



# Construction and Verification of a Glycolysis-Associated Gene Signature for the Prediction of Overall Survival in Low Grade Glioma

Wei Liu<sup>1†</sup>, Chunshan Liu<sup>1†</sup>, Chengcong Chen<sup>1†</sup>, Xiaoting Huang<sup>1</sup>, Qi Yi<sup>1</sup>, Yunhong Tian<sup>1\*</sup>, Biao Peng<sup>2\*</sup> and Yawei Yuan<sup>1\*</sup>

<sup>1</sup>Department of Radiation Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China,

<sup>2</sup>Department of Neurosurgery, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China

## OPEN ACCESS

### Edited by:

Li Xing,  
University of Saskatchewan, Canada

### Reviewed by:

Longbo Zhang,  
Yale University, United States  
Mingjie Wang,  
Shanghai Jiao Tong University, China

### \*Correspondence:

Yunhong Tian  
tianyuhong2020@gzhmu.edu.cn  
Biao Peng  
2747617668@qq.com  
Yawei Yuan  
yuanyawei@gzhmu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

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

Received: 26 December 2021

Accepted: 07 March 2022

Published: 23 March 2022

### Citation:

Liu W, Liu C, Chen C, Huang X, Yi Q,  
Tian Y, Peng B and Yuan Y (2022)  
Construction and Verification of a  
Glycolysis-Associated Gene Signature  
for the Prediction of Overall Survival in  
Low Grade Glioma.  
Front. Genet. 13:843711.  
doi: 10.3389/fgene.2022.843711

The overall survival of patients with lower grade glioma (LGG) that might develop into high-grade malignant glioma shows marked heterogeneity. The currently used clinical evaluation index is not sufficient to predict precise prognostic outcomes accurately. To optimize survival risk stratification and the personalized management of patients with LGG, there is an urgent need to develop an accurate risk prediction model. The TCGA-LGG dataset, downloaded from The Cancer Genome Atlas (TCGA) portal, was used as a training cohort, and the Chinese Glioma Genome Atlas (CGGA) dataset and Rembrandt dataset were used as validation cohorts. The levels of various cancer hallmarks were quantified, which identified glycolysis as the primary overall survival-related risk factor in LGGs. Furthermore, using various bioinformatic and statistical methods, we developed a strong glycolysis-related gene signature to predict prognosis. Gene set enrichment analysis showed that in our model, high-risk glioma correlated with the chemoradiotherapy resistance and poor survival. Moreover, based on established risk model and other clinical features, a decision tree and a nomogram were built, which could serve as useful tools in the diagnosis and treatment of LGGs. This study indicates that the glycolysis-related gene signature could distinguish high-risk and low-risk patients precisely, and thus can be used as an independent clinical feature.

**Keywords:** low-grade glioma, glycolysis, signature, prognosis, risk model

## INTRODUCTION

Glioma is the most common intracranial malignant tumor, which accounts for 80% of all intracranial malignancies, with 15,000–17,000 new cases every year in America (Ostrom et al., 2018). The World Health Organization defined grade II and III gliomas as diffuse lower grade gliomas (LGGs), occupying approximately 30% of intracranial tumors (Brat et al., 2015). LGGs predict a better prognosis than glioblastoma; however, despite their highly heterogeneous natural processes, most LGGs progress to aggressive high-grade secondary gliomas that ultimately cause death (Zhang et al., 2020). Despite great advances in treatment for LGGs being achieved, including neurosurgery, radiotherapy, and chemotherapy, the treatment of LGGs remains a challenge (Youland et al., 2017). Recently, our ability to diagnose and prognose LGGs has been enhanced by the discovery of various biomarkers. The classification of CNS tumors was revised by the World Health Organization in 2016, based on

morphological and molecular features, which highlighted the importance of molecular testing and signature construction for the diagnosis and prognosis of LGGs (Louis et al., 2016).

Compared with healthy cells, cancer cells have very different metabolic patterns, adjusting metabolism to sustain the biosynthetic demands of tumor proliferation and have a higher resistance to cell death pathways (Abbaszadeh et al., 2020). Changes in the tumor glucose metabolic environment are often associated with oxygen supply and demand, dysfunction of key enzymes, or mitochondrial dysfunction, frequently resulting in advanced tumor aggressiveness, poor prognosis, and limited efficacy of available therapeutic methods (Ancey et al., 2018). Although some studies have shown that tumor glucose metabolism disorders are closely associated with cancer progression and low survival in gliomas, there is no glycolysis-based approach to distinguish high-risk patients (Gabriely et al., 2017; Strickland and Stoll, 2017).

There has been significant progress in molecular markers; however, they are still not satisfactory and there is still much room for improvement (Zhang et al., 2017). The present study aimed to determine whether glycolysis is a primary survival-related risk factor for LGGs, to identify powerful biomarkers, and construct a glycolysis-related gene signature for LGG prognosis. Furthermore, the prognostic efficiency of the model was tested in two additional validation sets. Finally, we established a comprehensive model based on clinicopathological and genetic characteristics to enhance the signature's accuracy and predictive power, which might be applied to guide the clinical care and patient consultation.

## MATERIALS AND METHODS

### Preparation and Processing of Data

We downloaded the transcriptome profiles (HTSeq Fragments Per kilobase of transcript per Million mapped reads [FPKM]) and their associated clinical profiles from TCGA Biobank (Colaprico et al., 2016), which was used as The Cancer Genome Atlas (TCGA) training cohort ( $n = 501$ ) to establish the risk score signature. The *IDH1* (encoding isocitrate dehydrogenase 1) statuses were calculated using maftools (Mayakonda et al., 2018). The first validation cohort comprised RNA sequencing (RNA-seq) data from the Chinese Glioma Genome Atlas (CGGA) database ( $n = 552$ ). Gene expression was also normalized and calculated using the FPKM method. Rembrandt microarray dataset was downloaded from CGGA database and used as the second validation cohort ( $n = 123$ ) (Gusev et al., 2018). The 13,609 common genes in 3 independent cohorts are used for the following analysis.

The batch effect of RNA-seq data in the CGGA database, which comprised two independent cohorts, was removed using SVA packages (Brägelmann and Lorenzo Bermejo, 2019). Patients with grade II and III glioma with complete sex, age, *IDH1* status, and survival information were used in the following study. The patients with grade II and III glioma in the Rembrandt dataset with complete survival information were used for the follow-up study. For all the included RNA-seq and microarray data, normalization and log<sub>2</sub> transformation were performed. The RNA-seq data were adopted in FPKM methods, the microarray data were normalization using RMA methods by oligo package (Carvalho and Irizarry, 2010).

### Screening Candidate Genes and Building the Signature

The RNA-seq training data and hallmark gene sets downloaded from the Molecular Signatures database (MSigDB) (Liberzon et al., 2015) were used by a single-sample gene set enrichment analysis (ssGSEA) algorithm (Hänzelmann et al., 2013) as the basis to calculate the quantified score of each cancer hallmark. Univariate Cox analysis was used to calculate the significance of different cancer hallmarks in patients with LGG. With a network type of unsigned and a soft threshold of  $\beta = 5$  (scale free  $R^2 = 0.8997766$ ), the expression values of protein coding genes in the LGG samples were subjected to weighted gene co-expression network analysis (WGCNA) to construct a scale-free co-expression network, which was used to screen those genes that were most correlated with glycolysis, based on their ssGSEA scores (Langfelder and Horvath, 2008). Subsequently, 19 modules were identified by setting the merged threshold function at 0.25, the green module was identified the genes of significantly related module. The correlation between the gene expression profiles and module eigengenes was measured using module membership (MM), and the correlation between the glycolysis ssGSEA score and individual genes represented the gene significance (GS). We screened 407 extracted candidates from the "glycolysis module" using a cutoff  $p$  value of  $<0.0001$  and the univariate Cox regression calculated  $p$  value of  $<0.01$ . Then, a least absolute shrinkage and selection operator (Lasso) Cox regression algorithm were employed to identify the most significant prognostic markers (Friedman et al., 2010). We then constructed a glycolysis-related risk score (GRS) that included normalized gene expression values weighted by their LASSO Cox coefficients as follow:

$$GRS = \sum_{i=1}^n coefficient(mRNA_i) * expr(mRNA_i)$$

### Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Analysis Based on DEGs Between High GRS Score and Low GRS Score

Based on median value of GRS score, we divided the patients in training cohort into high-risk group and low-risk group, we obtain 1827 differentially expressed genes ( $p < 0.05$ ,  $|\log_2FC| > 1$ ) using the edgeR (Robinson et al., 2010) package of R between two group. Based on the differential genes that have been obtained, Gene ontology and Kyoto encyclopedia of genes and genomes enrichment analyses were launched to probe the potential biological functions and signaling pathways by clusterProfiler (Yu et al., 2012), enrichplot packages (Yu, 2021). Standards of significantly enriched terms were set as follow:  $q$ -value  $< 0.05$ .

