



Establishment and validation of a prognostic prediction model for glioma based on key genes and clinical factors

Yu Lin¹, Huining Li², Qiang Ge³, Dan Hua⁴

¹Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin, China; ²Department of Neurology, Tianjin Medical University General Hospital, Tianjin, China; ³Department of Military Preventive Medicine, Faculty of Health Services, Logistic University of People's Armed Police Force, Tianjin, China; ⁴Department of Neuropathology, Tianjin Neurological Institute, Tianjin Medical University General Hospital, Tianjin, China

Contributions: (I) Conception and design: D Hua, Y Lin; (II) Administrative support: D Hua; (III) Provision of study materials or patients: Y Lin; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dan Hua, MD, PhD. Department of Neuropathology, Tianjin Neurological Institute, Tianjin Medical University General Hospital, No. 154 Anshan Road, Heping District, Tianjin 300052, China. Email: kila1126@hotmail.com.

Background: Glioma is a common brain tumour that is associated with poor prognosis. Immunotherapy has shown significant potential in the treatment of gliomas. Herein, we proposed a new prognostic risk model based on immune- and mitochondrial energy metabolism-related differentially expressed genes (IR&MEMRDEGs) to enhance the accuracy of prognostic assessment in patients with glioma.

Methods: Data from samples from 671 glioma patients and 5 normal controls with available follow-up data and prognostic outcomes were downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. All data were downloaded on 13 November 2023. IR&MEMRDEGs were screened from the GeneCards website and published literature. Prognostic prediction models were constructed and analysed using Cox and Least Absolute Shrinkage and Selection Operator (LASSO) regression, Kaplan-Meier (KM) curve, and receiver operating characteristic (ROC) curve analyses. Single-sample gene set enrichment analysis (ssGSEA) was further performed to ascertain the percentage of immune cell infiltration in the glioma specimens.

Results: Bioinformatics analysis of the GEO and TCGA databases identified eleven MEMRDEGs with dysregulated expression in gliomas: *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *PKLR*, *AK2*, *ATP4A*, and *ALDH3B1*. Further analysis identified *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *AK2*, and *ALDH3B1* as separate predictive factors for glioma, among which *PGM2* and *AK2* exhibited superior accuracy [area under the ROC curve (AUC) >0.9], while *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *GPI*, and *ALDH3B1* demonstrated slightly lower accuracy (0.7 < AUC < 0.9), and *CD200* displayed poor accuracy (0.5 < AUC < 0.7). Among these genes, the levels of *AK2*, *ALDH3B1*, *EIF4EBP1*, *GPI*, *IDH1*, *PGM2*, and *TP53* were significantly higher in the high-risk group (HRG) compared with the low-risk group (LRG) ($P < 0.001$), indicating a negative association with patient prognosis. In contrast, *CD200* and *PRKCZ* were significantly downregulated in the HRG compared to the LRG ($P < 0.05$), indicating a potential correlation with patient outcomes. Subsequently, prognostic models were constructed based on IR&MEMRDEG and MEMRDEGs to anticipate the outcomes of glioma patients, while the predictive efficacy of the model was validated via KM and ROC curve analysis. The results revealed that *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *GPI*, *PGM2*, *ALDH3B1*, and *AK2* had superior accuracy in predicting glioma prognosis. The ssGSEA results showed that only *IDH1* was negatively linked to the amount of immune cell infiltration in the LRG, while displaying a positive connection in the HRG (r value > 0), indicating that the expression levels of *IDH1* may have a distinct influence on the tumour immune microenvironment.

Conclusions: The present study confirmed the significant predictive value of *IDH1* for glioma prognosis, which may guide immunotherapy for glioma treatment.

Keywords: Glioma; Gene Expression Omnibus (GEO); The Cancer Genome Atlas (TCGA); immune; prognostic model

Submitted Jun 21, 2024. Accepted for publication Nov 26, 2024. Published online Jan 20, 2025.

doi: 10.21037/tcr-24-1035

View this article at: <https://dx.doi.org/10.21037/tcr-24-1035>

Introduction

Glioma is the most common type of brain tumour worldwide. The World Health Organization (WHO) guidelines classify gliomas into four categories (1). Although the prognostic biomarkers and pathways associated with gliomas have been studied, there are currently no prognostic biomarkers or prognostic prediction models that can effectively improve the therapeutic outcomes of patients with gliomas (2). Currently, glioma, specifically high-grade glioma (i.e., grade 4 glioma), has an adverse prognosis, as evidenced by its median survival period of 15

months (3-5). Conventional treatments for gliomas include surgical resection, postoperative temozolomide (TMZ), and radiotherapy (6). However, these approaches have a poor prognosis (7). Despite substantial advances in cellular and molecular marker research, therapeutic strategies for glioma remain unsatisfactory.

Immunotherapy has emerged as a crucial component of cancer therapy, leading to its increased use in glioma treatment (8). Consequently, we proposed the creation of a novel prognostic risk model (PRM), and validated its efficacy, potentially assisting in the application of immunotherapy for glioma.

Mitochondria affect the efficacy of immunotherapy in tumours by modulating immune cell activity and establishing an immune microenvironment (9,10). Mitochondrial energy metabolism plays a critical role in the outcomes of immunotherapy. Currently, only a few studies have investigated the expression of genes linked to immunity and mitochondrial energy metabolism in patients with glioma, as well as their prognostic implications (11). Therefore, the present study aimed to construct and assess a PRM based on immune-related and mitochondrial energy metabolism-related differentially expressed genes (IR&MEMRDEGs), to enhance the accuracy of prognostic assessment in patients with glioma. Furthermore, the immune infiltration of these genes was analysed, serving as an independent prognostic factor and contributing to the advancement of immunotherapy strategies for glioma. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1035/rc>).

Methods

The flow chart of the experiments performed in this article is shown in *Figure 1*.

Data download

The GEOquery R package (12) was employed to download

Highlight box

Key findings

- We provided a new prognostic risk model based on immune- and mitochondrial energy metabolism-related differentially expressed genes (IR&MEMRDEGs) to enhance the accuracy of prognostic assessment in patients with glioma.

What is known and what is new?

- Immunotherapy has emerged as a crucial component of cancer therapy, leading to its increased use in glioma treatment. Mitochondria affect the efficacy of immunotherapy in tumors by modulating immune cell activity and establishing an immune microenvironment. Mitochondrial energy metabolism plays a critical role in the outcomes of immunotherapy. Currently, only a few studies have investigated the expression of genes linked to immunity and mitochondrial energy metabolism in patients with glioma, as well as their prognostic implications.
- We proposed the creation of a novel prognostic risk model (PRM), and validated its efficacy, potentially assisting in the application of immunotherapy for glioma. We constructed and assessed a PRM based on IR&MEMRDEGs, which has the potential to enhance the accuracy of prognostic assessment in patients with glioma. Furthermore, the immune infiltration of these genes was analyzed, serving as an independent prognostic factor and contributing to the advancement of immunotherapy strategies for glioma.

What is the implication, and what should change now?

- Our study confirmed the significant predictive value of IDH1 for glioma prognosis, which may guide immunotherapy for glioma treatment.

