Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

Identification and validation of a novel 9-gene signature of non-specific classification to predict prognosis in glioma patients

Guangzhao Li [1,2,#], Xiaowang Niu[3,#], Xiang Li[4], Bin Lin[2], Fei Yang[2], Zhong Wang[1,*]

[1] Department of Neurosurgery & Brain and Nerve Research Laboratory, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu Province, 215006, China
[2] Department of Neurosurgery, Hefei First People's Hospital, Hefei, Anhui Province, 230041, China
[3] Department of Neurosurgery, Suqian Hospital Affiliated to Xuzhou Medical University, Suqian, Jiangsu Province, 223800, China
[4] Department of Neurosurgery, Xinghua People's Hospital, Xinghua, Jiangsu Province, 225700, China

Article Info

Article history:

Received: November 12, 2023
Accepted: January 17, 2024
Published: January 31, 2024

Use your device to scan and read the article online

Abstract

This study aimed to identify and validate a 9-gene signature for predicting overall survival (OS) in glioma patients. Analysis of multiple gene expression datasets led to the identification of 135 candidate genes associated with OS in glioma patients. Further analysis revealed that IGFBP2, PBK, NRXN3, TGIF1, DNAJA4, and LGALS3BP were identified as risk factors for OS, while ENAH, PPP2R2C, and SPHKAP were found to be protective factors. Multifaceted validation using different databases confirmed their differential expression patterns in glioma tissues compared to normal brain tissue. By utilizing LASSO regression and multivariate Cox regression analysis, a risk score was developed based on the expression levels of the 9 crucial genes. The risk score showed a significant correlation with OS in both training and validation cohorts and yielded superior predictive accuracy compared to individual gene expression. Moreover, a predictive nomogram incorporating the risk score, WHO grade, age, IDH mutation, and 1p/19q co-deletion was constructed and validated, which exhibited high predictive capabilities for survival rates at different time points. Enrichment analysis revealed the involvement of extracellular matrix-related pathways and immune system signaling in glioma prognosis. Furthermore, the risk score showed a strong correlation with immune cell infiltration and immune checkpoint expression, suggesting its potential role in the tumor immune microenvironment. In conclusion, our study provides a robust 9-gene signature and a predictive nomogram for evaluating the prognosis of glioma patients, offering valuable insights into personalized treatment strategies.

Keywords: Glioma, Overall survival, Gene signature, Risk score, Nomogram, Immune cell infiltration

1. Introduction

Gliomas are a large group of brain and spinal cord tumors that originate from glial cells.

These brain tumors usually occur in people between the ages of four and sixty, although some types are more common in children. Men are slightly more likely to develop brain tumors [1,2].

History of radiation exposure is a risk factor for malignant glioma. Certain genetic disorders also increase the risk of these tumors in children, but they are rare in adults. Several lifestyle risk factors have been studied in association with malignant glioma, including smoking or cell phone use. The symptoms, prognosis, and treatment of malignant gliomas depend on the age of the patient, the exact type of tumor, and the location of the tumor in the brain. These tumors grow and invade normal brain tissue, making surgical removal very difficult and sometimes impossible and complicating treatment. Therefore, it is important to find a neurosurgeon with expertise in surgical diagnosis and treatment [1-3].

Despite thorough examination of glioma, the current therapies for this highly prevalent cancer of the brain and spinal cord are still insufficient. For the purpose of enhancing patient outcomes, it is imperative to determine biomarkers and therapeutic targets for glioma [1].

With the progress of high-throughput sequencing methods, the significance of bioinformatic analysis has grown in the quest for glioma oncogenes [2]. Several biomarkers have been discovered that can be utilized for forecasting the prognosis of cancer, providing an understanding of the molecular foundation of glioma and possible novel targets for therapy [3]. Numerous investigations have concentrated on the prognostic indicators of individual gene expression levels; nevertheless, these findings might lack credibility due to their susceptibility to external influences, thereby failing to accurately depict the overall survival (OS) of glioma patients [4]. Gene signature-based prognostic models have gained significant attention in recent years for their capacity to accurately predict patient outcomes by utilizing statistical methods to eliminate collinearity in gene expression [5]. Nevertheless, the existing gene signatures primarily originate from particular gene

* Corresponding author.
E-mail address: wangzhong_761@163.com (Z. Wang).
# These authors contributed equally
Doi: http://dx.doi.org/10.14715/cmb/2024.70.1.18

clusters, including those associated with copper metabolism, fatty acid metabolism, ferroptosis, pyroptosis, immunity, and inflammation [6]. Such models are limited in that they only consider certain aspects of glioma characteristics, and glioma heterogeneity may obscure the predictive value of the model. In order to tackle this problem, we examined the transcriptomic information and clinical data of glioma from publicly available data sources. Subsequently, we developed a risk score consisting of novel gene markers without bias, which could potentially offer fresh perspectives on the management of individuals with glioma.

2. Materials and methods

2.1. Acquisition of the data

GEO database (The Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo) offered the RNA sequencing of 395 brain tumors from the GSE4290, GSE50161, GSE74195, GSE104291, and GSE29796 datasets (Table 1). Additionally, the RNA sequencing and clinical information of 702 glioma patients were provided by the TCGA database (The Cancer Genome Atlas, https://portal.gdc.cancer.gov/). RNA sequencing and clinical data from 693 and 325 individuals were obtained from the CGGA database (The Chinese Glioma Genome Atlas, http://www.cgga.org.cn/), while data from 475 glioma patients were acquired from the Rembrandt database (Downloaded from the CGGA database). Furthermore, data from the control group were derived from the GTEx database (The Genotype-Tissue Expression, https://xenabrowser.net/datapages/) and CGGA database.

2.2. Screening of the candidate gene sets

By applying the screening criteria of | logFC |> 1 and padj < 0.05, a screening process was conducted to identify differentially expressed genes (DEGs) for GSE4290, GSE50161, GSE74195, GSE104291, and GSE29796. The intersection of these genes from all five datasets was defined as the GSE differential gene sets. Genes with significant prognostic value for TCGA-OS were identified using batch fit survival analysis using TCGA transcriptome data separated by median mRNA expression. A candidate gene set was compiled by combining the GSE differential gene set with the TCGA-OS-associated gene set.

2.3. Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment

Functional annotations and pathway enrichment analyses were performed using the R package 'clusterProfiler' to evaluate the biological functions of the candidate gene sets. The analysis included GO terms such as molecular function (MF), cellular component (CC), and biological process (BP), as well as KEGG pathways.

