E-ISSN 2250-0944
ISSN 2250-1150
doi: 10.31838/ijprt/16.02. 483

Research Article

Decoding Glioblastoma: An Integrated Prognostic Gene Signature and Molecular Landscape Analysis

Shruti Nikam[1], Vibha Patil[2*], Vaibhavi Patil[3], Sharvari Chavan[4]
[1]Senior Analyst at Illinois Institute of Technology.
[2*]Assistant Professor at ALARD School of Pharmacy, Pune.
[3]Research Scholar at Poona College of Pharmacy, Pune.
[4]Assistant Professor at Abhinav Education Society's College of Pharmacy.
Corresponding Author: Vibha Shivajirao Patil
Assistant Professor at ALARD School of Pharmacy, Pune.
Email: vibhap1414@gmail.com
Received: 03.07.26, Revised: 05.08.26, Accepted: 07.09.26

ABSTRACT
Background: Glioblastoma (GBM) exhibits extensive molecular heterogeneity and poor clinical outcomes, necessitating reliable molecular prognostic markers. This study aimed to develop and validate a transcriptome-based prognostic signature for GBM.
Methods: RNA-sequencing and clinical data from The Cancer Genome Atlas (TCGA) were analyzed using differential expression, Kaplan–Meier survival, univariate and multivariable Cox regression, and LASSO Cox regression. Functional enrichment, gene set enrichment analysis (GSEA), protein–protein interaction (PPI), and immune-infiltration analyses were performed. The Chinese Glioma Genome Atlas (CGGA) cohort was used for external validation.
Results: Differential expression analysis identified 15,098 DEGs, including 8,922 upregulated and 6,176 downregulated genes. After filtering, 41,864 genes from 283 patients were analyzed for survival. Univariate Cox analysis identified 76 significant genes, while Kaplan–Meier analysis identified 35 prognostically significant genes. Multivariable Cox regression demonstrated a C-index of 0.7101 (SE 0.0205). LASSO Cox regression selected an 18-gene signature at λmin = 0.055606. Using a risk-score cutoff of 1.135096, patients were classified into 142 high-risk and 141 low-risk groups. The model achieved AUCs of 0.7711, 0.8560, and 0.8864 for 1-, 3-, and 5-year overall survival, respectively. A nomogram incorporating age and risk score was developed, with coefficients of 0.027826 and 1.401713, respectively. Functional and immune analyses demonstrated distinct biological and immune characteristics between risk groups.
Conclusion: The 18-gene signature demonstrated promising prognostic discrimination and provides a potential framework for GBM risk stratification and biomarker discovery. Further experimental and prospective clinical validation is required.

Keywords: Glioblastoma, Prognostic Signature, Transcriptomic Analysis, Bioinformatics, Differentially Expressed Genes, Gene Set Enrichment Analysis (GSEA).

INTRODUCTION

Glioblastoma

With a median survival rate of roughly 20 months, glioblastoma (GBM) is the most common primary malignant brain tumour in adults[(1)]. Oncolytic virus (OV) therapy has surfaced as a uniquely mechanistic treatment approach aimed at addressing the shortcomings of traditional immunotherapies in glioblastoma[(2)]. Moreover, the tightly regulated blood-brain barrier (BBB) and the fluctuating blood-brain tumour barrier (BBTB), which together restrict drug entry into glioblastoma tissue, considerably reduce therapy efficacy[(3)]. Observational studies have yielded contradictory results, indicating both enhanced survival with fluoxetine and worse outcomes linked to antidepressant usage in glioblastoma patients[(4)]. These results highlight the necessity for the creation of more potent therapeutic approaches that can surmount the immunosuppressive tumour microenvironment and enhance clinical results in glioblastoma patients[(5)].

Epidemiology and Disease Burden

Despite developments in multimodal treatment options, population-based studies have shown only small gains in glioblastoma survival over recent decades, emphasising the disease's ongoing burden[(6)]. In both newly diagnosed and recurrent glioblastoma, the blood-brain barrier is a major obstacle that keeps therapeutic doses of the majority of systemic medications from getting to the tumour tissue[(7)]. Furthermore, glioblastoma's infiltrative growth

pattern and strong molecular heterogeneity contribute to therapy resistance, making long-term disease control particularly problematic[4]. Furthermore, the combination of radiotherapy and temozolomide has dramatically enhanced the five-year survival of patients with glioblastoma in comparison to radiotherapy alone[8]. These characteristics limit the ability of standard clinicopathological indicators to reflect the genomic variety underpinning prognosis, motivating the creation of reproducible molecular risk-stratification frameworks[9]. Overall, the high recurrence rate and the lack of standardised care techniques continue to contribute significantly to the disease burden of glioblastoma[10].