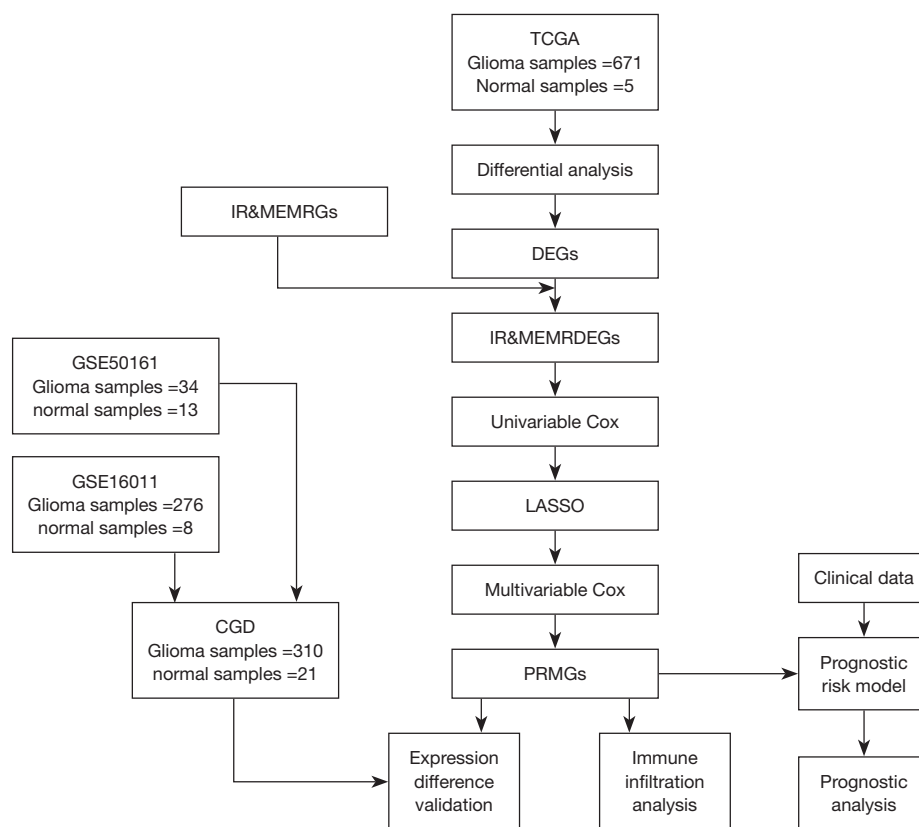


Figure 1 Flow chart of the experiments performed in this study. TCGA, The Cancer Genome Atlas; IR&MEMRGs, immune-related and mitochondrial energy metabolism-related genes; DEGs, differentially expressed genes; IR&MEMRDEGs, immune- and mitochondrial energy metabolism-related differentially expressed genes; CGD, Combined GEO Dataset; PRMGs, prognostic risk model genes; LASSO, Least Absolute Shrinkage and Selection Operator.

Table 1 The information of GEO microarray chip

Characteristics	GSE50161	GSE16011
Platform	GPL570	GPL8542
Type	Array	Array
Species	Homo sapiens	Homo sapiens
Tissue	Brain	Brain
Glioma samples in tumor group	34	276
Normal samples in normal group	13	8
Reference	(13)	(14)

GEO, Gene Expression Omnibus.

the glioma datasets (lower-grade glioma and glioblastoma multiforme, GBMLGG) GSE50161 (13) and GSE16011 (14) from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). All GEO data were

downloaded on 13th November 2023. Samples from both datasets were derived from Homo sapiens and originated from the brain tissue. The chip platforms GSE50161 and GSE16011 were GPL570 and GPL8542, respectively (Table 1). GSE50161 comprised 34 glioma samples and 13 normal samples, whereas GSE16011 comprised 276 glioma samples and eight normal samples. All samples from the glioma (tumour group, TG) and normal (NG) groups were included in this study.

The sva R package (15), was applied to conduct batch processing on the GSE50161 and GSE16011 datasets to obtain an integrated GEO dataset, hereafter referred to as the Combined GEO Dataset (CGD), comprising a total of 310 glioma and 21 NG specimens. Subsequently, CGD was standardised and annotated with probes employing limma, an R package (16). Principal component analysis (PCA) was performed on the expression matrix prior to and following batch effect elimination to assess the efficacy

of removing batch effects (17). PCA is a technique used for data dimensionality reduction that identifies feature vectors (ingredients) from high-dimensional data, which are subsequently transformed into lower-dimensional data, and visualized in two- or three-dimensional graphs. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of IR&MEMRDEGs in glioma

The glioma dataset was acquired from The Cancer Genome Atlas (TCGA) using TCGAAbiolinks, an R package referred to as the TCGA-GBMLGG dataset, which was utilised as a test set (18). All TCGA data were downloaded on 13th November 2023. Samples lacking clinical information were excluded, while the count format sequencing data of 671 glioma samples (tumour group) and five specimens in NG were obtained. Using the limma package in R, differential gene expression analysis was conducted between the TG and NG. Differentially expressed genes (DEGs) were identified using $|\log FC| > 1$ and $\text{adj. } P < 0.05$ was set as the limit of significance to identify DEGs. DEGs were categorised as elevated when $\log FC$ was > 1 with an $\text{adj. } P < 0.05$, or decreased when $\log FC$ was < 1 with an $\text{adj. } P < 0.05$. Subsequently, the GeneCards database (<https://www.genecards.org/>) (19) and published literature were searched, identifying 76 genes associated with immune and mitochondrial energy metabolism. To identify IR&MEMRDEGs in gliomas, 76 immune-related and mitochondrial energy metabolism-related genes (IR&MEMRGs) were intersected with DEGs from TCGA-GBMLGG dataset to create a Venn diagram.

Construction of a PRM for Glioma

Univariate Cox regression (UCR) and multivariate Cox regression (MCR) analyses were performed on the TCGA-GBMLGG dataset using the survival package in R to construct a PRM for gliomas (20). These analyses were based on clinical information to assess the impact of IR&MEMRDEGs on prognosis and ascertain their status as independent prognostic factors.

Initially, variables with a P value < 0.10 were discovered by UCR analysis and subsequently incorporated in the Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, which was conducted employing the glmnet package in R software, specifying family="cox" as a parameter (21). The analysis was performed on the

identified IR&MEMRDEGs from the single-factor Cox regression analysis, with the number of cycles set to 10. LASSO regression is a form of linear regression analysis that addresses overfitting and improves a model's capacity to be generalised by including a penalty term ($\lambda \times$ absolute value of the slope). The outcomes of the LASSO regression study were presented by employing PRM and variable trajectory diagrams. The LASSO risk score (RS) was computed as follows:

$$\text{Risk Score} = \sum_i \text{Coefficient}(\text{gene}_i) * \text{mRNA Expression}(\text{gene}_i) \quad [1]$$

Subsequently, MCR analysis was performed on the IR&MEMRDEG retained from the LASSO regression analysis. PRM genes (PRMGs) were identified, and the outcomes of the MCR analysis were presented using a forest plot. Finally, the high- (HRG) and low-risk group (LRG) classifications of glioma samples from the TCGA-GBMLGG dataset were ascertained using the median expression value of the LASSO RS derived from the tumour PRM.