### Gene Set Enrichment Analysis

We utilized the msigdb. v7.0. entrez.gmt gene sets from the MSigDB database and clusterProfiler packages to explore the therapeutic resistance and possible cellular pathways. We set following standards for significantly enriched terms: I. NOM  $p$ -value  $< 0.05$ ; II. FDR  $q$ -value  $< 0.05$ .

## Drug Sensitivity of the Members in GRS Score

The drug sensitivity of the members in GRS score were calculated by online webtool GSCALite (<http://bioinfo.life.hust.edu.cn/web/GSCALite/>). The results from CTRP database were adopted and displayed in our study.

## Development and Evaluation of Clinical Predictive Models

A decision tree was constructed using recursive partitioning analysis (RPA) in the “rpart” package for risk stratification (Terry and Beth, 2019). For the quantitative prediction of the prognosis of patients with LGG, we established a prognostic nomogram, comprising the glycolysis-based risk model and other clinical parameters, which was used to predict the probability of overall survival (OS) for 1, 3, and 5 years. The predictive ability of the model was demonstrated by plotting a calibration curve, in which a curve close to 45° indicates a good predictive ability.

## Bioinformatic and Statistical Analyses

Data processing and graph construction were carried out using R software (version 4.0.3, <http://www.r-project.org>). Adobe Illustrator 2020 was used to fine tune the graphics. Survival was evaluated using the Kaplan-Meier method. The differences between the high-risk group and low-risk group in the targeted cohorts were compared using a Log-rank test. To evaluate the risk signature’s predictive performance, receiver operating characteristic curve (ROC) and area under the curve (AUC) at 1, 3, and 5 years were calculated in all cohorts using the “timeROC” packages (Blanche et al., 2013). In the absence of a full-scale gene signature expression profile, each cohort was divided into different clusters according to the optimum k value using K-means consensus clustering in the “ConsensusClusterPlus” packages. The optimal cluster numbers were determined by constructing CDF curves respectively (Wilkerson and Hayes, 2010). In the pooled cohort, to assess the prognostic value, a meta-analysis ( $I^2 = 97\%$ , random-effect model) was used.

## RESULTS

### Schematic Diagram of Research Design

The information for the patients in each cohort who met the inclusion criteria is listed in **Supplementary Table S1**. First, among the various hallmarks of cancer, we identified glycolysis as the primary OS-related risk factor in patients with LGG (**Figure 1A**). Then, WGCNA, univariate Cox regression analysis, and the LASSO algorithm were used to screen candidate genes and construct a robust signature, which could be used to predict patient survival reliably (**Figure 1B**). Next, the glycolysis signature’s prognostic efficiency was assessed in the training set and two independent validation datasets. In addition, to confirm the signature’s prognostic accuracy, a meta-analysis was performed, and the usability of the signature in clinical practice was assessed by choosing the response to therapy as an important evaluation indicator (**Figure 1C**). Finally, we built a decision tree for prognostic precision improvement, and constructed a nomogram based on GRS and other

clinicopathological indicators to evaluate the level of risk and the individual patients’ survival probability (**Figure 1D**).

## Glycolysis Is the Primary OS-Related Risk Factor in LGGs

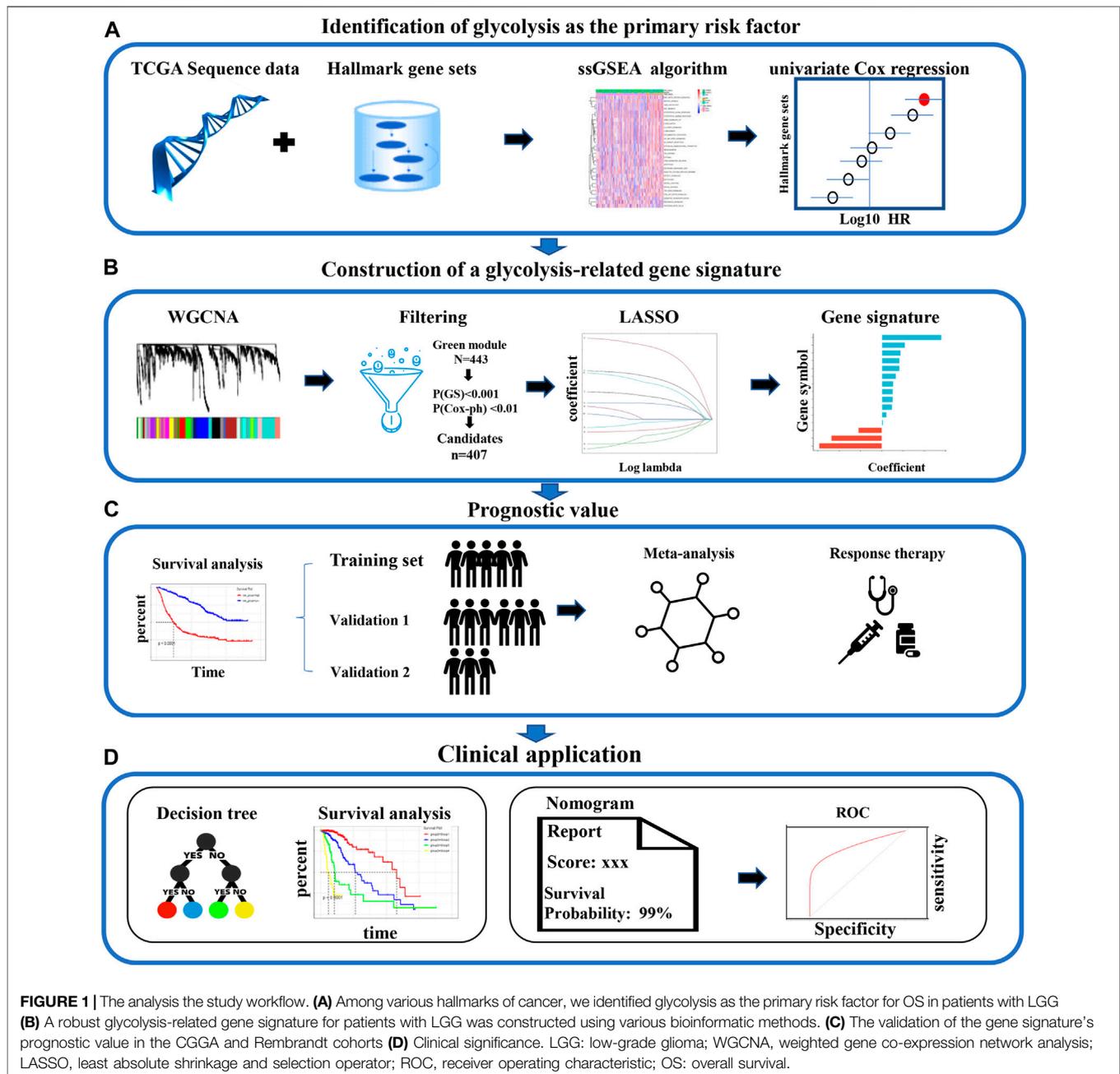
ssGSEA is an analytical algorithm that uses RNA expression data to score different cancer hallmarks in each sample. The score of 50 hallmark gene sets were calculated and 30 gene sets with a  $p$ -value < 0.05 were considered to represent significantly enriched pathways (**Figure 2A**). Compared with other cancer gene sets, such as those representing apical junction, apoptosis, and angiogenesis, the glycolysis-related gene set had the most powerful influence (highest HR) on survival (**Figure 2B**). Based on their median ssGSEA score, patients ( $n = 501$ ) could divide into a low score group and a high score group, which exhibited poorer OS compared with those in the low score group ( $p < 0.05$ ; **Figure 2C**).