2.4. Creating and validating a risk score

For the training cohort, the TCGA database was used, and CGGA and Rembrandt databases were used for the external validation cohort. The candidate gene set was confirmed again using univariate Cox regression. To select the best combinations of DEGs, we utilized the LASSO method. Afterwards, a multivariate Cox regression analysis was utilized to detect the crucial genes and construct a risk score using the expression levels of those genes and their regression coefficients. The risk score equation for every glioma patient was derived as Risk score = (βmRNA1*expmRNA1 + βmRNA2*expmRNA2 + … + βmRNAn*expmRNAn). Glioma patients were categorized into two groups, namely the high and low-risk score group. A Kaplan-Meier analysis was performed to compare the OS of patients in the high- and low-risk score groups. The Area Under of Curve (AUC) was calculated using the Receiver Operating Characteristic (ROC) method to evaluate the accuracy of the risk score. Furthermore, the precision of the established risk score was thoroughly evaluated in two external validation cohorts.

We assessed the suitability of the risk score by comparing the outcome of glioma patients within the identical clinical subgroup sourced from the TCGA database. Furthermore, we performed GO, KEGG, and Gene Set Enrichment Analysis (GSEA) on the DEGs between the high-and low-risk score groups, utilizing the Rembrandt database and applying the criteria of padj < 0.05 and | logFC |> 1.

2.5. Analyzing gene expression, survival value, and gene mutations of crucial genes

By utilizing the TCGA database, we analyzed the mRNA expression levels of crucial genes that contribute to the risk score, examining their correlation with clinical characteristics. To confirm the mRNA expression levels of these genes, we utilized single-cell sequencing data from GSE139448 in the TISCH2 database (Tumor Immune Single-cell Hub 2, http://tisch.comp-genomics.org). Furthermore, we examined the protein expression levels of the crucial genes in glioma using the UALCAN database (The University of Alabama at Birmingham Cancer Data Analysis Portal, https://ualcan.path.uab.edu/analysis-prot.html) and the HPA database (The Human Protein Atlas, https://www.proteinatlas.org). We performed K-M survival analysis to evaluate the impact of mRNA expression levels of the crucial genes on patients' OS. In order to enhance our comprehension of the genomics map of the crucial genes, we performed gene mutation analysis by utilizing the cBioPortal online database (http://www.cbioportal.org).

2.6. Construction of a Nomogram

We assessed if our established risk score could function

Table 1. Characteristics of Brain tumor patients in GEO sets.

GEO	platform	Sample
GSE4290	GPL570	Glioma:157, Brain tissue:23
GSE104291	GPL570	Glioma:24, Brain tissue:2
GSE29796	GPL570	Glioma:51, Brain tissue:21
GSE50161	GPL570	Brain tumor: 117, Brain tissue: 13
GSE74195	GPL570	Brain tumor: 46, Brain tissue: 5

as an independent prognostic factors in glioma patients in the TCGA database by employing Cox regression analysis. Using the 'rms' package, a prognostic nomogram was developed to forecast the OS at 1-, 3-, and 5- years in glioma patients, considering the clinical factors and risk score obtained from the multivariate Cox regression analysis. A calibration chart was utilized to evaluate the predictive capability of the nomogram. The Decision Curve Analysis (DCA) was implemented to evaluate the clinical net benefit using the 'rmda' package. The analysis was also confirmed in the external validation cohort CGGA database.

2.7. Evaluation of infiltration of immune cells in tumors

Using the Single-sample GSEA (ssGSEA) algorithm, 24 immune cell markers were computed to examine the infiltration of tumor immune cells. The matrix, immune, and estimate scores for both cohorts were determined using the Estimate algorithm. Additionally, we conducted a more in-depth examination of the correlation between 46 prevalent immune checkpoints, chemokines, and the risk score.

2.8. Statistical analysis

The R software (4.2.1) was utilized to compute all statistical analysis outcomes. To create a risk score, we utilized univariate Cox regression, LASSO, and multivariate Cox regression analysis. Pearson's correlation analysis was used to identify correlations, while the chi-square and t-test were employed to compare clinical features. The K-M method was utilized to generate OS curves, which were subsequently assessed using the log-rank analysis. Statistical significance was indicated by a P-value of less than 0.05, with all conducted statistical tests being two-sided.

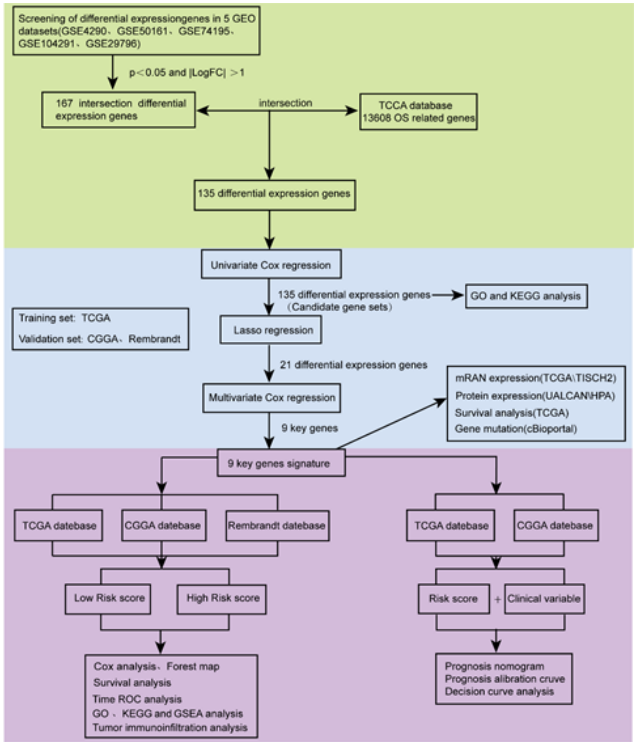

Fig. 1. Flow chart of this study. GEO: Gene Expression Omnibus. TCGA: The Cancer Genome Atlas. CGGA: The Chinese Glioma Genome Atlas. DEGs: Differentially Expressed genes. LASSO: the least absolute contraction and selection operator. GO: Gene Ontology. KEGG: Kyoto Encyclopedia of Genes and Genomes. GSEA: Gene Set Enrichment Analysis. ROC: Receiver Operating Characteristic. TISCH2: Tumor Immune Single-cell Hub 2 database. UALCAN: The University of Alabama at Birmingham Cancer Data Analysis Portal database. HPA: The Human Protein Atlas database.