Prognostic analysis and validation of the glioma PRM

To assess the difference in overall survival (OS) between the HRG and LRG of glioma samples within the TCGA-GBMLGG dataset, the survival package in R software was applied to conduct Kaplan-Meier (KM) curve analysis (22). The R software timeROC was employed to create a time-dependent receiver operating characteristic (ROC) curve by deploying the LASSO RS and OS. This analysis was conducted to predict the 1-, 2-, and 3-year survival rates of glioma samples within the TCGA-GBMLGG dataset, by calculating the area under the ROC curve (AUC). AUC values typically range between 0.5 and 1, with higher values closer to 1 indicating better diagnostic performance. AUC values greater than 0.5 indicate that the molecule's expression is associated with a propensity to facilitate the event's occurrence. AUC values between 0.5 and 0.7, 0.7 and 0.9, and above 0.9 indicate low, moderate, and high degrees of accuracy, respectively.

This study investigated the association and predictive capacity of the LASSO RS in relation to clinical outcomes, by employing UCR and MCR analyses, taking into account the LASSO RS, sex, age, and clinical staging variable (grade). A forest plot was subsequently applied to visualise the outcomes of both the UCR and MCR studies, which demonstrated the influence of LASSO RS and clinical factors. The nomogram graphically represents the functional connections between independent variables

by employing discrete line segments inside a rectangular coordinate system (23). A nomogram was created using the outcomes of the MCR analysis, conducted using the rms package in R software. The nomogram illustrates the connection between RS and several clinical factors in relation to survival rates at one, two, and three years.

Decision curve analysis (DCA) is a straightforward tool used to assess the performance of clinical predictive models, diagnostic trials, and molecular markers. In the present study, the ggDCA package of R software was applied to generate DCA plots derived from the nomogram model, and to evaluate and predict the 1-, 2-, and 3-year survival rates.

Validation of the differential expression and ROC curve analysis for IR&MEMRDEGs

The Mann-Whitney *U* test was applied to generate comparative graphs based on the TCGA-GBMLGG dataset and CGD to estimate the differential expression of PRMGs between TG and NG. Further, PRMGs expression levels in the high- and low-risk glioma specimen groups from both datasets were examined using the Mann-Whitney *U* test. Subsequently, the pROC package in the R software was applied to generate ROC curves for PRMGs in the TG and NG, while AUC values were computed to assess the diagnostic implications of PRMG expression levels in predicting glioma occurrence. Additionally, the pROC package was applied to generate ROC curves for PRMGs from the HRG and LRG of the glioma samples. The AUC of the ROC curve generally varies between 0.5 and 1, with greater values denoting superior prognostic performance. AUC values between 0.5 and 0.7, 0.7 and 0.9, and above 0.9 indicate low, moderate, and high degrees of accuracy, respectively.

Immune infiltration analysis

The single-sample gene set enrichment analysis (ssGSEA) approach was applied to ascertain the proportionate presence of immune cell infiltration in glioma specimens obtained from TCGA-GBMLGG database. Initially, various immune cell types were identified, including activated CD8⁺ T cells, activated dendritic cells, gamma delta T cells, natural killer cells, regulatory T cells, and different subtypes of human immune cells. Subsequently, enrichment scores computed by ssGSEA were used to indicate the proportional frequency of immune cell infiltration in each specimen, generating an immune cell

infiltration matrix for glioma samples within the TCGA-GBMLGG database. Subsequently, the ggplot2 package in R was used to estimate and contrast immune cell levels in glioma specimens from TCGA-GBMLGG dataset, specifically between the HRG and LRG. Immune cells displaying significant differences between the two groups were isolated for further examination. The correlation between immune cells was subsequently computed using the Spearman algorithm, and pheatmap, an R package, was used to generate a heatmap representing the correlation between immune cells, which represented the outcomes of the correlation study. Thereafter, correlation analysis between PRMGs and immune cells was performed using the Spearman algorithm, and ggplot2, an R package, was used to generate a correlation bubble chart for visual representation of the analysis results.

Statistical analysis

R software (version 4.3.1) was used for data processing and analyses. To determine the significance of variations in normally distributed continuous variables between the two groups, the independent Student's *t*-test was applied, unless stated otherwise. Disparities in variables that did not follow a normal distribution were assessed using the Mann-Whitney *U* test (Wilcoxon rank-sum test). The Kruskal-Wallis test was further applied to compare three or more groups. Spearman's correlation analysis was further used to determine the correlation coefficients between the various compounds. Unless explicitly stated, all *P* values in the statistical analysis were two-tailed, and any *P* value >0.05 was deemed significant.

Results

Integration of glioma datasets

To integrate the glioma datasets, SVA, an R package, was first used to eliminate batch effects in the GSE50161 and GSE16011 datasets, yielding a consolidated dataset, termed the CGD. The differential expression between the pre- and post-batch effect-removal results is illustrated in distribution box plots (Figure 2A,2B). Additionally, a PCA plot (Figure 2C,2D) was used to assess the distribution of low-dimensional features in the dataset prior to and following batch effect correction. The distribution box and PCA plots showed that the batch effect in the glioma specimens was effectively eliminated.

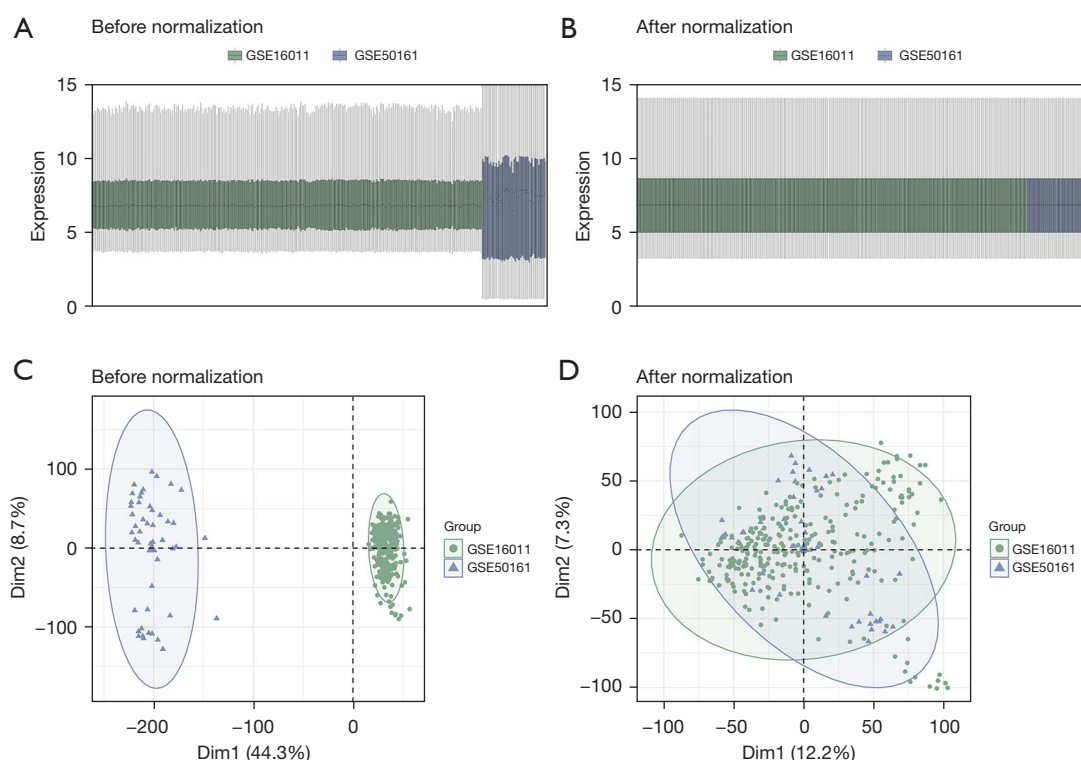


Figure 2 Removal of batch effects in datasets GSE50161 and GSE16011. (A) Box plot of the distribution of CGD prior to batch processing. (B) Boxplots depicting the distribution of the CGD post-batch processing. (C) 2D PCA plot depicting the dataset prior to batch processing. (D) 2D PCA diagram depicting CGD following batch processing. The GSE16011 and GSE50161 datasets are represented in green and blue, respectively. CGD, Combined GEO Dataset; PCA, principal component analysis; GEO, Gene Expression Omnibus.