## Construction of a Prognostic Glycolysis-Related Gene Signature

The top 50 percent of differentially expressed genes (6,805 genes), selected on the basis of an analysis of variance and their glycolysis ssGSEA score in the training set, were chosen for WGCNA. The optimal soft-thresholding power was 5, which ensured that the co-expression network was scale free (**Supplementary Figure S1A**). Specifically, 20 co-expression modules were identified after merging modules with similarities above 0.75 (**Figure 3A**). To assess the stability of each module identified in the training cohort, we divided the TCGA data into training and testing cohort to conduct module preservation analysis using the module preservation test (nPermutations = 200) in the WGCNA package; modules with Z-score >10 were considered to be strongly preserved (**Supplementary Figure S1B**). Among the modules with Z-score >10, the green module showed the highest correlation ( $r = 0.57$ ,  $p = 3e-44$ ) and was thus considered to correlate the most with glycolysis (**Figure 3B**). Using a  $p$  value <0.0001 as the filtering criterion for GS, univariate Cox regression analysis was performed on hub genes extracted from the green module. Based on a threshold  $p$  value < 0.01 for the univariate Cox regression analysis, 407 candidate genes were submitted to the LASSO Cox regression algorithm, which can identify the strongest prognostic markers for prognosis. Ten-fold cross-validation was used to abrogate over-filtering, by selecting an optimal  $\lambda$  value of 0.0684 (**Figure 3C**). Fifteen genes (*ARL3*, *EMP3*, *IGFBP2*, *PTGFRN*, *ADAMTS3*, *ARL9*, *SEMA4G*, *RYR3*, *TNFRSF11B*, *SSFA2*, *ABCC3*, *EMILIN3*, *IGF2BP2*, *KLHL9*, and *RHBDF1*) retained their individual non-zero LASSO coefficients (**Figure 3D**). The LASSO coefficients of each gene in the GRS. Thus, we established the GRS formula, among the members of the signature, *KLHL9*, *ARL9* and *SEMA4G* are protective factors, the others are risk factors (**Figure 3E** and **Supplementary Table S2**).

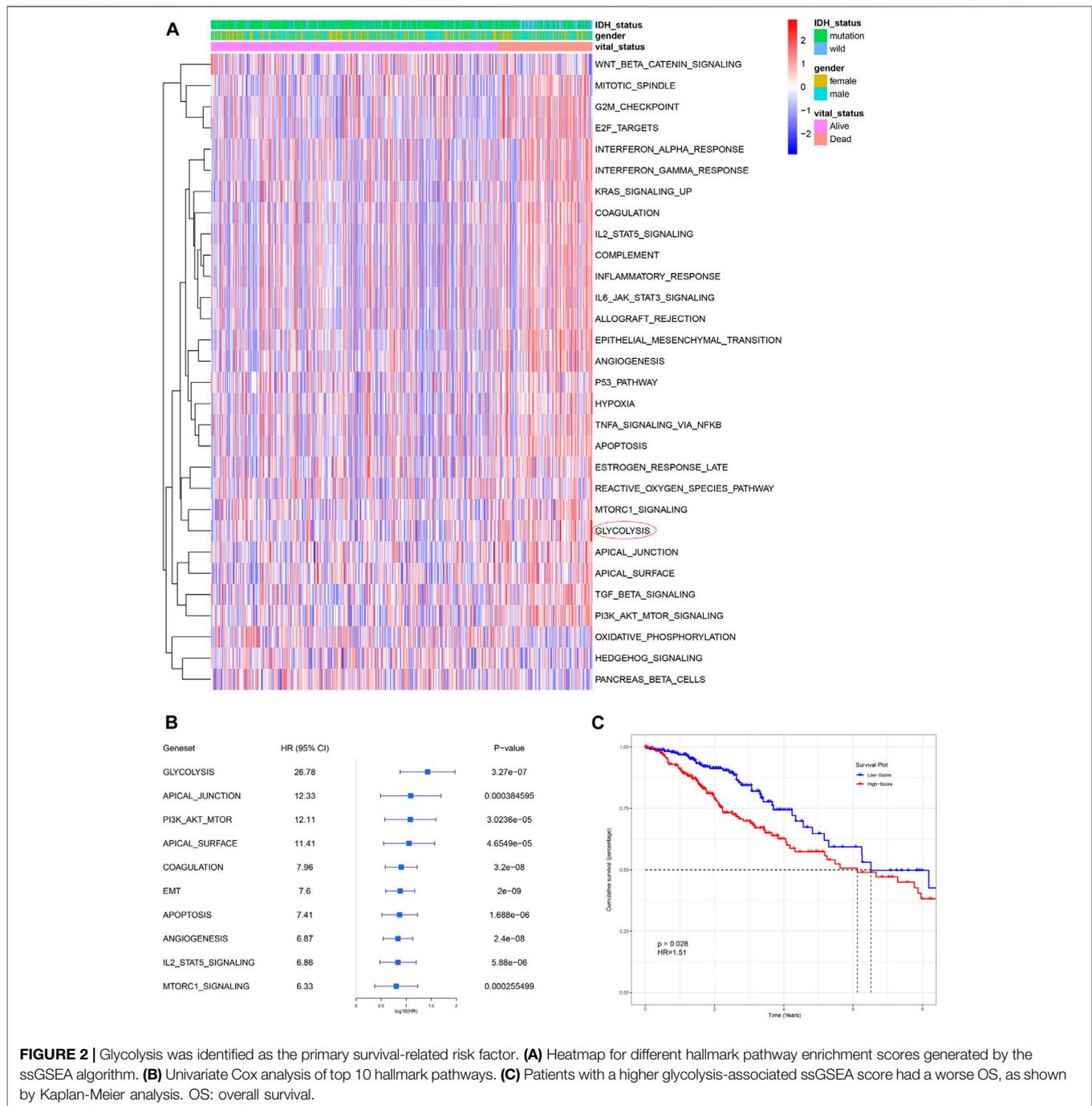
## GRS Is a Risk Factor for OS in Each Cohort

The median GRS risk score in the training set and two validation sets were used to categorize patients with LGG into low- and high-risk groups. In each cohort, patients in the high-GRS score group had worse OS than those with a low GRS score (**Figures 4A–C**,  $p <$



0.0001). Time-dependent ROC curves were used to evaluate the reliability of the GRS (Figures 4D–F). The AUC values were 0.78, 0.711, and 0.603 for 1-year, 3-years, and 5-years survival, respectively, in the TCGA training set, implying good reliability of the GRS to monitor survival. In the CGGA validation cohort, the AUC values were 0.731, 0.753, and 0.717; and in the Rembrandt cohort, the AUC values were 0.777, 0.842, and 0.749 for 1-year, 3-years and 5-years survival, respectively. Clustering was performed using a k-means algorithm which divided each cohort into different groups based on the best k value of genes in the GRS signature (Figures 4G–I), basing on the cumulative distribution function (CDF) curves