Current Standard Treatment and Its Limitations

For glioblastoma, the Stupp regimen is currently the accepted treatment. It entails maximal safe surgical resection, concurrent temozolomide radiation, and adjuvant temozolomide chemotherapy[11]. The maximum safe surgical resection followed by concurrent radiation and temozolomide chemotherapy is the current standard treatment for glioblastoma; nonetheless, the five-year survival rate is still only about 5% despite this multimodal strategy[4]. The development of a deeply immunorepressive niche marked by poor antigen presentation, an increase in immunosuppressive myeloid populations, and increasing T-cell depletion is a major obstacle to therapeutic success[5]. In older glioblastoma patients, performance status, treatment intensity, and MGMT promoter methylation status are significant predictors of survival, underscoring the need for integrated histo-molecular classification and individualised treatment planning[12]. Thus, merging improved imaging indicators with molecularly targeted treatment methods may improve clinical decision-making and optimise patient management in glioblastoma[13]. Overall, temozolomide resistance underlines the critical need for innovative combination therapies capable of engaging several regulated and immunogenic cell death mechanisms to improve long-term therapeutic results in glioblastoma[14].

Prognostic Biomarkers and Survival Prediction

Recent integrated bioinformatics investigations have revealed strong prognostic biomarkers and risk models that improve survival prediction and enable personalised therapy methods in glioblastoma[15]. Because of their high expression in tumour tissues and body fluids, particularly cerebrospinal fluid and blood, cathepsins are being studied as non-invasive biomarkers for diagnosis, prognosis, and therapy monitoring in glioblastoma[16]. A prognostic risk model that demonstrated excellent predictive accuracy for intermediate to prolonged survival in glioblastoma patients was developed thanks to the discovery of NGFR and L1CAM[15]. In addition, integrating radiomic, pathomic, genomic, transcriptomic, and proteomic data using a multimodal machine learning framework offered more accurate risk categorisation and a thorough characterisation of molecular heterogeneity in glioblastoma[17]. Furthermore, machine learning models trained on preoperative resting-state functional MRI connectivity showed that large-scale network changes can be used as reliable, non-invasive indicators to forecast survival without development while it support personalised clinical decision-making in glioblastoma[18]. These multimodal methodologies have significantly increased prognosis accuracy and allowed for a more thorough assessment of glioblastoma survival than standard single-modality methods[19]. Furthermore, combining radiomic characteristics with transcriptome and single-cell tumour microenvironment characterisation provides a noninvasive technique for capturing glioblastoma heterogeneity, improving prognostic classification, and uncovering personalised treatment possibilities[20]. Furthermore, a manganese metabolism-related gene profile displayed independent predictive value and was strongly connected with immune cell infiltration, indicating its potential relevance for risk stratification and personalised treatment decision-making in glioblastoma[21]. Integrating the pathomics risk score with clinical characteristics, such as Karnofsky Performance Status and extent of resection, enabled the development of a very accurate prognostic nomogram for individualised survival prediction in patients with IDH-wildtype glioma[22]. Overall, these findings show that integrating genetic indicators with advanced computational and clinicopathological models can significantly enhance prognostic accuracy and simplify personalised survival prediction in glioblastoma[23].

Survival analysis methodology

Glioblastoma survival analysis has evolved from conventional regression and hazard-based

models to sophisticated machine learning methods that can analyse high-dimensional data and capture complex nonlinear relationships for more precise result prediction[18]. These methods allow for the integration of various clinical, genetic, and imaging data to estimate progression-free survival (PFS) and overall survival (OS), facilitating precise prognostic classification and customised therapy management for glioblastoma patients [24]. Additionally, the integration of single-cell transcriptomics, bulk RNA sequencing, and machine learning has enabled the identification of novel prognostic biomarkers and risk models. This has enhanced prognostic assessment and facilitated the development of customised glioblastoma treatment regimens[15]. Overall, the evidence now available indicates that fluorescence-guided resection mostly improves the safety and extent of tumour removal; however, its independent contribution to long-term survival in high-grade glioma is still unknown and necessitates additional high-quality randomised research[25].