Figure 3 Venn diagram depicting the DEGs and IR&MEMRGs in the TCGA-GBMLGG dataset. DEGs, differentially expressed genes; IR&MEMRGs, immune-related and mitochondrial energy metabolism-related genes.

Glioma-related IR&MEMRDEGs

The TCGA-GBMLGG dataset was partitioned into the TG and NG to determine the variations in gene expression

levels. Differential analysis was conducted on the TCGA-GBMLGG dataset using limma, an R package, to identify DEGs between the TG and NG samples. The investigation identified 3,623 DEGs in the TCGA-GBMLGG dataset, which were chosen identified based on the following parameters: $|\log FC| > 1$ and $\text{adj. } P < 0.05$. Moreover, 76 IR&MEMRGs were identified by searching the GeneCards database and published literature. The IR&MEMRGs and DEGs were intersected and visually represented by a Venn diagram (Figure 3). The results identified 11 IR&MEMRDEGs, namely *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *PKLR*, *AK2*, *ATP4A*, and *ALDH3B1*. Among these, the expression of *EIF4EBP1*, *TP53*, *IDH1*, *PGM2*, *AK2*, and *ALDH3B1* was upregulated in the TG compared to that in the NG, whereas *PRKCZ*, *CD200*, *GPI*, *PKLR*, and *ATP4A* were downregulated in the TG.

Development of a PRM for glioma

To establish a PRM for glioma, UCR analysis was subsequently

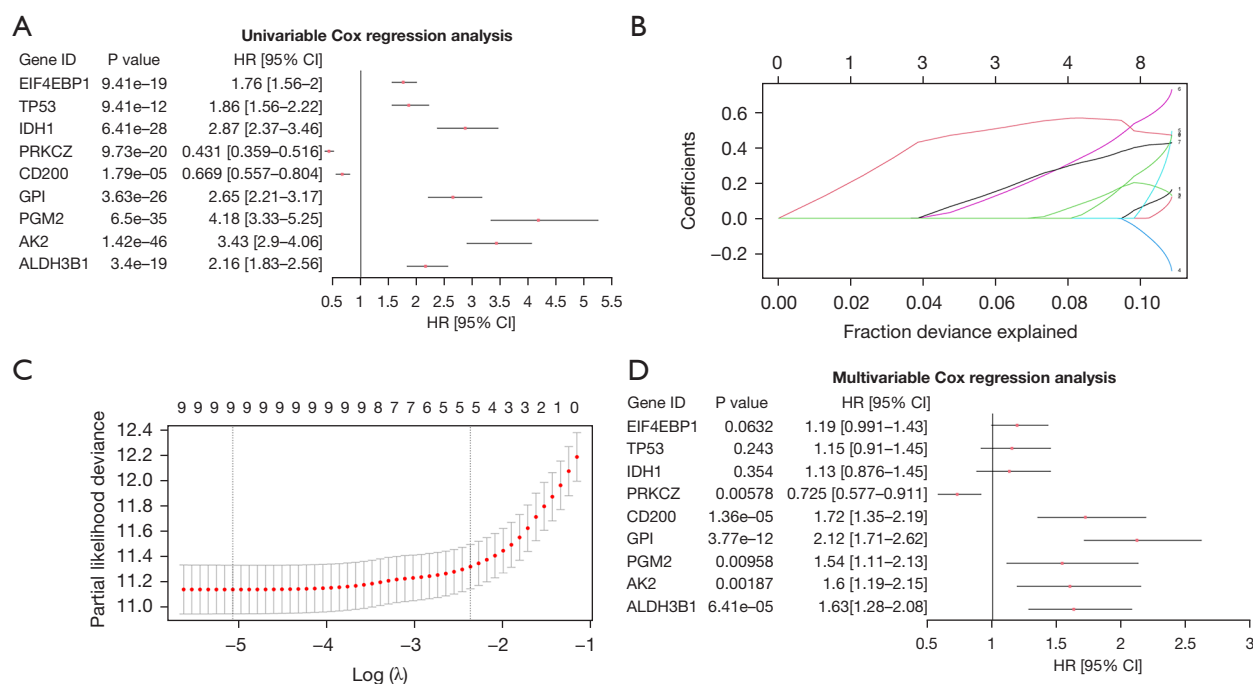


Figure 4 Outcomes of the LASSO and Cox regression analysis. (A) Forest plot of nine IR&MEMRDEGs in the UCR model. (B,C) The PRM diagram (B) and variable trajectory diagram (C) generated by the LASSO regression model. (D) Forest plot depicting three PRMGs in the MCR model. HR, hazard ratio; CI, confidence interval; LASSO, Least Absolute Shrinkage and Selection Operator; IR&MEMRDEGs, immune- and mitochondrial energy metabolism-related differentially expressed genes; UCR, univariate Cox regression; PRM, prognostic risk model; PRMGs, prognostic risk model genes; MCR, multivariate Cox regression.

performed based on the eleven IR&MEMRDEGs. The Forest plot (Figure 4A) presented all variables that were significant ($P < 0.10$ in the univariate analysis). Overall, nine IR and MEMRDEGs exhibited statistical significance in the single-factor Cox regression model ($P < 0.10$): *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *AK2*, and *ALDH3B1*.

To further assess the prognostic significance of IR&MEMRDEGs within the single-factor Cox regression model for glioma prognosis, LASSO regression analysis was conducted, as presented in the LASSO regression model diagram (Figure 4B), and the LASSO variable trajectory diagram (Figure 4C). Nine genes, referred to as PRMGs, were identified in the LASSO regression model: *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *AK2*, and *ALDH3B1*. Subsequently, an MCR analysis was performed, employing the nine PRMGs to ascertain the connection between LASSO RS expression levels and clinical prognosis (Figure 4D). The LASSO RS was computed using the following formula:

$$\begin{aligned} \text{Risk Score} = & \text{EIF4EBP1} * (0.154) + \text{TP53} * (0.11) + \text{IDH1} * (0.132) \\ & + \text{PRKCZ} * (-0.279) + \text{CD200} * (0.457) + \text{GPI} * (0.713) \\ & + \text{PGM2} * (0.427) + \text{AK2} * (0.472) + \text{ALDH3B1} * (0.452) \end{aligned} \quad [2]$$

Thereafter, the glioma specimens obtained from the TCGA-GBMLGG dataset were allocated to the HRG and LRG using the LASSO RS median as the criterion.

Prognostic analysis and validation of a PRM for glioma

The diagnostic performance of LASSO RS for OS was evaluated using the survival package in R to analyse and plot KM curves (Figure 5A). The results revealed a significant variation in OS between the HRG and LRG ($P < 0.05$). Additionally, ROC curves for 1-, 2-, and 3-year intervals are displayed for glioma samples from the TCGA-GBMLGG dataset (Figure 5B). The outcomes showed that LASSO RS could predict prognosis, with the best predictive effect in the third year (AUC = 0.898).