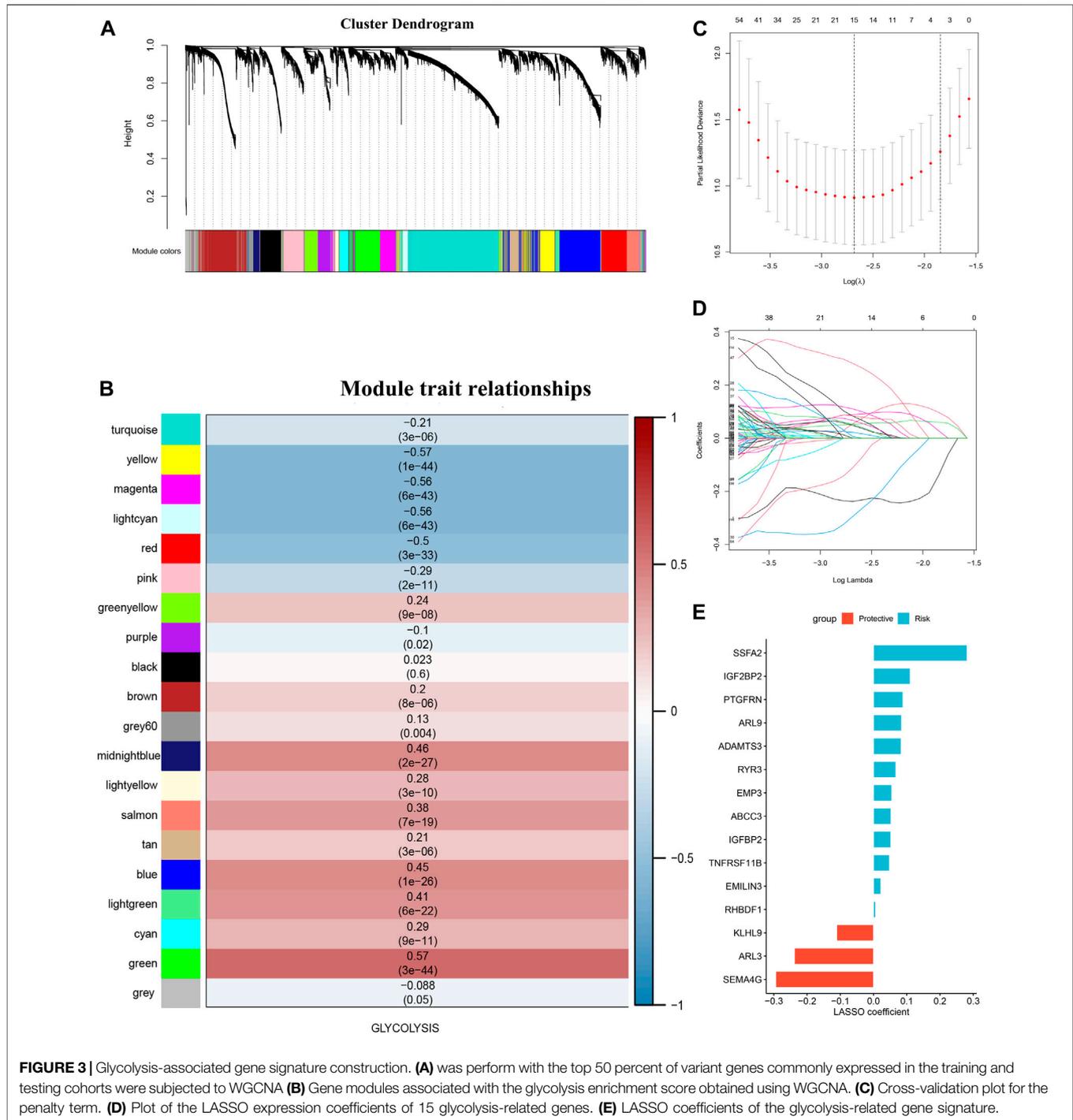
(Supplementary Figure S2A: TCGA, Supplementary Figure S2B: CGGA, Supplementary Figure S2C: Rembrandt). The results showed that the OS differed significantly among the k-means derived groups, which suggested that the classification means based on genes in the signature could be used directly for tumor subtyping (Figure 4J–L). Moreover, a meta-analysis of the prognostic value of the GRS in the pooled cohort, which integrated the TCGA training cohort and the two validation cohorts, showed that among all the 1,176 patients in three cohorts, those with a lower GRS had a better prognosis than those with a high GRS (pooled HR = 1.23, 95% confidence interval (CI): 1.09–1.39; Figure 4M).



## GRS Score Was Associated With the Clinicopathological Characteristics

To identifying the relationship between the GRS score and the clinicopathological characteristics, we explored the corresponding clinical information of LGG cases in the TCGA training cohort and the CGGA testing cohort (The clinical information in the Rembrandt cohort was

incomplete). The GRS score was associated significantly with age, survival status, and *IDH1* status. The results showed that the GRS score was high in “Dead (survival status)”, “> 40 (Age)”, “*IDH1* wild-type”, non-codel (1p19q status) unmethylated (MGMT status) group, the respective *p* values were below 0.05 (Figures 5A–F), the same trend was observed in the CGGA cohort (Figure 5G–L).

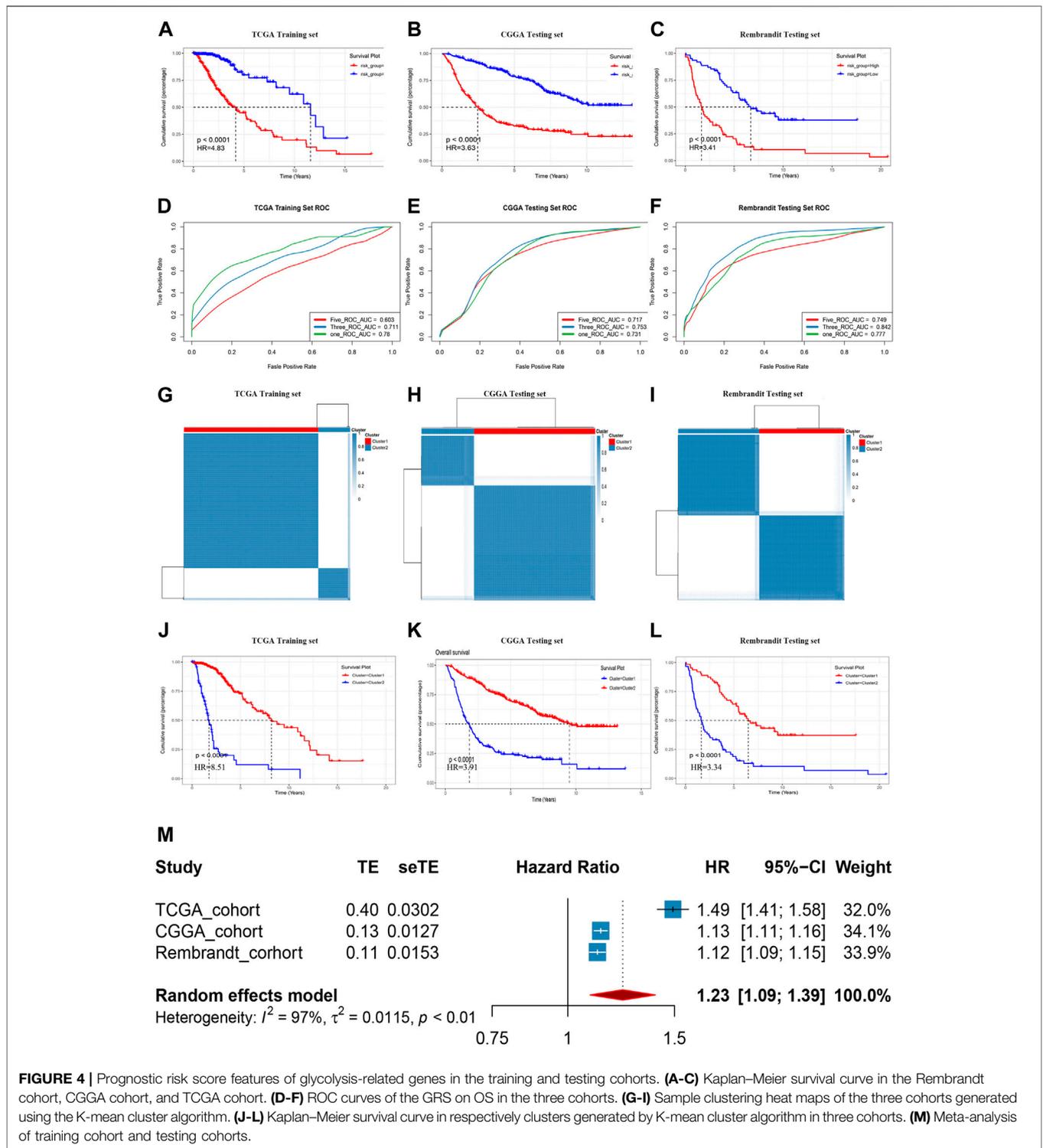


**FIGURE 3 |** Glycolysis-associated gene signature construction. **(A)** was performed with the top 50 percent of variant genes commonly expressed in the training and testing cohorts were subjected to WGCNA **(B)** Gene modules associated with the glycolysis enrichment score obtained using WGCNA. **(C)** Cross-validation plot for the penalty term. **(D)** Plot of the LASSO expression coefficients of 15 glycolysis-related genes. **(E)** LASSO coefficients of the glycolysis-related gene signature.