METHODOLOGY

TCGA

Glioblastoma (GBM) sufferers' RNA sequencing (RNA-seq) counts information and associated medical records were acquired from The Cancer Genome Atlas (TCGA) using the TCGAbiolinks tool in R. Transcriptomic count data was obtained by accessing the Genomic Statistics Commons (GDC) website using the GDCquery, GDCdownload, and GDCprepare functions. The associated clinical data for further bioinformatics analysis was obtained using the GDCquery_clinic function[26].

Downstream bioinformatics analyses, such as survival analysis, Cox proportional hazards regression, Kaplan-Meier analysis, machine learning-based prognostic model development, and cell-type deconvolution, were performed on the downloaded transcriptome and clinical datasets. Cox regression analysis in standard form was used to identify independent prognostic parameters, and time-depending receiver operating characteristic (ROC) analysis was used to evaluate the prognostic outcome of the constructed models[27].

CGGA

The Chinese Glioma Genome Atlas (CGGA) database provided RNA sequencing (RNA-seq) transcript as well as related clinical information. The downloaded datasets served as an independent validation cohort for further bioinformatics analyses, and patients with complete clinical and survival data were kept[28]. To confirm the biological and prognostic significance of the discovered biomarkers, the downloaded CGGA transcriptomic and clinical datasets were then processed for downstream analyses, such as immune cell infiltration analysis, correlation analysis, Kaplan-Meier survival analysis, and comparative transcriptomic evaluation[29].

Differential Expression Analysis

The processed RNA sequencing count matrix from the TCGA glioblastoma cohort was used to perform differential expression analysis. Annotated gene expression matrices were utilised as input to determine differentially expressed genes (DEGs) after gene expression data was filtered and normalised before statistical Inspection. Differential expression analysis was performed by R program limma. To find genes that differed in expression between the predetermined comparison groups, linear models with empirical Bayes moderation were used. Genes having a rate of false discovery (FDR)-modified p value < 0.05 and $|\log_2$ fold change $(\log_2 FC)| > \log_2(1.5)$ were considered significantly differentially expressed[30].

Kaplan–Meier Survival Analysis

Clinical data from the TCGA glioblastoma cohort, as well as the filtered gene expression matrix, were integrated using matched patient identities. The differentially expressed genes discovered in the preceding analysis were integrated with overall survival data to create a uniform dataset for survival analysis. First, genes that had a strong correlation with overall survival were found using a univariate Cox proportional hazards regression model. Genes that met the predefined significance criteria were classed as risk or protective, based on their hazard ratios. Following Kaplan-Meier survival analysis of the relevant prognostic genes, patients were categorised into high- and low-expression groups based on median gene expression levels. The Kaplan-Meier method was used to create survival curves; log-rank test was used to analyse group differences; and statistically relevant genes were ranked based on how well they predicted outcomes[31].

Cox Proportional Hazards Regression

Using matching patient identities, prepared clinical data and the filtered gene expression

matrix from the TCGA glioblastoma cohort were combined. For Cox proportional hazards regression, genes that showed a significant correlation with overall survival during Kaplan-Meier survival analysis were selected as potential prognostic biomarkers. To begin, a univariate Cox proportional hazards regression analysis was conducted assessing a correlation among overall survival and candidate gene expression. Variables with statistically significant relationships were then incorporated within multivariate Cox proportional hazards regression model for find independent prognostic biomarkers. Each variable was assigned a hazard ratio (HR), 95% confidence intervals (CIs), as well as matching p-value. After a proportional hazards assumption was verified before an interpretation of this final multivariable model, and statistically significant genes were kept as the final prognostic biomarkers[31,32].

LASSO Cox Regression

Genes strongly related with overall survival in the previous Kaplan-Meier and Cox proportional hazards studies were chosen as possible prognostic genes. The LASSO Cox regression analysis is being carried out employing the matched clinical survival data and gene expression patterns based on TCGA glioblastoma cohort. The most helpful prognostic genes were found using R's glmnet utility, which also reduced multicollinearity and model overfitting. The LASSO model included gene candidates chosen by univariate Cox regression, and 10-fold cross-validation has been used to calculate the ideal consequences(λ). Genes with non-zero regression coefficients at optimal λ value had been utilized for the development of prognostic risk model. Total of the products of the gene expression values and the corresponding regression coefficients had been utilised for estimating the risk score for every individual(33,34).