3. Results

3.1. Investigating and selecting candidate genes.

Figure 1 displays a diagrammatic depiction of our investigation. We identified 167 intersection genes from the 5 GEO datasets (Figure 2A). Subsequently, 13,608 genes with significant influence on OS in glioma patients were identified from the TCGA database. We found 135 candidate genes that were shared between the GSE differential gene set and the TCGA-OS-associated gene set (Figure 2B). The impact of these 135 candidate genes on the OS of TCGA patients was reaffirmed through univariate Cox regression analysis (Figure S1A). In addition, the analysis of these 135 potential genes using GO and KEGG analysis revealed their strong association with the segregation of sister chromatids during mitosis, chromosomal region, binding to microtubules, and the cell cycle (Figure S1B-C).

3.2. Construction of a risk score

In order to determine the best gene combinations, we utilized LASSO regression on the 135 genes identified through univariate Cox regression, ultimately selecting 21 genes (Figure 2C-D). A risk score was established by selecting 9 crucial genes using multivariate Cox regression analysis (Figure 2E). In patients with glioma, IGFBP2, PBK, NRXN3, TGIF 1, DNAJA4, and LGALS3BP were identified as risk factors for OS, whereas ENAH, PPP2R2C, and SPHKAP were found to be protective factors. These 9 crucial genes also showed positive or negative correlations with each other (Figure 2F).

The risk score for every individual was computed by utilizing the regression coefficients (β) obtained from the multivariate Cox regression analysis. Risk score = 0.4721*IGFBP2 + 0.7673*PBK -0.6397*PPP2R2C + 0.7849*NRXN3 + 0.3540*DNAJA4 + 1.0905*TGIF1 - 0.6315*SPHKAP - 0.7273*ENAH + 0.4514*LGALS3BP.

3.3. Multifaceted validation the 9 crucial genes

The TCGA database showed significant increases in mRNA expression in IGFBP2, PBK, TGIF1, ENAH, and LGALS3BP, while PPP2R2C, NRXN3, DNAJA4, and SPHKAP were significantly decreased (Figure S2A). In the CGGA database, the NRXN3 and DNAJA4 mRNA expression levels reduced, but the ENAH expression was not significantly different (Figure S2B). As for the highly expressed genes, excluded ENAH, IGFBP2, PBK, TGIF, and LGALS3BP all increased with the WHO grade. The genes PPP2R2C, NRXN3, and SPHKAP showed a decrease in expression with the WHO grade. Additionally, a significant difference was observed between 9 crucial genes and the presence of IDH mutation and 1p/19q co-deletion, respectively (Figure S2C-K).

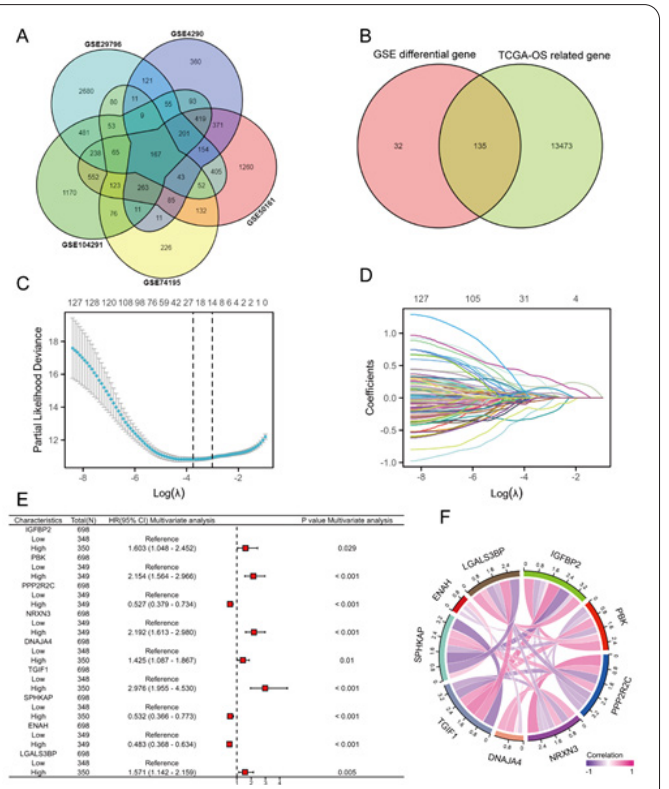

Fig. 2. Screening of the crucial genes. (A): The intersection of 5 GEO datasets yields 167 differentially expressed genes in GSE. (B): The intersection of DEGs in GSE and DEGs in the TCGA database gives a total set of 135 candidate genes. (C): Cross-validation for optimal parameter selection in the LASSO regression model. (D): Track plot of the prognostic lasso variables. (E): Multivariate Cox regression analysis selected 9 crucial genes to build a prognostic risk score. (F): Chords chart of the correlation of the 9 key genes. GEO: Gene Expression Omnibus. DEGs: DEGs: Differentially Expressed genes. TCGA: The Cancer Genome Atlas.

Survival analysis revealed that glioma patients with elevated levels of IGFBP2, PBK, TGIF1, DNAJA4, and LGALS3BP exhibited an unfavorable prognosis, whereas those with high expression of PPP2R2C, NRXN3, ENAH, and SPHKAP had a prolonged survival period, suggesting their potential role as tumor suppressors (Figure S3A-I). Interestingly, NRXN3 paradoxically emerged as a prognostic risk factor in the Cox regression analysis, while survival analysis indicated its possible protective effect. Consequently, we confirmed that increased NRXN3 expression correlates with longer overall survival in the CGGA database (693 and 325_mRNAseq cohorts) (Figure S3J-K). Additionally, NRXN3 was identified as a potential cancer inhibitor in the Rembrandt database (Figure S3L).

According to Figure S4, the UALCAN database revealed that 9 crucial genes exhibited protein expression levels that were in line with their mRNA expression levels. In glioma, the levels of IGFBP2, PBK, TGIF1, ENAH, and LGALS3BP were markedly elevated, whereas PPP2R2C, NRXN3, DNAJA4, and SPHKAP exhibited significant reduction. Protein expression was confirmed using immunohistochemistry, as shown in Figure S5. Glioma tissues exhibited higher staining for PBK, TGIF1, ENAH, and LGALS3BP antibodies compared to normal brain tissue, while staining for PPP2R2C, NRXN3, DNAJA4, and SPHKAP antibodies was reduced.

In order to confirm the aforementioned findings, the expression of 9 crucial genes in the TISCH2 database's single-cell transcriptome data aligned with the results obtained from the TCGA database (Figure S6A-J). The 9 crucial genes identified as significantly differently expressed in glioma samples have a low occurrence of mutations in the cBioPortal online database, indicating that the risk score was relatively cautious and exhibited strong stability (Figure S6K).