## Pathway Enrichment Analysis and Visualization of Differentially Expressed Gene (DEGs) Between the High and Low GRS Score Groups

To explore the association of gene expression with GRS score, gene expression in 501 patients with LGG in the training set was

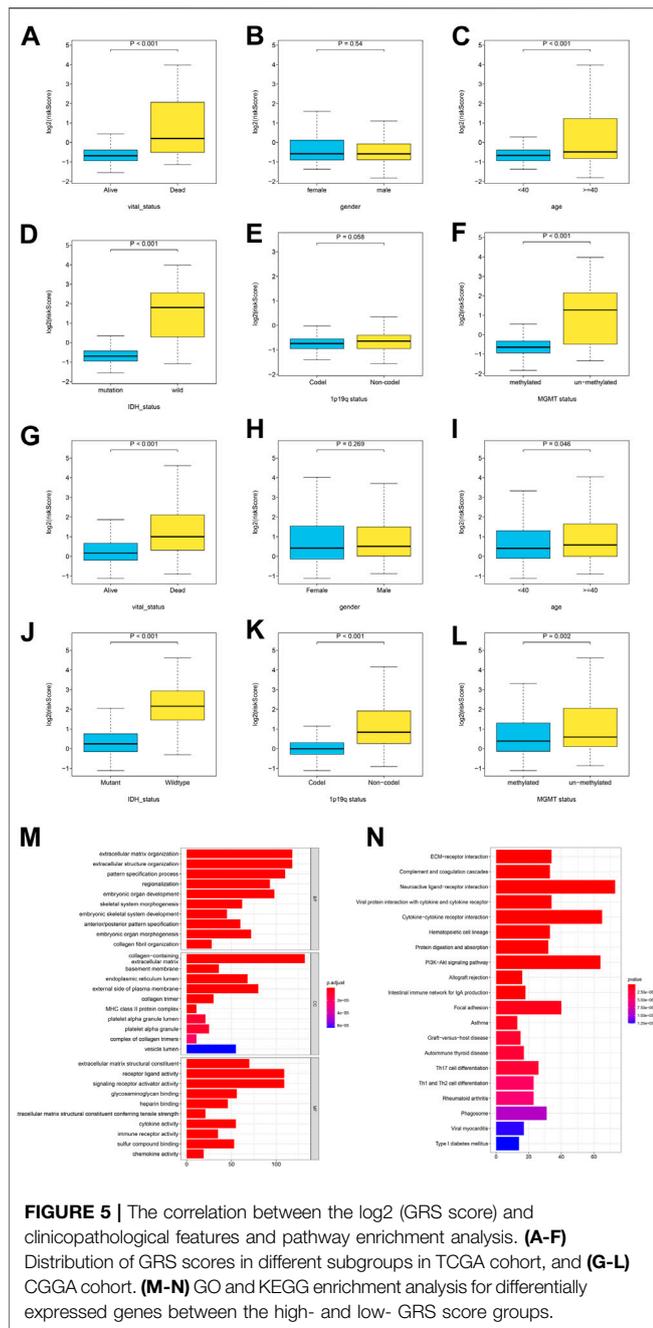
compared between the high and low GRS score groups, to identify differentially expressed genes (DEGs). We obtain 1827 differentially expressed genes ( $p < 0.05$ ,  $|\log_2FC| > 1$ ), Gene ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed on the DEGs. The GO analysis showed that a higher GRS score was associated with



**FIGURE 4 |** Prognostic risk score features of glycolysis-related genes in the training and testing cohorts. **(A-C)** Kaplan–Meier survival curve in the Rembrandt cohort, CGGA cohort, and TCGA cohort. **(D-F)** ROC curves of the GRS on OS in the three cohorts. **(G-I)** Sample clustering heat maps of the three cohorts generated using the K-mean cluster algorithm. **(J-L)** Kaplan–Meier survival curve in respectively clusters generated by K-mean cluster algorithm in three cohorts. **(M)** Meta-analysis of training cohort and testing cohorts.

extracellular matrix organization and extracellular structure organization (Figure 5M). The KEGG analysis indicated that a higher GRS score was related to ECM–receptor interaction, Complement and coagulation cascades

(Figure 5M). Taken together, these results suggested that the proteins encoded by the GRS genes have important functions in tumor microenvironment (TME) remodeling in patients with LGG.



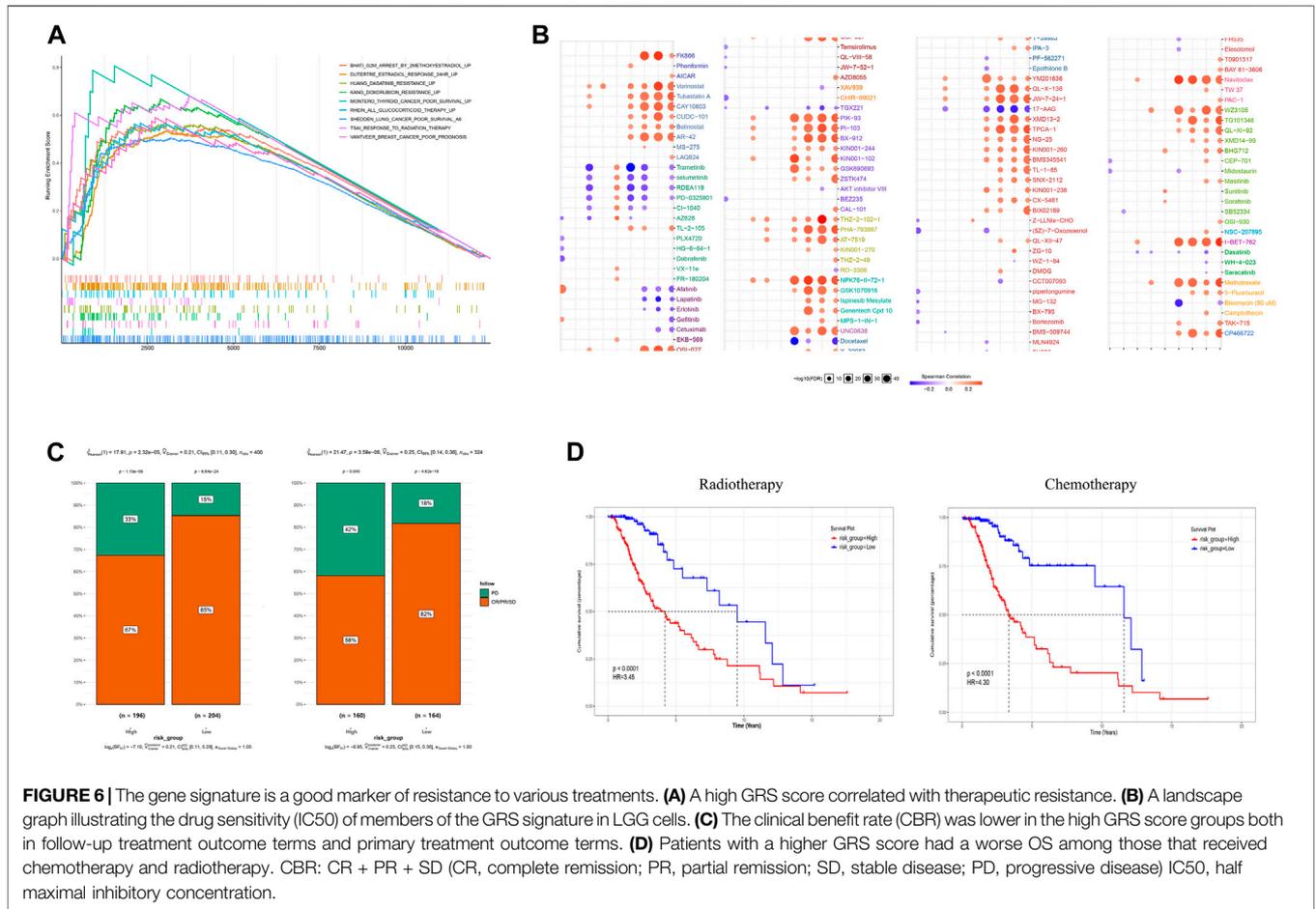
## GRS Acts as an Indicator of Therapeutic Resistance and Potential Treatments for Patients With High GRS Score Patients