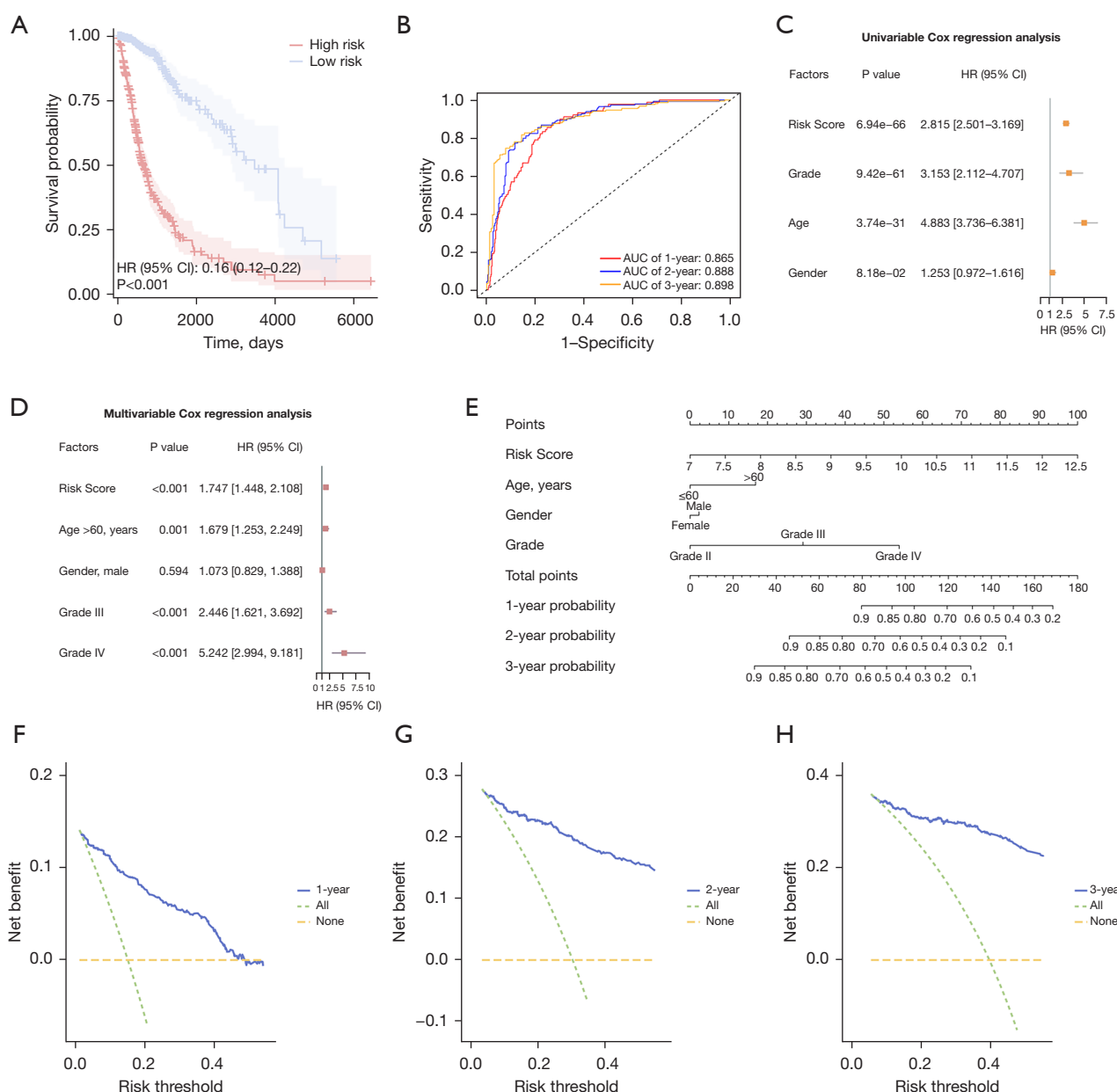


Figure 5 Analysis and validation of the PRM for glioma. (A) The connection between the LASSO RS in the HRG LRG and OS was illustrated in the KM curve. (B) ROC curves for the LASSO RS at 1-, 2-, and 3-year intervals. (C) Forest plot depicting the outcomes of UCR analysis, depending on RS, age, gender, and clinical stage (grade). (D) Forest plot depicting the outcomes of the MCR analysis of the PRM for glioma, including the RS, age, gender, and clinical stage (grade). (E) A nomogram representing the PRM. (F–H) The 1- (F), 2- (G), and 3-year (H) DCA graphs depicting the glioma PRM showed AUCs ranging from 0.7 to 0.9. HR, hazard ratio; AUC, area under the ROC curve; CI, confidence interval; PRM, prognostic risk model; LASSO, Least Absolute Shrinkage and Selection Operator; RS, risk score; HRG, high-risk group; LRG, low-risk group; OS, overall survival; KM, Kaplan-Meier; UCR, univariate Cox regression; MCR, multivariate Cox regression; DCA, decision curve analysis; ROC, receiver operating characteristic.

Table 2 Results of univariable and multivariable Cox analysis for TCGA-GBMLGG dataset

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P	Hazard ratio (95% CI)	P
Gender	671	1.253 (0.972–1.616)	0.082		
Female	283			Reference	
Male	388			1.073 (0.829–1.388)	0.594
Age (years)	671	4.883 (3.736–6.381)	<0.001		
≤60	531			Reference	
>60	140			1.679 (1.253–2.249)	0.001
Grade	671	3.153 (2.112–4.707)	<0.001		
Grade II	248			Reference	
Grade III	255			2.446 (1.621–3.692)	<0.001
Grade IV	168			5.242 (2.994–9.181)	<0.001
Risk score	671	2.815 (2.501–3.169)	<0.001	1.747 (1.448–2.108)	<0.001

TCGA, The Cancer Genome Atlas; GBMLGG, lower-grade glioma and glioblastoma multiforme; CI, confidence interval.

UCR analysis was performed on the LASSO RS and prognostic clinical data of glioma specimens sourced from TCGA-GBMLGG dataset, to ascertain the association and predictive capacity of LASSO RS with clinical outcomes. *Table 2* presents the outcomes of UCR and MCR analyses. UCR analysis used LASSO RS, age, sex, and clinical stage (grade) as variables. The MCR analysis subsequently encompassed variables that had a significance level of $P < 0.10$ in the univariate analysis. The results of both analyses are shown in a forest plot (*Figure 5C, 5D*). Furthermore, a nomogram was created based on PRM, incorporating LASSO RS, age, sex, and grade variables, illustrating their correlations (*Figure 5E*). The outcomes showed that LASSO RS exhibited notably greater predictive efficacy within the PRM for glioma than the other variables, whereas sex displayed significantly lower efficacy. The clinical utility of the glioma PRM was assessed using DCA at 1-, 3-, and 5-year. The PRM developed in this study demonstrated a superior clinical prediction effect at 3 years compared with that at 2 and 1 year (*Figure 5F–5H*).