3.4. Assessment and verification of the predictive accuracy

Both univariate and multivariate Cox regression analysis indicated a significant correlation between the risk score and OS in patients (Figure 3A). According to the K-M analysis, individuals with a decreased risk score ex-

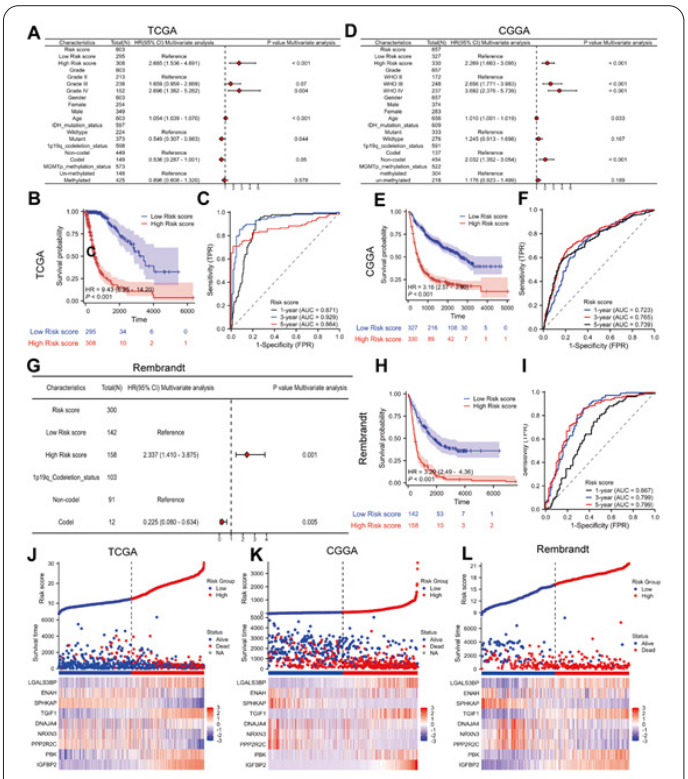

Fig. 3. Construction and validation of the risk score for 9 crucial genes. (A-B): Multivariate Cox regression analysis shows that the risk score is an independent risk factor for patient outcome in the TCGA and CGGA databases. (C): Survival analysis of the TCGA database showed a significant reduction in OS in the high-risk score group. (D): Time-dependent ROC curve analysis shows that the AUC of risk score OS prediction for patients 1-,3- and 5- year in TCGA database is 0.871,0.929 and 0.864, respectively. (E): Survival analysis of the CGGA database showed a similarly significantly lower OS of patients in the high-risk score group. (F): Time-dependent ROC curve analysis shows that the AUC of risk score prediction for patient 1-, 3- and 5-year OS in CGGA database is: 0.723, 0.765, and 0.739, respectively. (G): Multivariate Cox regression analysis showed that the risk score was a risk factor independent of patient outcome in the Rembrandt database. (H): Survival analysis of the Rembrandt database showed significantly lower OS in the high-risk score group. (I): Time-dependent ROC curve analysis shows that the AUC of risk score prediction for patient 1-, 3- and 5-year OS in Rembrandt database is 0.667, 0.799 and 0.799, respectively. (J-L): Risk factors in the TCGA, CGGA, and the Rembrandt database. The corresponding heat map shows the expression distribution of the 9 crucial genes. TCGA: The Cancer Genome Atlas. CGGA: The Chinese Glioma Genome Atlas. OS: Overall survival. ROC: Receiver Operating Characteristic. AUC: Area Under the Curve.

hibited a notably superior rate of survival in comparison to those with an increased risk score (Figure 3B). The predictive accuracy of the risk score for patient 1-, 3-, and 5-year OS was further confirmed through time-dependent ROC analysis, showing AUC values of 0.871, 0.829, and 0.864, respectively (Figure 3C).

To assess the precision of the risk score in predicting prognosis in additional groups, we performed an external validation to evaluate its predictive capability using the CGGA and Rembrandt databases. The risk score exhibited AUC values exceeding 0.6 at 1-, 3-, and 5-year intervals for OS in both validation cohorts, demonstrating exceptional predictive ability (Figure 3D-I). The risk score for OS had a higher prediction performance in both TCGA and CGGA databases compared to a single crucial gene (Figure S7).

Furthermore, the risk score of every patient in both the training and validation cohorts was assessed and ordered. As the risk score increased, the glioma patients experienced a decrease in survival time and an increase in mortality events. Figure 3J-L displays the distribution of gene expression for the 9 crucial genes among the high- and low-risk score groups, as depicted by the heat map.

3.5. Creating a predictive nomogram

Using the TCGA database, a predictive nomogram was developed to estimate the survival rates of glioma patients for 1, 3, and 5 years. Figure 4A displayed the inclusion of risk score, WHO grade, age, IDH mutation, and 1p/19q co-deletion in the model. The calibration plot indicated that the nomogram exhibited a high level of agreement with the real observations, with a C-index of 0.867 (0.855-0.879) (Figure 4B). Furthermore, the value of the nomogram for clinical decisions was assessed through DCA analysis, revealing that the 1, 3, and 5-year nomogram provided superior net benefits compared to solely relying on clinical features (Figure 4C-E).

We constructed a prognostic nomogram as an external validation using the same variables from the CGGA database. The findings suggested that the nomogram successfully predicted the OS of the validation set, as shown in Figure 4F. According to the calibration plot, the nomogram accurately predicted the OS of glioma patients at 1, 3, and 5 years, aligning with the real-life observations. Additionally, the C-index was recorded as 0.764(0.752-0.776) in Figure 4G. The analysis of the DAC also reflected similar findings to the training group (Figure 4H-J).