Tumor glycolysis increases therapy resistance; therefore, we determined if the GRS could function as an indicator of therapeutic resistance (Hamilton et al., 2018). In addition to being associated with poor survival, a higher GRS score correlated significantly with resistance to various therapies, such as targeted therapy, radiation therapy, and chemotherapy, according to

GSEA in the training cohort (Figure 6A). The online tool GSCALite was used to draw landscape plot (bubble heatmap), which demonstrated the relationships between gene signature members and drug responses (Figure 6B). Genes in the signature correlated significantly with the half-maximal inhibitory concentration (IC<sub>50</sub>) data in LGG cells. The genes *ABCC3*, *RHBDF1*, and *PTGFRN* conferred drug resistance, which were consistent with the results shown in Figure 3E. Next, clinical outcomes and treatment information from the training cohort were used to verify these predictions. Following surgery, the clinical benefit rate (CR, complete remission; PR, partial remission; SD, stable disease) was remarkably lower ( $p < 0.0001$ ) in the higher GRS group, both in terms of follow-up treatment outcome and primary treatment outcome (Figure 6C). Furthermore, among patients who had been treated with adjuvant therapies (chemotherapy and radiotherapy), patients with a higher GRS score had worse ( $p < 0.0001$ ) OS than those with a lower GRS score (Figure 6D).

## Integrating the Glycolysis Signature With Clinicopathological Features to Improve Risk Stratification and Survival Prediction

A decision tree was constructed to increase the stratification of risk using the three parameters available age ( $\geq 40$  or  $< 40$ ), IDH status (wild and mutant), and the GRS (high and low), the results of which showed that all factors (nsplit = 3, xerror is minimum) remained in the decision tree and four different risk subgroups were identified (Figure 7A). The GRS score played an important role in the model. Kaplan–Meier curves showed that OS differed significantly among the four risk subgroups. Patients in the high GRS score, wild-type *IDH1* status, advanced age (age  $> 40$ ) subgroups had the highest risk (Figure 7B). The patients in the training cohort with age, sex, *IDH1* status, and GRS scores were used for further study. Univariate Cox regression analysis showed that age, *IDH1* status, and GRS score were associated significantly with the prognosis of LGG (Figure 7C,  $p < 0.05$ ). Multivariate Cox regression analysis identified GRS as an independent prognostic factor combined with other clinicopathological factors (Figure 7D,  $p < 0.2$ ). These three factors (The GRS score was scaled) were constructed into a nomogram to quantify the risk assessment and survival probability for individual patients with LGG (Figure 7E). The forecast curve of the calibration analysis (black line) of the nomogram for 3-years and 5-years survival probability closely resembled that of the ideal performance (the grey line in Figures 7G,H), indicating that the nomogram was highly accurate. Finally, the nomogram's reliability was evaluated using time-dependent ROC curves (Figure 7F). The AUC values were 0.875, 0.893, 0.799, and 0.762 for 1-, 3-, 5-, and 10-years survival, respectively, indicating a good potential in clinical practice for monitoring survival. We conduct same analysis in the CGGA cohort, the result was exhibited in Supplementary Figure S3, which also suggest high reliability of our model.

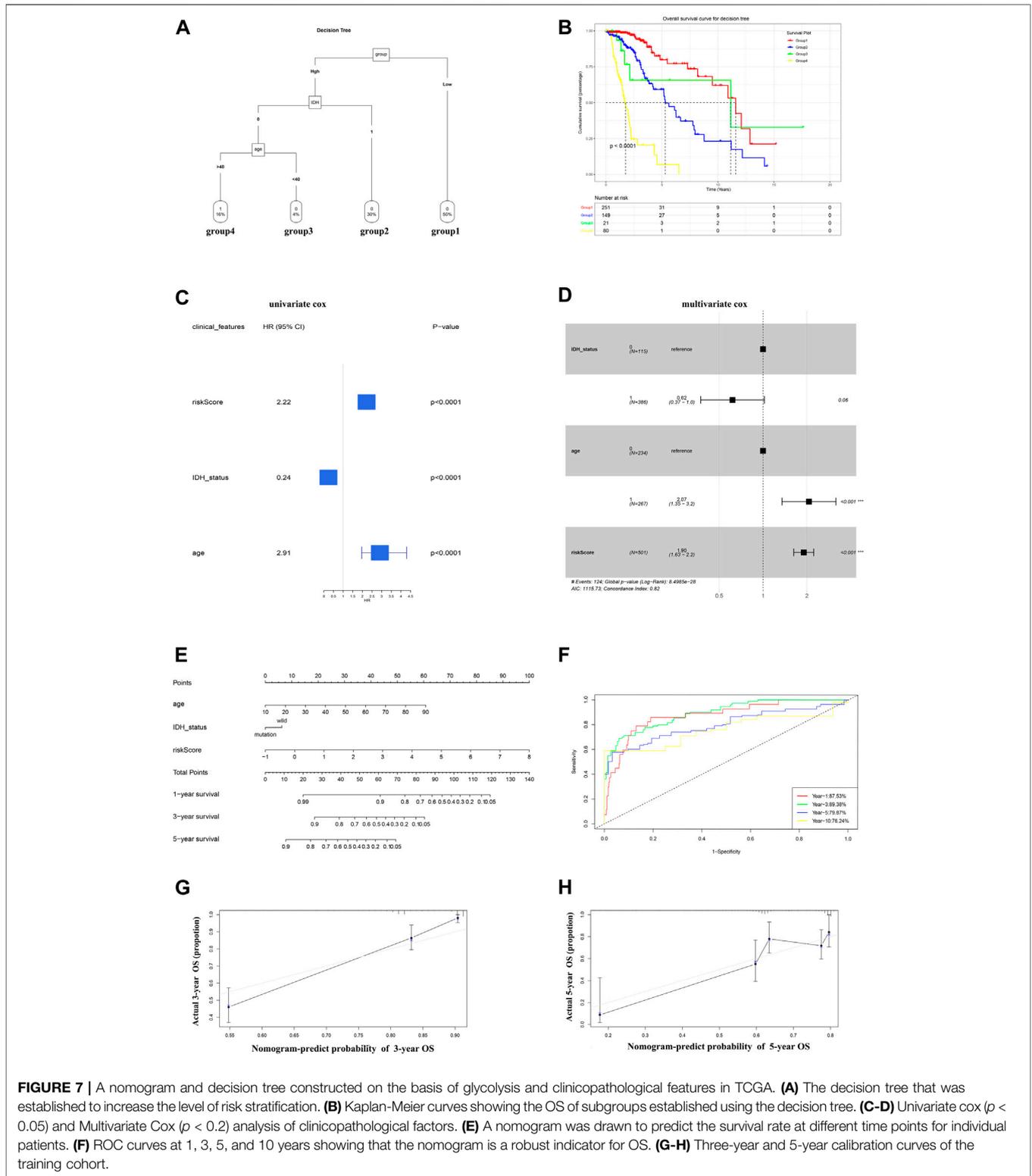


## DISCUSSION

Compared with normal cells, tumor cells have many unique hallmarks, which are of great significance in the precise treatment of tumors. Metabolic change is one of the most intriguing areas of tumor research (Pavlova and Thompson, 2016). To meet their high metabolism, tumor cells must struggle to get enough nutrients (Ou and Lv, 2020). More than 90 years ago, Warburg et al. found that tumor cells maintain active glycolysis levels even in the absence of oxygen. This landmark discovery is known as the “Warburg Effect” (Liberti and Locasale, 2016). Since then, the study of glucose metabolism in tumors has been in full swing. Positron emission tomography (PET) imaging based on cancer glucose metabolism has been widely used in clinical and plays an important role in the diagnosis and monitoring of tumors (Walker et al., 2020). Therapies targeting glucose metabolism, such as glucose transporter (GLUT) inhibitors, also hold great promise in the treatment of tumors (Reckzeh and Waldmann, 2020). The role of glucose metabolism in cancer is becoming increasingly attractive.