Verification of gene expression differences in PRMs

To depict the disparities in the expression levels of the nine PRMGs between TG and NG of the TCGA-GBMLGG dataset and the CGD, a group comparison chart was constructed (*Figure 6A, 6B*). Statistically significant differences were observed in the nine PRMGs expression

levels between the NG and TG in the TCGA-GBMLGG dataset ($P < 0.01$, $P < 0.001$; *Figure 6A*). Specifically, *AK2*, *ALDH3B1*, *EIF4EBP1*, *IDH1*, *PGM2*, and *TP53* exhibited significantly higher expression levels in the TG than in the NG ($P < 0.01$, $P < 0.001$, respectively; *Figure 6A*). In contrast, TG significantly mitigated *CD200*, *GPI*, and *PRKCZ* expression levels compared to the NG ($P < 0.001$; *Figure 6A*). Additionally, eight PRMGs expression levels in the TG and NG of the CGD showed significant variation ($P < 0.05$, $P < 0.001$; *Figure 6B*). Notably, *AK2*, *ALDH3B1*, *EIF4EBP1*, *IDH1*, *PGM2*, and *TP53* exhibited significantly higher expression in the TG than in the NG ($P < 0.05$, $P < 0.001$; *Figure 6B*), whereas *CD200* and *PRKCZ* displayed significantly lower expression in the TG ($P < 0.001$; *Figure 6B*). Furthermore, glioma samples from the TCGA-GBMLGG dataset and CGD were classified into the HRG and LRG, depending on the median expression value of the LASSO RS of PRMGs. The analysis further revealed significant variations in nine PRMGs expression levels between the HRG and LRG in both datasets ($P < 0.001$; *Figure 6C, 6D*). Specifically, *AK2*, *ALDH3B1*, *EIF4EBP1*, *GPI*, *IDH1*, *PGM2*, and *TP53* exhibited significantly higher expression levels in the HRG than in the LRG for both datasets ($P < 0.001$; *Figure 6C, 6D*). Conversely, the expression levels of *CD200* and *PRKCZ* were significantly lower in the HRG than in the LRG in both datasets ($P < 0.001$; *Figure 6C, 6D*).

The prognostic accuracy of the nine genes was assessed

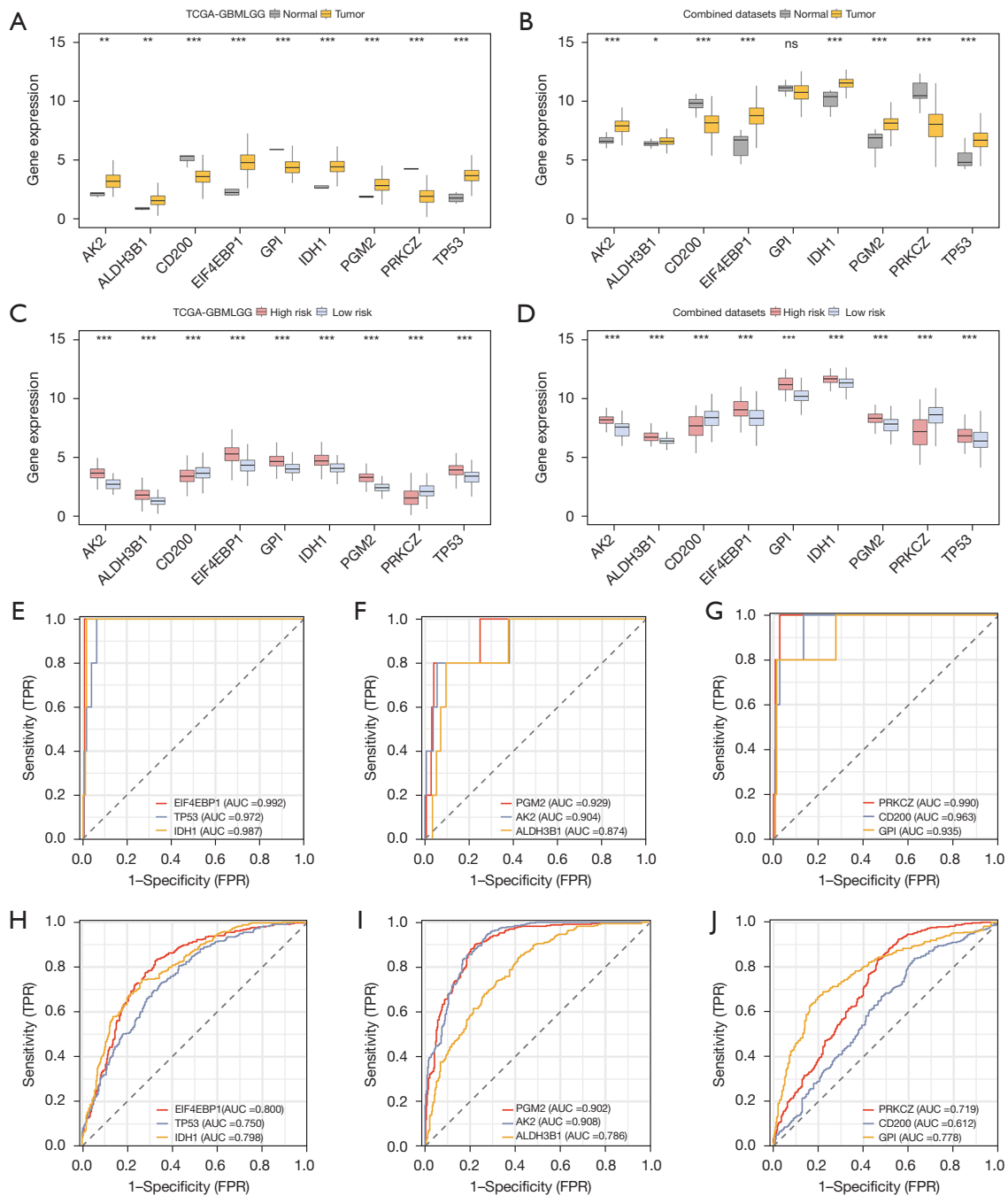


Figure 6 Validation of gene expression differences within PRMs and the ROC curve. (A,B) The comparison of PRMGs expression levels in the TG and NG from both the TCGA-GBMLGG dataset and the CGD. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, ns represents no statistical significance. (C,D) Comparison of PRMG expression levels in the HRG and LRG from the TCGA-GBMLGG dataset and the CGD. ***, $P < 0.001$. (E-J) ROC curves of PRMGs were analyzed to compare the differences in their prognostic accuracy between the TG and NG (E-G), as well as between the HRG and LRG (H-J) in the TCGA-GBMLGG dataset. AUC values between 0.5 and 0.7, 0.7 and 0.9, and above 0.9 indicate low, moderate, and high degrees of accuracy, respectively. TCGA, The Cancer Genome Atlas; GBMLGG, lower-grade glioma and glioblastoma multiforme; AUC, area under the ROC curve; TPR, true positive rate; FPR, false positive rate; PRMs, prognostic risk models; ROC, receiver operating characteristic; PRMGs, prognostic risk model genes; TG, tumour group; NG, normal group; CGD, Combined GEO Dataset; HRG, high-risk group; LRG, low-risk group.

using the pROC package in R. The ROC curve results revealed that eight of these PRMGs exhibited high prognostic accuracy ($AUC > 0.9$) in distinguishing between TG and NG within the TCGA-GBMLGG dataset, including *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, and *AK2* (Figure 6E-6G). However, *ALDH3B1* demonstrated a marginally lower accuracy ($0.7 < AUC < 0.9$; Figure 6E-6G). In addition, the analysis revealed that two PRMGs, *PGM2* and *AK2*, exhibited superior accuracy ($AUC > 0.9$) in distinguishing between the HRG and LRG of glioma samples within the TCGA-GBMLGG dataset (Figure 6H-6J). Conversely, six PRMGs (*EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *GPI*, and *ALDH3B1*) demonstrated slightly lower accuracy ($0.7 < AUC < 0.9$; Figure 6H-6J), while *CD200* displayed poor accuracy ($0.5 < AUC < 0.7$; Figure 6H-6J) in discriminating between HRG and LRG in the TCGA-GBMLGG dataset.