3.6. Performing enrichment analysis on DEGs

The prognosis of glioma patients in different subgroups, including age, gender, IDH mutation, and 1p/19q co-deletion, was assessed through survival analysis, revealing a negative correlation between a higher risk score and prognosis (Figure 5A-I). Analysis of the DEGs in the high- and low-risk score groups using the GO function and KEGG pathway revealed a correlation between extracellular matrix organization and BP. The extracellular matrix structural components and inhibitors of enzyme activity were greatly enriched in MF. The extracellular matrix containing collagen was enriched in CC. Furthermore, the KEGG pathway analysis revealed significant enrichment in the pathways of ECM-receptor interaction, Focal adhesion, and complement and coagulation cascades (Figure 5J, Table 2). According to the GSEA findings, the group

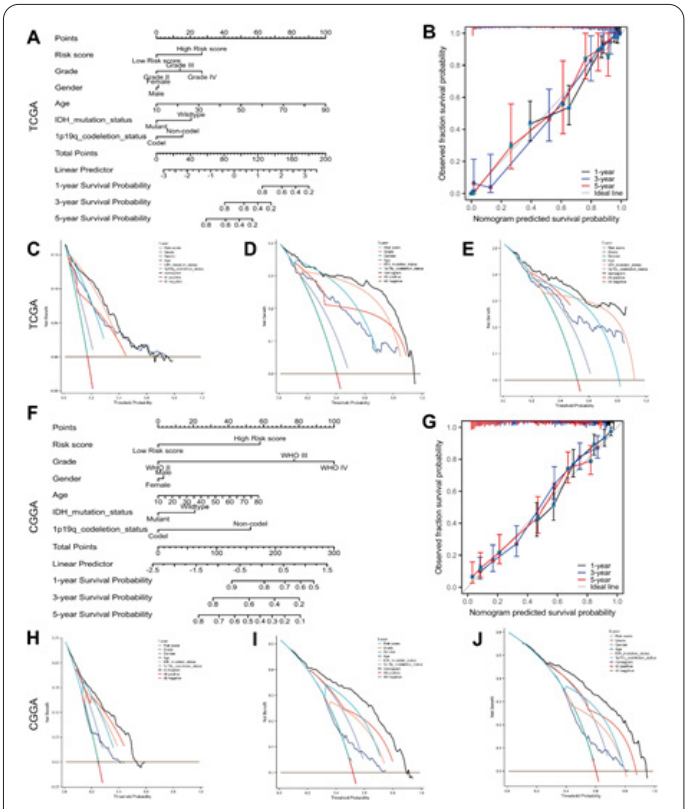

Fig. 4. Construction, evaluation and validation of Nomogram based on risk score. (A): Prognostic nomogram constructed in the TCGA database based on the variables showing significant prognostic significance in the multivariate Cox regression analysis; (B): The calibration graph of the nomogram shows that the nomogram has good value for patient OS prediction. (C-E): Decision curve analysis plot shows the clinical utility of the nomogram for patients at 1, 3 and 5 years than the other separate variables. (F): The constructed nomogram was verified in CGGA database. (G): The calibration graph of the nomogram shows that the nomogram still has good value for patient OS prediction. (H-J): Decision curve analysis plot shows the clinical utility of the nomogram for patients at 1, 3 and 5 years than the other separate variables. TCGA: The Cancer Genome Atlas. OS: Overall survival. CGGA: The Chinese Glioma Genome Atlas.

at high risk showed significant enrichment in Focal adhesion, Cell cycle mitosis, cytokine signaling in the immune system, extracellular matrix organization, and NABA matrisome. In Figure 5K and Table 3, it was observed that the low-risk group had a higher presence of the neuronal system, transmission through chemical synapses, interaction with neuroactive ligand receptors, and transmission of neurotransmitter receptors and postsynaptic signals.

3.7. The correlation between the risk score and immune cells

Figure 6 illustrates the association between the risk score formulated for 9 crucial genes and the presence of immune cells infiltrating the tumor. The high-risk score group exhibited increased stromal, immune, and estimate scores (Figure 6A-C). According to the TCGA database, the risk score showed a strong correlation with 20 out of 24 immune cells (Figure 6D). Figure 6F, and 6H shows that the CGGA database was associated with 21 immune cells, while the Rembrandt database was associated with 20 immune cells in relation to the risk score. The comprehensive training and validation cohort analysis yielded findings indicating that individuals in the high-risk score

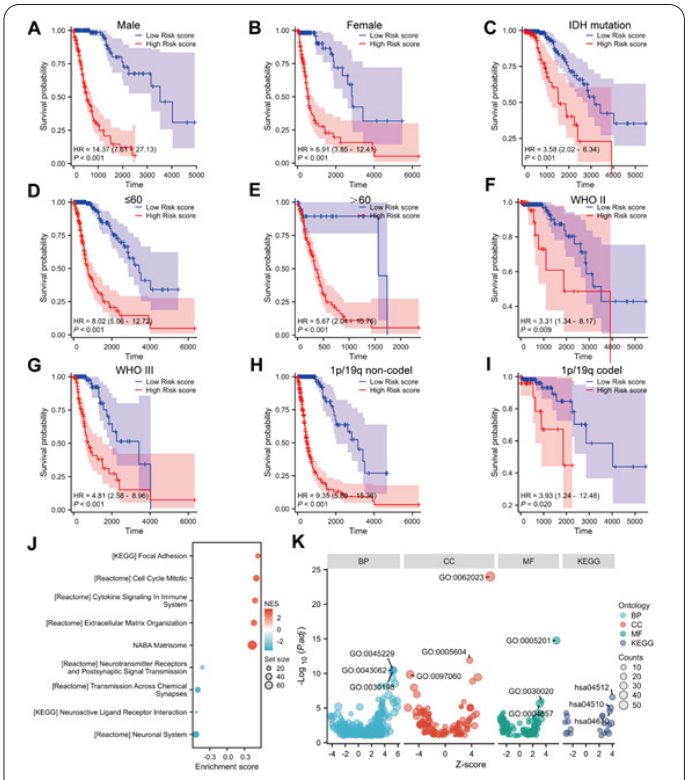

Fig. 5. The relationship between the risk score and the clinical features. Kaplan-Meier survival analysis: (A): Male. (B): Female. (C): IDH mutation. (D): ≥60 years old. (E): <60 years old. (F): WHOII. (G): WHO III. (H):1p/19q non-codel. (I):1p/19q codel. (J): GSEA enrichment analysis of differential genes between high-and low-risk score groups. (K): GO and KEGG analysis of the differential genes between the high-and low-risk score groups. GSEA: Gene Set Enrichment Analysis. GO: Gene Ontology. KEGG: Kyoto Encyclopedia of Genes and Genomes.

category exhibited reduced levels of B cell, Tcm, Tem, Tfh, Tgd, and NK56 bright cell while showing elevated expression of NK cell, macrophages, neutrophils, and Th2 cell. Figures 6E, 6G, and 6I displayed the arrangement of 24 immune cells in the different groups within the training and validation cohorts, as depicted by the radar plots.