Time-dependent ROC Analysis

Risk scores from the LASSO Cox prognostic model, in addition to the matching total survival duration and survival score of TCGA glioblastoma patients, were employed in receiver operating characteristic (ROC) analysis that was time-sensitive. R's time-dependent receiver operating characteristic (ROC) analysis was used to evaluate the prediction ability of the prognostic signature. To assess the risk model's discriminatory ability to forecast overall survival at one, three, and five years, AUC, or the area under the ROC curve, was computed. Greater AUC scores imply that the constructed risk model has superior prediction accuracy and prognostic performance[35,36].

Nomogram Construction

A prognostic nomogram for the TCGA glioblastoma cohort was produced by combining the independent predictive variables found in the multivariable Cox regression analysis. To give personalised survival prediction, a prognostic nomogram was built in R using the multivariable Cox proportional hazards model. Each independent prognostic factor was given a weighted value based on its regression coefficient, and the cumulative score was used to assess the likelihood of overall survival. The nomogram was constructed using the rms program and acted as a graphical tool for individual prognostic assessment[37,38].

Calibration Curve

Using calibration curves generated by the calibrate() function of the rms package in R, the calibration effectiveness of the prognostic nomogram was evaluated. For internal validation, predicted survival probability were compared with observed survival outcomes at 1- and 3-
year time points using bootstrap resampling (1,000 iterations). It was believed that accurate model calibration and dependable predicted performance were shown by the calibration curve's strong agreement with the 45° reference line[39,40].

Decision Curve Analysis (DCA)

The clinical value a typical predictive nomogram has been assessed by Decision Curve Analysis (DCA). The overall advantage had been computed over a variety of clinically relevant threshold probabilities in order to compare the prediction model to a standard strategy for either medicating every individual or no treatment of them. A ggDCA package in R was used to build decision curves, and a higher net benefit across threshold probabilities indicated that the prognostic model could be more clinically helpful[41,42].

GO & KEGG Enrichment Analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis was performed on the discovered prognostic genes using the R package clusterProfiler to investigate their biological roles and signaling pathways. GO enrichment

was done in groupings: biological process (BP), cellular component (CC), and molecular function (MF). KEGG pathway enrichment assessment was used to find biological pathways that were considerably enriched. Evaluating many hypotheses provided applying the Benjamini-Hochberg method. Pathways have been taken into account statistically significant if their modified P value fell below 0.05. An enriched GO keywords and KEGG pathways were visualised using dot and bar plots generated in R[43].

Gene Set Enrichment Analysis (GSEA)

A ranking gene list Were linked with considerably enhanced biological pathways using R's clusterProfiler utility. The Hallmark and Gene sets for canonical pathways was included in a Molecular Signatures Database (MSigDB), which was used for enrichment analysis after genes were ranked according to log2 fold-change values. The importance of the enrichment was evaluated using normalised enrichment scores (NES). The significantly enriched pathways were identified by their modified P value < 0.05 and an FDR < 0.25. Dot plots and enrichment plots created with R were used to visualise the highly enriched pathways[44].

STRING Protein–Protein Interaction Network

Using The Finding Engine to Ensure the Acquisition of Interacting Genes/Proteins (STRING) database, the functional links between the differentially expressed genes were investigated. Gene symbols were submitted into an STRING database, with Homo sapiens serving for reference organism. To design the interaction network, a minimum interaction score (combined score) of 0.4 (medium confidence) was used. A network that emerged were analyzed to identify highly connected hub genes by node degree, and the interaction network was exported for additional visualisation and downstream analysis. The STRING interaction network was exported and imported into Cytoscape for topological and visual analysis. Using Cytoscape network analysis tools, hub genes were discovered by node degree, and subnetworks were further studied based on connectedness[45].

Immune Infiltration

The CIBERSORT method had been utilised for determine the relative proportions of 22 immune cell types in each sample. Normalised gene expression matrix was compared to the LM22 immune cell signature matrix as a reference. Immune cell infiltration variations across groups with high and low risk have been evaluated, and immune cell fractions were calculated for each sample. A correlation among prognostic risk ratings as well as invasion of immune cells Spearman's correlation analysis had been utilised to evaluate levels; $p < 0.05$ was deemed statistically significant[46].