Immune infiltration analysis in the HRG and LRG

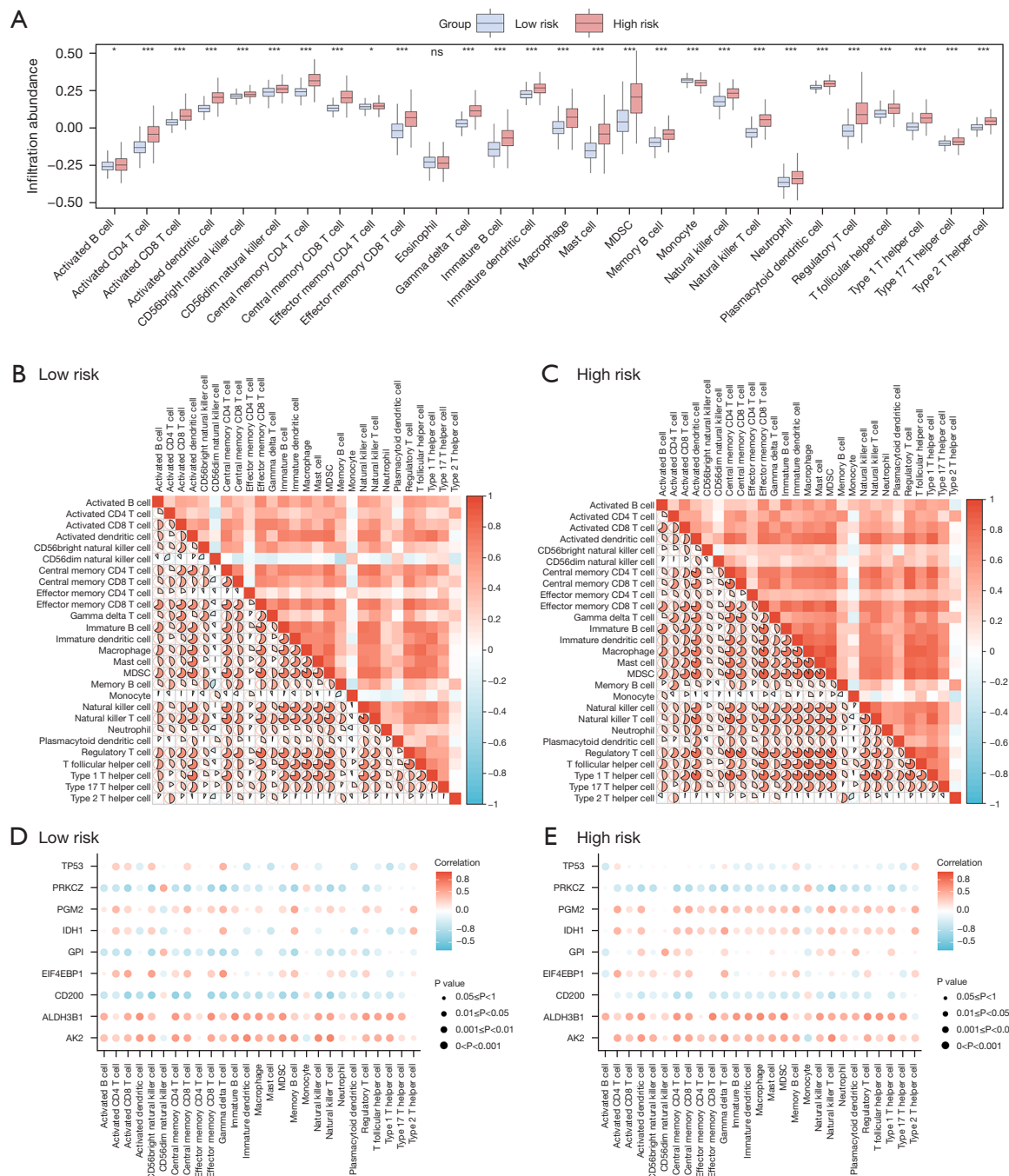
Immune infiltration was compared between the HRG and LRG using the expression matrix of glioma samples from the TCGA-GBMLGG dataset. The ssGSEA was applied to determine the infiltrations of 28 immune cells in both groups. Initially, immune cells were screened based on a significance threshold of $P < 0.05$, using comparison analysis, which revealed disparities in the levels of immune cell infiltration between the two groups. The group comparison plot (Figure 7A) indicated that 27 immune cells exhibited significant disparities ($P < 0.05$, $P < 0.001$), the only exception being monocytes. The infiltration abundances of all other cell types were significantly higher in the HRG than in the LRG. These cell types included activated cells, such as B, CD4 T, CD8 T, dendritic, and CD56bright dendritic cells; CD56dim natural killer, central memory CD4 T, central memory CD8 T cells, effector memory CD4 T cells, effector memory CD8 T cells, gamma delta T, immature B cells, immature dendritic cells, macrophages, mast cells, myeloid-derived suppressor cells (MDSCs), memory B, and natural killer cells. Subsequently, correlation analysis of the infiltration abundance of 27 immune cells in the immune infiltration assessment between the HRG and LRG was visually represented using a correlation heat map (Figure 7B, 7C). These findings revealed a predominantly positive correlation between immune cells in the HRG and the LRG. Additionally, the connection between PRMGs and the amount of immune cell infiltration is illustrated by a correlation dot plot (Figure 7D, 7E). Correlation plot

analysis further showed that within the glioma samples of the TCGA-GBMLGG dataset, *ALDH3B1* and *AK2* positively correlated with the majority of immune cells in both the HRG and LRG ($r > 0$). Conversely, *CD200* negatively correlated ($r < 0$) with most immune cells in both the HRG and LRG.

Discussion

Gliomas are the most prevalent primary brain tumours. This malignancy is associated with a high mortality rate and unfavourable prognosis, particularly in cases of malignant glioblastoma, where the OS period is only 15 months (22). Research has indicated that the prognosis of gliomas is intricately linked to alterations in gene expression within tumour cells (24). However, no standardised risk-prediction model has yet been established based on IR&MEMRDEGs for gliomas. Therefore, in the present study, we performed a bioinformatics analysis of the GEO and TCGA databases to identify 11 genes that displayed dysregulated expression in gliomas, and were associated with the immune response and mitochondrial energy metabolism. These genes included *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *PKLR*, *AK2*, *ATP4A*, and *ALDH3B1*. The TCGA-GBMLGG dataset and CGD had significantly greater *AK2*, *ALDH3B1*, *EIF4EBP1*, *IDH1*, *PGM2*, and *TP53* expression levels in the TG than in the NG ($P < 0.01$, $P < 0.001$), whereas the TG showed a significant hindrance in *CD200* and *PRKCZ* expression levels compared with the NG ($P < 0.001$). In the TCGA-GBMLGG dataset, TG showed a significant reduction in *GPI* expression compared to NG. Nevertheless, no significant variation in *GPI* expression was observed between the TG and NG in the CGD.