As research progressed, the role of glucose metabolism in tumors was found to be related to more than just energy supply.

Some intermediates of glucose metabolism, such as lactic acid, may be involved in improving the tumor microenvironment and mediating immune reprogramming (Vaupel et al., 2019). Clinicians can also be able to choose and tailor treatments based on well-established disease models. Zhang et al. constructed immune infiltrating cells-derived risk signature, which can well describe immune characteristics and predict prognosis in glioma patients (Zhang et al., 2021). Therefore, the current research on glucose metabolism is far from enough. In addition, several clinical studies have found that high glucose metabolism is associated with poor prognosis in some tumors. Zhang et al. constructed a prognostic signature of glycolysis-related genes in lung adenocarcinoma (Zhang et al., 2019), in which patients in the high-risk group had a worse prognosis. Yu et al. also constructed a prognostic model of glycolysis-related genes in gastric cancer and found that glycolysis was related to prognosis and immune infiltration in gastric cancer patients (Yu et al., 2020). However, for the low-grade glioma, a tumor with a very poor prognosis, no glycolysis-related prognostic model has been established. Therefore, it is necessary to evaluate the prognostic significance and therapeutic guidance value of glycolysis in low-grade glioma.



**FIGURE 7** | A nomogram and decision tree constructed on the basis of glycolysis and clinicopathological features in TCGA. **(A)** The decision tree that was established to increase the level of risk stratification. **(B)** Kaplan-Meier curves showing the OS of subgroups established using the decision tree. **(C-D)** Univariate cox ( $p < 0.05$ ) and Multivariate Cox ( $p < 0.2$ ) analysis of clinicopathological factors. **(E)** A nomogram was drawn to predict the survival rate at different time points for individual patients. **(F)** ROC curves at 1, 3, 5, and 10 years showing that the nomogram is a robust indicator for OS. **(G-H)** Three-year and 5-year calibration curves of the training cohort.

In this study, using the ssGSEA algorithm and univariate regression model, glycolysis was identified as the primary OS-related risk factor in patients with LGG. Combining algorithm

WGCNA, univariate Cox regression and LASSO Cox regression model, we obtained robust prognostic candidates and constructed a glycolysis-related gene signature (GRS). Next, the gene

signature's prognostic value was confirmed by multiple analyses in two other cohorts. The clinical characteristics of patients with LGG were analyzed to determine the correlation between the GRS score and clinicopathological features, which showed that the GRS score was elevated significantly in patients with older age (>40), wild-type *IDH1* status, and those who had died. Thereafter, GO, KEGG enrichment analysis based on DEGs between the high and low GRS score groups were carried out to explore what aspects of the high GRS score affected the TME of patients with LGG. In addition, patients with LGG with a high GRS score experienced poorer survival compared with those in the low GRS score group in the adjuvant therapy groups, which might have been caused by therapeutic resistance induced by members of the signature, which suggested that the GRS could also be used as a reliable tool to predict therapeutic resistance in patients with LGG. Finally, a decision tree was constructed to improve the stratification of risk, which was integrated with the clinicopathological characteristics. In the decision tree, the GRS score functioned as the major decisive factor. Meanwhile, the GRS was confirmed as an independent prognostic factor after adjusting for other clinicopathological features using multivariate Cox regression analysis. These results indicated that the GRS is a reliable risk factor for OS in patients with LGG. To increase the predictive ability of OS for LGG in a quantitative manner, a prognostic nomogram containing scale-GRS scores and other clinical features was established to predict the probability of 1-, 3-, 5-years OS. Calibration analysis demonstrated the accurate predictive ability of the nomogram, which was in accord with actual survival. Furthermore, ROC analysis demonstrated that the nomogram model exhibited high accuracy to predict survival in the timeline of follow up.

Some candidates included in the signature, have been reported to be involved in the glucose metabolic process of low-grade glioma cancer. ARL3 cycling between an in active GDP-bound and an active GTP-bound form, is involved in energy metabolic process, which may be associated with glucose metabolic process (Zhou et al., 2006; Veltel et al., 2008). Rahman et al. demonstrated that IGF1R could induce the increase of glioma invasion and malignancy by activating PTEN and AKT pathways, which is enhanced by HIF- $\alpha$  associated with promoting glycolysis related activities (Rahman and Thomas, 2011; Das et al., 2013). Li et al. reported that the express of RHBDF1 inhibited the RACK1 induced HIF- $\alpha$  degradation in breast cancer, which aggravated the hypoxic environment of the tumor environment and promoted the activation of the glycolysis pathway (Zhou et al., 2014). Based on comprehensive bioinformatic analysis, this study further screened some potentials which have not been reported in glucose metabolic process, which may provide new insight for the further research in glucose metabolic process of glioma.

Meanwhile, glycolysis related signature shows high accuracy in predicting the survival rate in 3 cohorts in our study, furthermore the signature displays strong correlation with clinicopathological features. All the results indicate GRS can serve as a reliable indicator in clinical application. To test the correctness of the model and improve the prediction ability of the model, more independent cohorts should be involved in our study. Go and KEGG analysis reveal fact glycolysis is a tumor

hallmark involving gene networks (all the members of GRS) rather than some individual "glycolysis genes", which is involved in tumor microenvironment remodeling. It gives us new enlightenment for the treatment of glioma which should be considered as a whole from multiple targets. In addition, patients with high GRS score exhibited poorer survival treated with traditional adjuvant therapy methods, which implicated urgent need for new means for treatment. we exhibited the drug sensitivity (IC50) of members of the GRS signature in LGG cells, which is need further biological experiments to validate the effectiveness *in vitro* and *vivo*.

At last, this study constructed a robust predictive model which can increase the abilities of risk stratification and survival prediction. However, the clinical information of cases in incorporated datasets were inconsistent. We established a predictive model based on the information available in TCGA training set, not including some important clinical features which can further improve the predictive ability of the model, such as p/19q status, WHO grade. In addition, the further verifications are needed by biological experiments and clinical studies.

## CONCLUSION

The present study has the advantages of incorporating different cohorts from the TCGA and CGGA databases to construct a robust glycolysis signature that can predict patient survival and therapeutic resistance with high accuracy in patients with LGG.

Combining the GRS signature with clinicopathological characteristics allowed us to construct a decision tree and a nomogram, which increased the abilities of risk stratification and survival prediction. The GRS might function as a reliable clinical prediction tool and might aid the future development of therapeutic targets.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: This study used online resources, which are available from the TCGA database at <https://portal.gdc.cancer.gov/> and the CGGA database at <http://www.cgga.org.cn/>.

## AUTHOR CONTRIBUTIONS

WL and CC conceived the study. WL and CL wrote the paper and drew the figures. WL XH, and QY analyzed the data. YY, YT, and BP edited and reviewed the paper. YY, CC, and XH were responsible for funding acquisition. All the authors contributed to the study and approved the submitted version of the article.

## FUNDING

This work was supported by the National Natural Science Foundation of China (grant number 81773354 and 82102971),

the Guangdong Basic and Applied Basic Research Foundation, China (grant number 2020A1515111074), and the Guangzhou Science and Technology Plan Project (grant number 202102020034).