R Statistical Computing Environment and RStudio

Key analyses were carried out with the TCGAbiolinks, limma, survival, glmnet, timeROC, rms, ggDCA, clusterProfiler, and CIBERSORT software. R is a free, open-source programming language intended for statistical assessment along with data analysis, offering a reproducible and adaptable framework for implementing computational algorithms. RStudio is an integrated development environment that improves the usability of R by providing tools for script development, project management, data visualisation, package management, and result interpretation, promoting efficient and organised data analysis[47].

RESULTS AND DISCUSSION

TCGA Dataset Characteristics and Quality Assessment

For glioblastoma (GBM), the Cancer Genome Atlas (TCGA) successfully supplied RNA sequencing (RNA-seq) statistical information as well as associated clinical data. The downloaded dataset included 599 clinical patients, 391 RNA sequencing samples, and 60,660 genes annotated using GENCODE version 36 (Table 6).

The clinical characteristics of the TCGA cohort showed that the average patient age was 58.35 years, 59.45 years old is the median age. At time of data collection, 147 of the 599 patients were still alive, while 446 had passed away. The age distribution revealed that the majority of patients were diagnosed between 50 and 75 years old, with the largest frequency occurring in the sixth decade of life (Figure 1A). Gender distribution analysis revealed that male patients outnumbered female patients, with only a small number of samples lacking gender information (Figure 1B). Similarly, the vital status distribution showed that deceased patients outnumbered surviving patients, demonstrating the poor overall prognosis of glioblastoma (Figure 1C).

To create a dataset appropriate for integrated

transcriptome and survival studies, the clinical and RNA sequencing datasets were matched with unique patient identification. Of the 599 clinical patients and 293 RNA samples used for matching, 287 had complete clinical and transcriptomic information, 312 had only clinical information, and 6 had only RNA sequencing data (Table 5). These matched instances were then used for further prognostic analysis.

Quality assessment revealed that the RNA sequencing data had acceptable sequencing depth throughout the study population. The majority of samples had 4-6 × 10^7 sequencing reads, with only a few outliers at low and high depths (Figure 2A). In line with this discovery, the library size boxplot revealed reasonably equal sequencing depth across samples, indicating adequate data quality for downstream transcriptome analysis (Figure 2B).

The expression matrix consists of 60,660 genes evaluated from 391 RNA sequencing samples (Table 2). After $\log_2$ translation, expression values varied from 0 to 23.615, with a mean of 3.695, median of 2.000, and standard deviation of 4.166 (Table 2). The $\log_2$ expression boxplots revealed consistent expression distributions across samples, indicating minimal technical variance after preprocessing (Figure 2C). Similarly, the density plots revealed significantly overlapping expression profiles between samples, showing overall consistency in transcriptome data distribution (Figure 2D). The histogram also revealed that the majority of genes had relatively low expression levels, with fewer genes showing moderate to high expression levels, as is typical of RNA sequencing datasets (Figure 2E).

An examination of gene annotations revealed 60,660 annotated genes representing various transcript types. Protein-coding genes were the most numerous gene category (19,962), followed by long non-coding RNAs (16,901) and processed pseudogenes (10,167). Unprocessed pseudogenes, miscellaneous RNAs, short nuclear RNAs, microRNAs, transfer RNA-related transcripts, and a variety of additional low-abundance gene biotypes were also included (Table 3). Figures 2F and 2G show the distribution of the key gene categories in a bar plot and a pie chart.

The RNA sequencing dataset's integrity was confirmed by the quality control assessment. 57,954 of the 60,660 identified genes showed detectable expression, while 2,706 genes had no counts across all samples. There were no duplicate gene identifiers, sample identifiers, missing expression values, or zero-count samples. Although 1,233 duplicate gene names were discovered, they represented numerous annotations that shared the same gene symbol rather than redundant gene identifiers (Table 4). Overall, these findings demonstrated that the TCGA RNA sequencing dataset was of adequate quality for subsequent differential expression, survival, and functional enrichment analysis.

