of biology. *Nat Genet.* 2000;25(may):25-29.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999;27(1):29-34. doi:10.1093/nar/27.1.29
- Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G: Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 2007;35(SUPPL.2): 193-200. doi:10.1093/nar/gkm226
- Therneau T. Survival analysis. *Cran.* Published online 2016. doi: 10.1007/978-1-4419-6646-9
- Fox J. Cox proportional-hazards regression for survival data the cox proportional-hazards model. *Most.* 2002;2008(June):1-18. doi:10.1016/j.carbon.2010.02.029
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves.

- BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
17. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005; 102(43):15545-15550. doi:10.1073/pnas.0506580102
  18. Consortium GTE. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653
  19. Chen WL, Wang YY, Zhao A, et al. Enhanced fructose utilization mediated by SLC2A5 is a unique metabolic feature of acute myeloid leukemia with therapeutic potential. *Cancer Cell*. 2016; 30(5):779-791. doi:10.1016/j.ccell.2016.09.006
  20. Zhang YA, Zhou Y, Luo X, et al. SHOX2 is a potent independent biomarker to predict survival of WHO Grade II-III Diffuse Gliomas. *EBioMedicine*. 2016;13:80-89. doi:10.1016/j.ebiom.2016.10.040
  21. Chen B, Liang T, Yang P, et al. Classifying lower grade glioma cases according to whole genome gene expression. *Oncotarget*. 2016;7(45):74031-74042. doi:10.18632/oncotarget.12188
  22. Zhao Y, Zhang W, Guo Z, et al. Inhibition of the transcription factor Sp1 suppresses colon cancer stem cell growth and induces apoptosis in vitro and in nude mouse xenografts. *Oncol Rep*. 2013;30(4):1782-1792. doi:10.3892/or.2013.2627
  23. Zhao T, Wang H, Ma H, Wang H, Chen B, Deng Y. Starvation after cobalt-60 $\gamma$ -ray radiation enhances metastasis in U251 glioma cells by regulating the transcription factor SP1. *Int J Mol Sci*. 2016;17(4):386. doi:10.3390/ijms17040386
  24. Chen YT, Tsai HP, Wu CC, Chen CY, Chai CY, Kwan AL. High-level sp1 is associated with proliferation, invasion, and poor prognosis in astrocytoma. *Pathol Oncol Res*. 2019;25(3):1003-1013. doi:10.1007/s12253-018-0422-8
  25. Zhang JP, Zhang H, Wang HB, et al. Down-regulation of Sp1 suppresses cell proliferation, clonogenicity and the expressions of stem cell markers in nasopharyngeal carcinoma. *J Transl Med*. 2014;12(1):1-12. doi:10.1186/s12967-014-0222 -1
  26. Muramoto K, Tange R, Ishii T, Miyauchi K, Sato T. Downregulation of transcription factor Sp1 suppresses malignant properties of A549 human lung cancer cell line with decreased  $\beta$ 4-galactosylation of highly branched N-glycans. *Biol Pharm Bull*. 2017;40(8):1282-1288. doi:10.1248/bpb.b17-00212
  27. Wong CF, Barnes LM, Dahler AL, et al. E2F suppression and Sp1 overexpression are sufficient to induce the differentiation-specific marker, transglutaminase type 1, in a squamous cell carcinoma cell line. *Oncogene*. 2005;24(21):3525-3534. doi:10.1038/sj.onc.1208372
  28. Nonet GH, Stampfer MR, Chin K, Gray JW, Collins CC, Yaswen P. The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. *Cancer Res*. 2001;61(4):1250-1254.
  29. Ollins COC, Ommens JOMR, Owbel DAK, et al. Positional cloning of ZNF 217 and NABC 1: genes amplified at 20q13. 2 and overexpressed in breast carcinoma. *Proc Natl Acad Sci U S A*. 2000;95(July 1998):8703-8708.
  30. Jiang X, Zhang C, Qi S, et al. Elevated expression of ZNF217 promotes prostate cancer growth by restraining ferroportin-conducted iron egress. *Oncotarget*. 2016;7(51):84893-84906. doi:10.18632/oncotarget.12753
  31. Bellanger A, Donini CF, Vendrell JA, et al. The critical role of the ZNF217 oncogene in promoting breast cancer metastasis to the bone. *J Pathol*. 2017;242(1):73-89. doi:10.1002/path.4882
  32. Rahman MT, Nakayama K, Rahman M, et al. Gene amplification of ZNF217 located at chr20q13.2 is associated with lymph node metastasis in ovarian clear cell carcinoma. *Anticancer Res*. 2012; 32(8):3091-3095.
  33. Shida A, Fujioka S, Kurihara H, et al. Prognostic significance of ZNF217 expression in gastric carcinoma. *Anticancer Res*. 2014; 34(9):4813-4817.
  34. Vendrell JA, Thollet A, Nguyen NT, et al. ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion. *Cancer Res*. 2012;72(14): 3593-3606. doi:10.1158/0008-5472.CAN-11-3095
  35. Cimino D, Fuso L, Sfiligoi C, et al. Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. *Int J Cancer*. 2008;123(6):1327-1338. doi:10.1002/ijc.23660
  36. Chen XF, Li CF, Lü L, Mei ZC. Expression and clinical significance of aquaglyceroporins in human hepatocellular carcinoma. *Mol Med Rep*. 2016;13(6):5283-5289. doi:10.3892/mmr.2016.5184
  37. Dai C, Charlestin V, Wang M, et al. Aquaporin-7 regulates the response to cellular stress in breast cancer. *Cancer Res*. 2020; 80(19):4071-4086. doi:10.1158/0008-5472.CAN-19-2269
  38. Vartuli RL, Zhou H, Zhang L, et al. Eya3 promotes breast tumor-associated immune suppression via threonine phosphatase-mediated PD-L1 upregulation. *J Clin Invest*. 2018;128(6): 2535-2550. doi:10.1172/JCI96784
  39. Dong C, Sun J, Ma S, Zhang G. K-ras-ERK1/2 down-regulates H2A.X(Y142 ph) through WSTF to promote the progress of gastric cancer. *BMC Cancer*. 2019;19(1):530. doi:10.1186/s12885-019-5750-x
  40. Robin TP, Smith A, McKinsey E, Reaves L, Jedlicka P, Ford HL. EWS/FLI1 regulates EYA3 in Ewing sarcoma via modulation of miRNA-708, resulting in increased cell survival and chemoresistance. *Mol Cancer Res*. 2012;10(8):1098-1108. doi:10.1158/1541-7786.MCR-12-0086