The heatmap of co-expression reveals a significant correlation between the risk score and most of the immune checkpoints. In the TCGA and CGGA databases, CD274, CTLA4, PDCD1, LAG3, and CD276 exhibited a favorable association with the risk score, as depicted in Figure 7A-D. In the Rembrandt database, there was a positive correlation between the risk score and CXCL11, CXCL10,

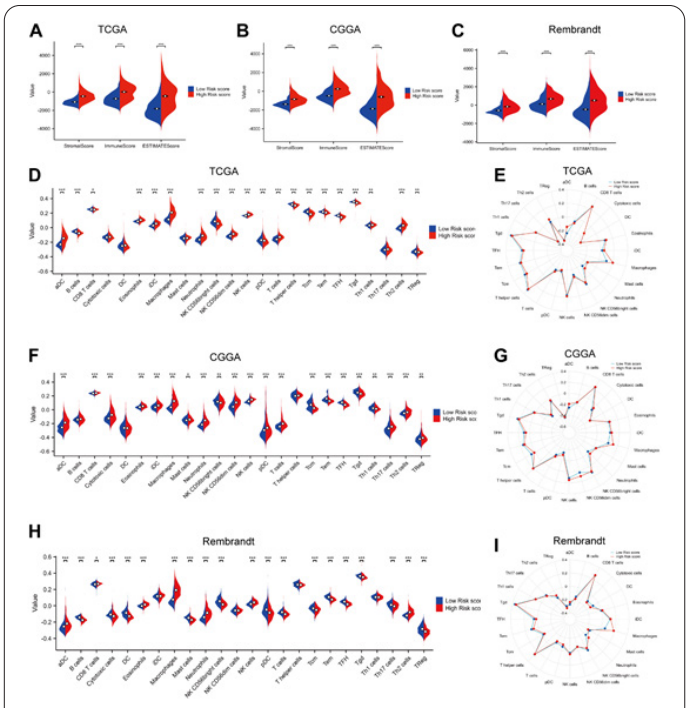

Fig. 6. A significant relationship between risk score and tumor immune cell infiltration. (A-C): The immune infiltration Estimate algorithm showed that the stromal, immune, and estimate scores of patients in the high-risk score group in TCGA, CGGA and Rembrandt databases were higher than those of the low-risk group. (D): The immune infiltration ssGSEA algorithm shows the differential expression of the 24 immune infiltration cells in the TCGA database in the high-low-risk groups. (E): Radar plot showing the expression distribution of 24 immune infiltrating cells in the TCGA database in high- low-risk groups. (F): Differential expression of 24 immune infiltrating cells in the high-and low-risk groups. (G): Radar plot showing the expression distribution of 24 immune infiltrating cells in CGGA database in the high-and low-risk groups. (H): Differential expression of the 24 immune-infiltrating cells in the Rembrandt database in the high- low-risk groups. (I): Radar diagram showing the expression distribution of 24 immune infiltrating cells in Rembrandt database in the high- low-risk groups. TCGA: The Cancer Genome Atlas. CGGA: The Chinese Glioma Genome Atlas. ssGSEA: Single-sample Gene Set Enrichment Analysis.

Table 2. GO and KEGG enrichment analysis of DGEs between high and low groups of risk score in the Rembrandt database.

Ontology	ID	Description	p-value	p.adjust	z-score
BP	GO:0030198	extracellular matrix organization	1.87e-14	3.46e-11	5.24
BP	GO:0043062	extracellular structure organization	2.07e-14	3.46e-11	5.24
BP	GO:0045229	external encapsulating structure organization	2.51e-14	3.46e-11	5.24
CC	GO:0062023	collagen-containing extracellular matrix	2.19e-27	9.79e-25	6.8279
CC	GO:0005604	basement membrane	5.65e-15	1.26e-12	3.7097
CC	GO:0097060	synaptic membrane	9.4e-13	1.4e-10	-5.3333
MF	GO:0005201	extracellular matrix structural constituent	2.54e-18	1.76e-15	5.5678
MF	GO:0030020	extracellular matrix structural constituent conferring tensile strength	2.82e-09	9.74e-07	3.3166
MF	GO:0004857	enzyme inhibitor activity	1.02e-08	2.35e-06	3.0533
KEGG	hsa04512	ECM-receptor interaction	9.52e-10	2.61e-07	4
KEGG	hsa04510	Focal adhesion	7.15e-08	9.79e-06	3.7097
KEGG	hsa04610	Complement and coagulation cascades	3.15e-07	2.87e-05	3.6056

DEGs: differential expression genes. GO: Gene Ontology. KEGG: Kyoto Encyclopedia of Genes and Genomes. BP: Biological Process. CC: Cellular Component. MF: Molecular Function.

Table 3. GSEA enrichment analysis of DGEs between high and low groups of risk score in the Rembrandt database.

ID	P.adjust	NES
matrisome	6.10487e-07	2.980223948
cell cycle mitotic	6.07551e-05	2.748959235
cytokine signaling in immune system	0.000396947	2.481442202
extracellular matrix organization	0.000398244	2.514589626
focal adhesion	0.000398244	2.482655469
neuronal system	3.40929e-07	3.540004944
transmission across chemical synapses	6.07551e-05	3.049618319
neuroactive ligand-receptor interaction	0.001483216	2.318012295
neurotransmitter receptors and postsynaptic signal transmission	0.007069078	2.085540992

GSEA: Gene Set Enrichment Analysis. DEGs: differential expression genes.

CCL2, and CXCR7, while CCL19 and CXCR3 showed a negative correlation with the risk score (Figure 7E-F). To summarize, the risk score can indicate the level of tumor immune cell infiltration and the immunosuppressive state within the tumor immune microenvironment.

4. Discussion

Glioma genesis is an extremely complex process that involves the functional alteration of different genes [7]. Thus, Models that utilize a combination of correlated genes have been proven to be more accurate in forecasting the prognosis of glioma patients compared to models that depend on a solitary gene [8]. As a result, we have created a novel risk assessment score utilizing 9 crucial genes without any particular categorization. The glioma patients were categorized into high- and low-risk score groups by this model, and it was observed that the former had a considerably reduced OS. Furthermore, we developed a nomogram that integrates a risk score and clinical characteristics to forecast the OS of patients at 1, 3, and 5 years. The calibration curve, which relied on the TCGA and CGGA databases, demonstrated a remarkable similarity between the predicted and observed values, affirming the exceptional precision of the nomogram. The improved nomogram can be utilized to forecast survival status and provide personalized treatment, surpassing the initial clinical characteristics.