Table 01: Clinical Summary

Parameter	Value
Total Patients	599
Alive	147
Dead	446
Mean Age	58.35
Median Age	59.44658

Table 02: Expression Summary

Statistic	Value
Genes	60660
Samples	391
Minimum	0
Maximum	23.615
Mean	3.695
Median	2
Standard Deviation	4.166

Table 03: Gene Type Summary

Gene Type	Count	Gene Type	Count
Protein coding	19962	rRNA	47
lncRNA	16901	IG D gene	37
Processed pseudogene	10167	TR V pseudogene	33

Unprocessed pseudogene	2614	Mt tRNA	22
Misc RNA	2212	IG J gene	18
snRNA	1901	pseudogene	18
miRNA	1881	IG C gene	14
TEC	1057	IG C pseudogene	9
snoRNA	943	ribozyme	8
Transcribed unprocessed pseudogene	939	TR C gene	6
Transcribed processed pseudogene	500	sRNA	5
rRNA pseudogene	497	TR D gene	4
IG V pseudogene	187	TR J pseudogene	4
IG V gene	145	IG J pseudogene	3
Transcribed unitary pseudogene	138	Mt rRNA	2
TR V gene	106	Translated processed pseudogene	2
Unitary pseudogene	98	IG pseudogene	1
TR J gene	79	scRNA	1
scaRNA	49	Translated unprocessed pseudogene	1
Polymorphic pseudogene	48	Vault RNA	1

Table 04: Sample QC

Parameter	Value
Total Genes	60660
Expressed Genes	57954
Total Samples	391
Duplicate Gene IDs	0
Duplicate Gene Names	1233
Duplicate Samples	0
Missing Values	0
Zero Count Genes	2706
Zero Count Samples	0

Table 05: Clinical–Rna Matching Summary

Parameter	Value
Clinical Patients	599
RNA Samples	293
Matched Patients	287
Clinical Only	312
RNA Only	6

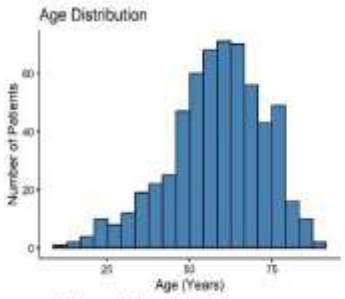


Figure 1A: Age Histogram

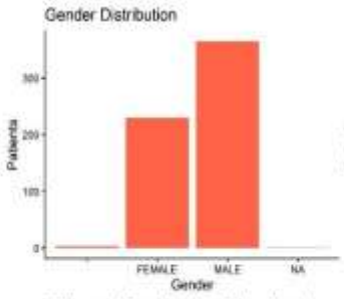


Figure 1B: Gender Distribution

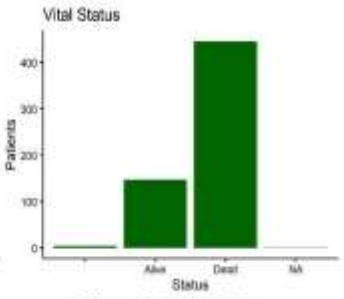


Figure 1C: Vital Status

Table 06: Tcga Dataset Summary

Parameter	Value
Clinical Patients	599
RNA Samples	391
Genes	60660
Gene Annotation Records	60660

Sample Metadata Records	391
Gene Annotation Version	GENCODE v36

Figure 1: Clinical Characteristics of the TCGA Glioblastoma Cohort.

(A) Age Distribution of Patients Included In the TCGA Cohort. (B) Gender Distribution of the TCGA Cohort. (C) Vital Status Distribution Showing the Proportion of Alive and Deceased Patients.

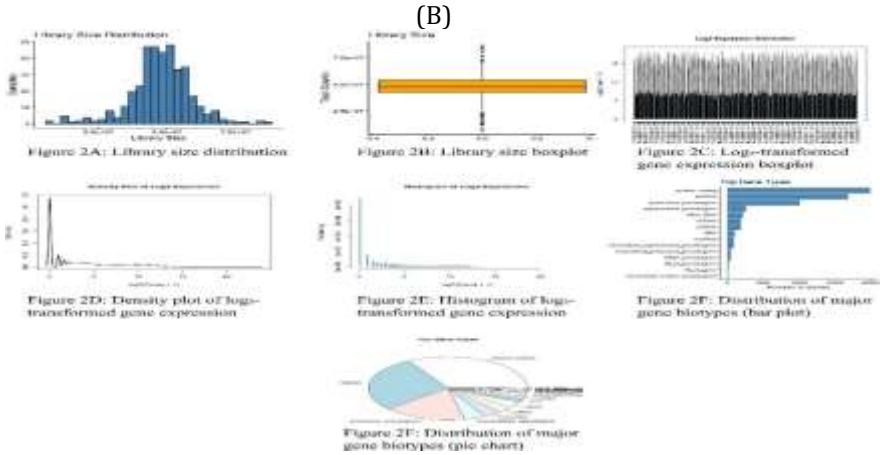


Figure 2: Quality Assessment and Transcriptomic Characteristics of the TCGA RNA Sequencing Dataset