UCR, LASSO regression, and forest plot analyses identified *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *AK2*, and *ALDH3B1* as independent prognostic factors for glioma. A predictive model for glioma prognosis was constructed based on these factors. The samples were subsequently stratified into HRG and LRG using LASSO RS, while the expression differences of the nine genes in these groups were calculated and visualized. The results revealed a statistically significant elevation in the expression levels of *AK2*, *ALDH3B1*, *EIF4EBP1*, *GPI*, *IDH1*, *PGM2*, and *TP53* in the HRG compared to those in the LRG ($P < 0.001$), indicating that the expression of these seven genes was negatively linked to patient prognosis. Conversely, HRG significantly mitigated *CD200* and *PRKCZ* expression levels as opposed to LRG in both



datasets ($P < 0.05$). These findings demonstrate a favourable correlation between *CD200* and *PRKCZ* expression levels and patient prognosis.

Moreover, the predictive efficacy of models based on IR and MEMRDEG for prognostic outcomes in glioma patients was validated by KM and ROC curve analyses utilising clinical data sourced from TCGA and UCSC Xena databases. These findings suggest that our PRM utilising nine genes (*EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *CD200*, *GPI*, *PGM2*, *ALDH3B1*, and *AK2*) demonstrated high accuracy in predicting the prognosis of patients with glioma (AUC > 0.7). Further investigation of these genes revealed that *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *GPI*, *PGM2*, *ALDH3B1*, and *AK2* displayed superior predictive accuracy (AUC > 0.7) in distinguishing between HRG and LRG of glioma samples. Overall, *EIF4EBP1*, *TP53*, *IDH1*, *PRKCZ*, *GPI*, *PGM2*, *ALDH3B1*, and *AK2* were identified as independent prognostic factors. PRMs based on these factors have demonstrated superior accuracy in predicting prognosis in the treatment of gliomas.

In addition to playing an essential role in facilitating energy production, mitochondrial metabolism impacts tumour immunotherapy by modulating immune cell functions within the tumour microenvironment. Prior research has shown that the efficacy of immunotherapy depends on adequate immune cell infiltration of the tumor microenvironment (25). This study further explored the correlation between nine distinct prognostic factors and immune cell infiltration in groups with high and low expression levels, with the aim of identifying potential targets for immunotherapy and prognostic markers. The results indicated that *IDH1* expression negatively correlated with the number of immune cells infiltrating the LRG, but a positive correlation was observed in the HRG ($r > 0$). These results further indicate that *IDH1* expression levels may have a distinct influence on the tumour immune microenvironment.

Conclusions

In conclusion, the results of the present study confirmed the significant predictive value of *IDH1* in assessing the prognosis of glioma, demonstrating high accuracy as an independent prognostic factor in distinguishing between TG and NG (AUC = 0.987) and displaying moderate accuracy in distinguishing between HRG and LRG within the TG (AUC = 0.798). Furthermore, our results highlight the involvement of *IDH1* in modulating the abundance of

immune-infiltrating cells in the immune microenvironment. Ultimately, *IDH1* plays a dual role in the prognostic prediction of gliomas and influences immunotherapy for glioma treatment.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1035/rc>

Peer Review File: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1035/prf>

Funding: This project was supported by the National Natural Science Foundation of China (Nos. 82302908 to D.H., and 82303926 to Y.L.); the Programs of Science and Technology Commission Foundation of Tianjin Municipality (16JCQNJC13400 to D.H.), the China Scholarship Council (201606945004 to D.H.), the Fundamental Research Funds for the Central Universities (3332021084 to D.H.), and the “Excellent New Star” Cultivation Project of Tianjin Medical University General Hospital (2020 to D.H.).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1035/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* 2021;23:1231-51.
- Zhang Q, Guan G, Cheng P, et al. Characterization of an endoplasmic reticulum stress-related signature to evaluate immune features and predict prognosis in glioma. *J Cell Mol Med* 2021;25:3870-84.
- Xu S, Tang L, Li X, et al. Immunotherapy for glioma: Current management and future application. *Cancer Lett* 2020;476:1-12.
- Lee M, Yoo JH, Kim I, et al. The compartment-specific manipulation of the NAD(+)/NADH ratio affects the metabolome and the function of glioblastoma. *Sci Rep* 2024;14:20575.
- Xia J, Yang Q, Wang C, et al. Comprehensive analysis to identify the relationship between CALD1 and immune infiltration in glioma. *Transl Cancer Res* 2024;13:3354-69.
- Goff KM, Zheng C, Alonso-Basanta M. Proton radiotherapy for glioma and glioblastoma. *Chin Clin Oncol* 2022;11:46.
- Nabors LB, Portnow J, Ammirati M, et al. NCCN Guidelines Insights: Central Nervous System Cancers, Version 1.2017. *J Natl Compr Canc Netw* 2017;15:1331-45.
- Li R, Chen Y, Yang B, et al. Integrated bioinformatics analysis and experimental validation identified CDCA families as prognostic biomarkers and sensitive indicators for rapamycin treatment of glioma. *PLoS One* 2024;19:e0295346.
- Qin H, Abulaiti A, Maimaiti A, et al. Integrated machine learning survival framework develops a prognostic model based on inter-crosstalk definition of mitochondrial function and cell death patterns in a large multicenter cohort for lower-grade glioma. *J Transl Med* 2023;21:588.
- Sainero-Alcolado L, Liaño-Pons J, Ruiz-Pérez MV, et al. Targeting mitochondrial metabolism for precision medicine in cancer. *Cell Death Differ* 2022;29:1304-17.
- Liu Y, Cai L, Wang H, et al. Novel mitochondrial-related gene signature predicts prognosis and immunological status in glioma. *Transl Cancer Res* 2024;13:3338-53.
- Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007;23:1846-7.
- Griesinger AM, Birks DK, Donson AM, et al. Characterization of distinct immunophenotypes across pediatric brain tumor types. *J Immunol* 2013;191:4880-8.
- Gravendeel LA, Kouwenhoven MC, Gevaert O, et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* 2009;69:9065-72.
- Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882-3.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Ben Salem K, Ben Abdelaziz A. Principal Component Analysis (PCA). *Tunis Med* 2021;99:383-9.
- Colaprico A, Silva TC, Olsen C, et al. TCGAblinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44:e71.
- Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1-1.30.33.
- Therneau T. A package for survival analysis in R. R package version 3.5-0. Available online: <https://CRAN.R-project.org/package=survival>
- Engelbrechtsen S, Böhlin J. Statistical predictions with glmnet. *Clin Epigenetics* 2019;11:123.
- Rich JT, Neely JG, Paniello RC, et al. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* 2010;143:331-6.
- Wu J, Zhang H, Li L, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun (Lond)* 2020;40:301-12.
- Papacoea SI, Vrinceanu D, Dumitru M, et al. Molecular Profile as an Outcome Predictor in Glioblastoma along with MRI Features and Surgical Resection: A Scoping Review. *Int J Mol Sci* 2024;25:9714.
- Chae YK, Arya A, Iams W, et al. Current landscape and future of dual anti-CTLA4 and PD-1/PD-L1 blockade immunotherapy in cancer; lessons learned from clinical trials with melanoma and non-small cell lung cancer (NSCLC). *J Immunother Cancer* 2018;6:39.

Cite this article as: Lin Y, Li H, Ge Q, Hua D. Establishment and validation of a prognostic prediction model for glioma based on key genes and clinical factors. *Transl Cancer Res* 2025;14(1):240-253. doi: 10.21037/tcr-24-1035