Extensive research has been conducted on insulin-like growth factor binding protein 2 (IGFBP2) as a potential biomarker for the detection of cancer and as a target for immunotherapy [9]. High-grade gliomas exhibit an elevated level of IGFBP2 expression, whereas IDH mutant gliomas display a reduced level of expression. Studies have demonstrated that IGFBP2 promotes the formation of the glioma cell network by increasing the concentrations of CD144 and MMP2 [10]. IGFBP2 activating integrin α5 and β1/ERK has been discovered to enhance the malignant characteristic of glioma cells (10). Furthermore, IGFBP2 stimulates the EGFR/STAT3 pathway, promoting the nuclear accumulation of EGFR and driving tumor growth [11]. The immunosuppressive actions also involve IGFBP2. IGFBP2's modulation of invasion and progression-associated gene CD24 contributes to increased invasiveness of GBM cells. PPP2R2C, which is a regulatory subunit of PP2A, has been suggested as a possible gene that suppresses tumors and is observed to be decreased in glioma cells [12]. Increased levels of PPP2R2C expression have been found to impede tumor growth by suppressing

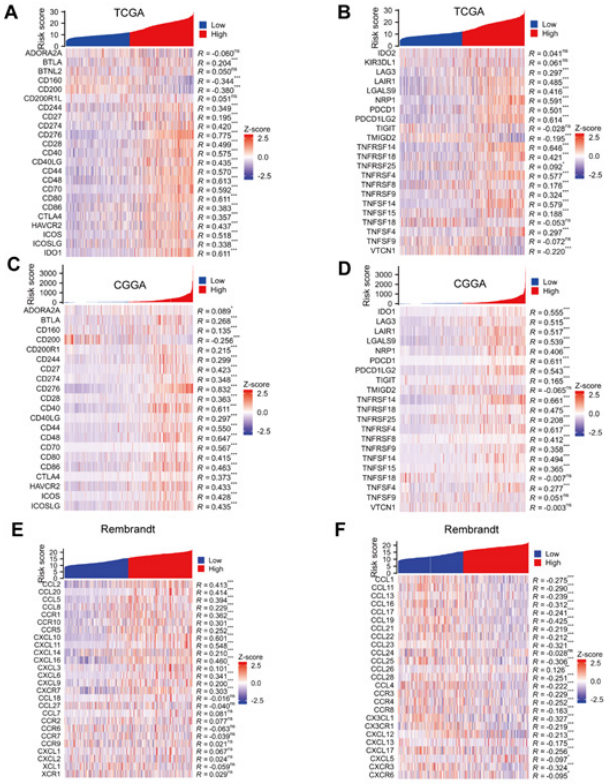

Fig. 7. Correlation analysis of risk score in common immune checkpoints. (A-B): Heatmap of the correlation between risk score and immune checkpoints in the TCGA database. (C-D): Heatmap of the correlation between risk score and immune checkpoints in CGGA database. (E-F): Heatmap of the correlation between risk score and chemokines in Rembrandt database. TCGA: The Cancer Genome Atlas database. CGGA: The Chinese Glioma Genome Atlas database. ns: no signification, *:P<0.05, **:P<0.01, ***:P<0.001.

the activity of the S6K enzyme in the mechanistic target of rapamycin (mTOR) pathway. Studies have shown that blocking TGIF1, a homeobox 1 factor that induces transforming growth factor β, can greatly decrease the growth and infiltration of glioma cells [13]. The utilization of LGALS3BP, a protein that binds to Galectin 3, shows promise as a biomarker for the early identification of glioma and enhanced patient survival rates [14]. The PBK, which is also referred to as the TOPK or T-LAK cell-originated protein kinase, plays a role in signal transduction pathways that control cell cycle and proliferation [15]. It is involved in the development of tumors and the spread of cancer. Research has indicated that the expression of PBK is elevated

in glioblastoma, which is linked to a lower rate of survival [16]. Our results were consistent with previous studies, which confirmed the importance of our selected IGFBP2, PPP2R2C, TGIF1, LGALS3BP and PBK in glioma.

Glioma patients with reduced levels of DNAJA4 have a poorer prognosis. This implies that DNAJA4 is a protective factor in glioma patients. However, our findings indicated that individuals with a decreased expression of DNAJA4 experienced a longer OS. NRXN3, also known as Neurexin 3, is an adhesion molecule found in the presynaptic region [17]. The presence of this protein is linked to the emergence of neuropsychiatric disorders and malignancies. The reduced expression of this protein in gliomas impedes the growth, migration, and infiltration of glioma cells [18]. These findings align with the outcomes of our multivariate Cox regression analysis. However, the significant upregulation of NRXN3 observed in our analysis of patient survival suggests that NRXN3 may function as a tumor suppressor gene, leading to increased OS. Hence, further investigation is necessary to ascertain the precise function of DNAJA4 and NRXN3 in glioma.

Our study shows that ENAH expression was elevated in gliomas, and patients with high expression live longer. SPHKAP expression was reduced in gliomas, and patients with low expression have shorter survival. The ENAH is a member of the Ena/VASP family and is associated with numerous processes that entail alterations to the cytoskeleton and cell polarity [19]. Its carcinogenic role has been investigated in gastric, breast, and esophageal cancers, yet its role in glioma is still unclear. SPHKAP may serve as a potential intersection between cAMP and sphingosine signaling, playing a regulatory function in SPHK1. The extent to which it affects glioma remains to be determined and must be studied further.

The examination of the tumor microenvironment (TME) has uncovered that the presence of immune cells plays a vital part in the advancement of tumors and has the potential to impact the prognosis of individuals with cancer [20]. Tumor cells frequently utilize immune checkpoints to avoid immune detection, thereby maintaining tumor immune tolerance [21]. The findings from GSEA indicated a heightened activation of pathways related to immune response in the group with high-risk scores, which inspired us to investigate the correlation between the risk score and the presence of immune cells within the tumor. The findings revealed that the levels of expression of most immune infiltrating cells varied significantly between the two groups of patients. B cells and T cells were found to be lower in the high-risk score group, while other non-lymphocytes exhibited an increasing pattern. Moreover, there was a positive correlation between the group with a high-risk score and the majority of immune checkpoints, indicating that the risk score has the ability to indicate the level of tumor immunosuppression. The Rembrandt database revealed a connection, whether it be positive or negative, between the risk score and chemokines, highlighting the intricate nature of glioma. Further exploration is necessary to acquire a more profound understanding of the function of immune cells that infiltrate tumors and the process of evading the immune system.

5. Conclusions

To summarize, this study has performed an extensive bioinformatics examination of a risk score that exhibits excellent predictive efficacy for individuals with glioma. By merging information from various sources, the constraints of a limited sample size can be overcome, allowing for the utilization of existing data for additional analysis. The investigation has discovered multiple possible targets for further research on the molecular processes and predictive indicators of glioma. However, additional molecular biology experiments and the validation of a larger number of clinical samples are necessary to ascertain the role of the crucial genes in individuals with varying subtypes of glioma and the clinical significance of the risk score.