## ACKNOWLEDGMENTS

We would like to thank the Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA) working groups for

kindly sharing of data. We also thank R Development Core Team and R community (<http://www.r-project.org>) for the time and effort they put into to create and optimize R.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.843711/full#supplementary-material>

## REFERENCES

- Abbaszadeh, Z., Çeşmeli, S., and Biray Avcı, Ç. (2020). Crucial Players in Glycolysis: Cancer Progress. *Gene* 726, 144158. doi:10.1016/j.gene.2019.144158
- Ancay, P. B., Contat, C., and Meylan, E. (2018). Glucose Transporters in Cancer - from Tumor Cells to the Tumor Microenvironment. *FEBS J.* 285, 2926–2943. doi:10.1111/febs.14577
- Blanche, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-Dependent Areas under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Statist. Med.* 32, 5381–5397. doi:10.1002/sim.5958
- Brägelmann, J., and Lorenzo Bermejo, J. (2019). A Comparative Analysis of Cell-Type Adjustment Methods for Epigenome-Wide Association Studies Based on Simulated and Real Data Sets. *Brief Bioinform* 20, 2055–2065. doi:10.1093/bib/bby068
- Brat, D. J., Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* 372, 2481–2498. doi:10.1056/NEJMoa1402121
- Carvalho, B. S., and Irizarry, R. A. (2010). A Framework for Oligonucleotide Microarray Preprocessing. *Bioinformatics* 26, 2363–2367. doi:10.1093/bioinformatics/btq431
- Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAAbiolinks: An R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44, e71. doi:10.1093/nar/gkv1507
- Das, S. K., Bhutia, S. K., Azab, B., Kegelman, T. P., Peachy, L., Santhekadur, P. K., et al. (2013). MDA-9/Syntenin and IGFBP-2 Promote Angiogenesis in Human Melanoma. *Cancer Res.* 73, 844–854. doi:10.1158/0008-5472.can-12-1681
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01
- Gabriely, G., Wheeler, M. A., Takenaka, M. C., and Quintana, F. J. (2017). Role of AHR and HIF-1 $\alpha$  in Glioblastoma Metabolism. *Trends Endocrinol. Metab.* 28, 428–436. doi:10.1016/j.tem.2017.02.009
- Gusev, Y., Bhuvaneshwar, K., Song, L., Zenklusen, J.-C., Fine, H., and Madhavan, S. (2018). The REMBRANDT Study, a Large Collection of Genomic Data from Brain Cancer Patients. *Sci. Data* 5, 180158. doi:10.1038/sdata.2018.158
- Hamilton, K. E., Rekmán, J. F., Gunnink, L. K., Busscher, B. M., Scott, J. L., Tidball, A. M., et al. (2018). Quercetin Inhibits Glucose Transport by Binding to an Exofacial Site on GLUT1. *Biochimie* 151, 107–114. doi:10.1016/j.biochi.2018.05.012
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *Bmc Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7
- Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *Bmc Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559
- Liberti, M. V., and Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* 41, 211–218. doi:10.1016/j.tibs.2015.12.001
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cel Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004
- 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
- Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118
- Ostrom, Q. T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J. S. (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015. *Neuro Oncol.* 20, iv1–iv86. doi:10.1093/neuonc/now131
- Ou, X., and Lv, W. (2020). Metabolic Changes and Interaction of Tumor Cell, Myeloid-Derived Suppressor Cell and T Cell in Hypoxic Microenvironment. *Future Oncol.* 16, 383–393. doi:10.2217/fon-2019-0692
- Pavlova, N. N., and Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cel Metab.* 23, 27–47. doi:10.1016/j.cmet.2015.12.006
- Rahman, M. S., and Thomas, P. (2011). Characterization of Three IGFBP mRNAs in Atlantic Croaker and Their Regulation during Hypoxic Stress: Potential Mechanisms of Their Upregulation by Hypoxia. *Am. J. Physiology-Endocrinology Metab.* 301, E637–E648. doi:10.1152/ajpendo.00168.2011
- Reckzeh, E. S., and Waldmann, H. (2020). Development of Glucose Transporter (GLUT) Inhibitors. *Eur. J. Org. Chem.* 2020, 2321–2329. doi:10.1002/ejoc.201901353
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Strickland, M., and Stoll, E. A. (2017). Metabolic Reprogramming in Glioma. *Front. Cel Dev. Biol.* 5, 43. doi:10.3389/fcell.2017.00043
- Terry, T., and Beth, A. (2019). Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-15. Available at: <https://CRAN.R-project.org/package=rpart>.
- Vaupel, P., Schmidberger, H., and Mayer, A. (2019). The Warburg Effect: Essential Part of Metabolic Reprogramming and central Contributor to Cancer Progression. *Int. J. Radiat. Biol.* 95, 912–919. doi:10.1080/09553002.2019.1589653
- Veltel, S., Kravchenko, A., Ismail, S., and Wittinghofer, A. (2008). Specificity of Arl2/Arl3 Signaling Is Mediated by a Ternary Arl3-Effector-GAP Complex. *Febs Lett.* 582, 2501–2507. doi:10.1016/j.febslet.2008.05.053
- Walker, S. M., Lim, I., Lindenberg, L., Mena, E., Choyke, P. L., and Turkbey, B. (2020). Positron Emission Tomography (PET) Radiotracers for Prostate Cancer Imaging. *Abdom. Radiol.* 45, 2165–2175. doi:10.1007/s00261-020-02427-4
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170
- Youland, R. S., Schomas, D. A., Brown, P. D., Parney, I. F., and Laack, N. N. I. (2017). Patterns of Care and Treatment Outcomes in Older Adults with Low Grade Glioma: a 50-year Experience. *J. Neurooncol.* 133, 339–346. doi:10.1007/s11060-017-2439-3
- Yu, G. (2021). Enrichplot: Visualization of Functional Enrichment Result. Available at: <https://yulab-smu.top/biomedical-knowledge-mining-book/>.

- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* 16, 284–287. doi:10.1089/omi.2011.0118
- Yu, S., Hu, C., Cai, L., Du, X., Lin, F., Yu, Q., et al. (2020). Seven-Gene Signature Based on Glycolysis Is Closely Related to the Prognosis and Tumor Immune Infiltration of Patients with Gastric Cancer. *Front. Oncol.* 10, 1778. doi:10.3389/fonc.2020.01778
- Zhang, C., Cheng, W., Ren, X., Wang, Z., Liu, X., Li, G., et al. (2017). Tumor Purity as an Underlying Key Factor in Glioma. *Clin. Cancer Res.* 23, 6279–6291. doi:10.1158/1078-0432.ccr-16-2598
- Zhang, H., Li, X., Li, Y., Chen, B., Zong, Z., and Shen, L. (2020). An Immune-Related Signature for Predicting the Prognosis of Lower-Grade Gliomas. *Front. Immunol.* 11, 603341. doi:10.3389/fimmu.2020.603341
- Zhang, L., Zhang, Z., and Yu, Z. (2019). Identification of a Novel Glycolysis-Related Gene Signature for Predicting Metastasis and Survival in Patients with Lung Adenocarcinoma. *J. Transl. Med.* 17, 423. doi:10.1186/s12967-019-02173-2
- Zhang, N., Zhang, H., Wang, Z., Dai, Z., Zhang, X., Cheng, Q., et al. (2021). Immune Infiltrating Cells-Derived Risk Signature Based on Large-Scale Analysis Defines Immune Landscape and Predicts Immunotherapy Responses in Glioma Tumor Microenvironment. *Front. Immunol.* 12, 691811. doi:10.3389/fimmu.2021.691811
- Zhou, C., Cunningham, L., Marcus, A. I., Li, Y., and Kahn, R. A. (2006). Arl2 and Arl3 Regulate Different Microtubule-Dependent Processes. *Mol. Biol. Cell* 17, 2476–2487. doi:10.1091/mbc.e05-10-0929
- Zhou, Z., Liu, F., Zhang, Z.-S., Shu, F., Zheng, Y., Fu, L., et al. (2014). Human Rhomboid Family-1 Suppresses Oxygen-Independent Degradation of Hypoxia-Inducible Factor-1 $\alpha$  in Breast Cancer. *Cancer Res.* 74, 2719–2730. doi:10.1158/0008-5472.can-13-1027

**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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Liu, Chen, Huang, Yi, Tian, Peng and Yuan. 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.