Data availability statement

The availability of data in this research involved downloading from various databases including GEO, TCGA, CGGA, Rembrandt, and GTEx. Furthermore, we employed various online databases including TISCH2, UALCAN, HPA, and cBioPortal to validate our findings.

Conflict of Interests

The author has no conflicts with any step of the article preparation.

Consent for publications

The author read and approved the final manuscript for publication.

Ethics approval and consent to participate

No human or animals were used in the present research.

Informed Consent

The authors declare not used any patients in this research.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' contributions

Guangzhao Li, Xiaowang Niu and Zhong Wang designed the study and performed the experiments, Xiang Li and Bin Lin collected the data, Xiang Li, Bin Lin and Fei Yang analyzed the data, Guangzhao Li, Xiaowang Niu and Zhong Wang prepared the manuscript. All authors read and approved the final manuscript.

Funding

Non.

References

1. Jiang X, Zhou X, Zhang L, Chen G, Li S, Cao Y (2022) Long-stranded non-coding RNA HCG11 regulates glioma cell proliferation, apoptosis and drug resistance via the sponge MicroRNA-144COX-2 axis. Cell Mol Biol 67:62-67. doi: 10.14715/cmb/2021.67.9
2. Jin R, Xu Y, Jiang P (2022) Inhibitory Effect of Nano-targeted Micelle Administration Combined with in Vitro Radiotherapy on Glioma Based on Nuclear Magnetic Resonance Technology. Cell Mol Biol 68:171-176. doi: 10.14715/cmb/2022.68.7.28
3. Liu ZY, Lan T, Tang F, He YZ, Liu JS, Yang JZ et al (2023) ZDHHC15 promotes glioma malignancy and acts as a novel prognostic biomarker for patients with glioma. Bmc Cancer 23:420. doi: 10.1186/s12885-023-10883-6
4. Ming Y, Luo C, Ji B, Cheng J (2023) ARPC5 acts as a potential

prognostic biomarker that is associated with cell proliferation, migration and immune infiltrate in gliomas. Bmc Cancer 23:937. doi: 10.1186/s12885-023-11433-w

5. Yan T, Tian D, Chen J, Tan Y, Cheng Y, Ye L et al (2021) FCGBP Is a Prognostic Biomarker and Associated With Immune Infiltration in Glioma. Front Oncol 11:769033. doi: 10.3389/fonc.2021.769033
6. Ye Z, Zhang S, Cai J, Ye L, Gao L, Wang Y et al (2022) Development and validation of cuproptosis-associated prognostic signatures in WHO 2/3 glioma. Front Oncol 12:967159. doi: 10.3389/fonc.2022.967159
7. Zhang R, Jiang W, Liu Z, Hou P, Zhang S (2022) MiR-218 Targeted Regulation of Robol Expression Regulates Proliferation, Invasion and Migration of Glioma Cells. Cell Mol Biol 68:202-206. doi: 10.14715/cmb/2022.68.5.27
8. Dong CY, Cui J, Li DH, Li Q, Hong XY (2018) HOXA10-AS: A novel oncogenic long non-coding RNA in glioma. Oncol Rep 40:2573-2583. doi: 10.3892/or.2018.6662
9. Lindstrom MS (2019) Expanding the scope of candidate prognostic marker IGFBP2 in glioblastoma. Bioscience Rep 39:BSR20190770. doi: 10.1042/BSR20190770
10. Liu Y, Li F, Yang YT, Xu XD, Chen JS, Chen TL et al (2019) IGFBP2 promotes vasculogenic mimicry formation via regulating CD144 and MMP2 expression in glioma. Oncogene 38:1815-1831. doi: 10.1038/s41388-018-0525-4
11. Chua CY, Liu Y, Granberg KJ, Hu L, Haapasalo H, Annala MJ et al (2016) IGFBP2 potentiates nuclear EGFR-STAT3 signaling. Oncogene 35:738-747. doi: 10.1038/onc.2015.131
12. Fan YL, Chen L, Wang J, Yao Q, Wan JQ (2013) Over expression of PPP2R2C inhibits human glioma cells growth through the suppression of mTOR pathway. Febs Lett 587:3892-3897. doi: 10.1016/j.febslet.2013.09.029
13. Wang B, Ma Q, Wang X, Guo K, Liu Z, Li G (2022) TGIF1 overexpression promotes glioma progression and worsens patient prognosis. Cancer Med-Us 11:5113-5128. doi: 10.1002/cam4.4822
14. Rana R, Chauhan K, Gautam P, Kulkarni M, Banarjee R, Chugh P et al (2021) Plasma-Derived Extracellular Vesicles Reveal Galectin-3 Binding Protein as Potential Biomarker for Early Detection of Glioma. Front Oncol 11:778754. doi: 10.3389/fonc.2021.778754
15. Liu Y, Liu H, Cao H, Song B, Zhang W, Zhang W (2015) PBK/TOPK mediates promyelocyte proliferation via Nrf2-regulated cell cycle progression and apoptosis. Oncol Rep 34:3288-3296. doi: 10.3892/or.2015.4308
16. Dong C, Fan W, Fang S (2020) PBK as a Potential Biomarker Associated with Prognosis of Glioblastoma. J Mol Neurosci 70:56-64. doi: 10.1007/s12031-019-01400-1
17. Zheng JJ, Li WX, Liu JQ, Guo YC, Wang Q, Li GH et al (2018) Low expression of aging-related NRXN3 is associated with Alzheimer disease: A systematic review and meta-analysis. Medicine 97:e11343. doi: 10.1097/MD.0000000000011343
18. Sun HT, Cheng SX, Tu Y, Li XH, Zhang S (2013) FoxQ1 promotes glioma cells proliferation and migration by regulating NRXN3 expression. Plos One 8:e55693. doi: 10.1371/journal.pone.0055693
19. Chen D, Xu L, Li X, Chu Y, Jiang M, Xu B et al (2019) Author Correction: Enah overexpression is correlated with poor survival and aggressive phenotype in gastric cancer. Cell Death Dis 10:366. doi: 10.1038/s41419-019-1573-6
20. Bikfalvi A, Da CC, Avril T, Barnier JV, Bauchet L, Brisson L et al (2023) Challenges in glioblastoma research: focus on the tumor microenvironment. Trends Cancer 9:9-27. doi: 10.1016/j.trecan.2022.09.005
21. Downs-Canner SM, Meier J, Vincent BG, Serody JS (2022) B Cell Function in the Tumor Microenvironment. Annu Rev Immunol 40:169-193. doi: 10.1146/annurev-immunol-101220-015603