(A) Histogram showing the distribution of sequencing library sizes across TCGA samples. (B) Boxplot illustrating variation in library sizes among samples. (C) Boxplot of log_2-transformed gene expression values across all samples. (D) Density plot showing the distribution of normalized gene expression profiles. (E) Histogram illustrating the overall distribution of log_2-transformed gene expression values. (F) Bar plot showing the distribution of gene biotypes in the TCGA RNA sequencing dataset. (G) Pie chart illustrating the proportion of different gene biotypes.

CGGA Database Download and Quality Assessment

To evaluate the predictive model created using the TCGA cohort, transcriptome relevant clinical information was gathered based on Chinese Glioma Genome Atlas (CGGA). Downloaded clinical dataset included 693 patients with 13 clinical characteristics, and no duplicate patient records were found, verifying the clinical information's integrity (Table 7). The demographic features revealed a mean patient age of 43.28 years (median: 43 years), with 393 men and 290 females, showing a little male predominance in the group (Table 11). The age distribution revealed that the majority of patients were between 35 and 55 years old, with a higher proportion of male patients (Figure 3A-B).

The clinical annotations revealed that most demographic variables were essentially comprehensive, but numerous treatment- and molecular-related factors lacked information. The MGMT promoter methylation status had the highest proportion of missing values (151 cases), followed by 1p/19q codeletion status (70 cases), IDH mutation status (51 cases), radiotherapy status (47 cases), chemotherapy status (46 cases), overall survival (36 cases), and censoring information (30 cases), with only a single missing value for histology, tumour grade, and age (Table 8). Despite the missing annotations, the dataset is still sufficient for external validation analysis.

The raw RNA sequencing read count matrix included 55,523 genes from 693 samples, with no duplicate or missing results. A total of 2,333 genes had no read counts and were termed non-expressed (Table 9). Similarly, the processed FPKM expression matrix had 23,987 genes assessed across 693 samples, with no duplicates, missing values, or zero-expression genes, exhibiting high dataset completeness and consistency (Table 10).

Descriptive analysis of the gene expression matrix revealed that expression values varied from 0 to 24.437, with a mean of 3.144, median of 1, and standard deviation of 3.98, indicating the

typical right-skewed distribution seen in RNA sequencing datasets (Table 12). The log_2-transformed expression boxplot showed a consistent expression distribution across all samples, indicating satisfactory normalisation (Figure 4C). Similarly, the density plot and

expression histogram revealed a comparable global expression profile, with the majority of genes having relatively low expression levels and a smaller number showing high expression (Figure 4D–E).

Quality assessment validated the CGGA dataset's resilience. The expression matrix contains 55,523 genes and 693 samples, with no duplicates, duplicate samples, or missing data. Although 2,333 genes had zero counts, no samples demonstrated complete absence of expression (Table 13). Furthermore, the library size histogram and boxplot showed very equal sequencing depth across samples, with only a few predicted outliers (Figure 4A-B), indicating that the sequencing data is reliable. Overall, the rigorous quality control evaluations revealed that the CGGA dataset was of good quality and suitable for further external validation of the glioblastoma prognostic model.

Table 07: Clinical Summary

Parameter	**Value**
Total Patients	693
Variables	13
Duplicate Samples	0

Table 8: Clinical Missing Values

Variable	**Missing**
Sample ID	0
PRS type	0
Histology	1
Grade	1
Gender	0
Age	1
OS	36
Censor (alive = 0; dead = 1)	30
Radio status (treated = 1; un-Treated = 0)	47
Chemo status (TMZ Treated = 1; un-treated = 0)	46
IDH mutation status	51
1p19q codeletion status	70
MGMTp methylation status	151

Table 9: Read Count Matrix Summary

Parameter	**Value**
Total Genes	55523
Samples	693
Duplicate Genes	0
Missing Values	0
Zero Count Genes	2333

Table 10: FPKM Expression Matrix Summary

Parameter	**Value**
Total Genes	23987
Samples	693
Duplicate Genes	0
Missing Values	0
Zero Expression Genes	0

Table 11: Clinical Characteristics Summary

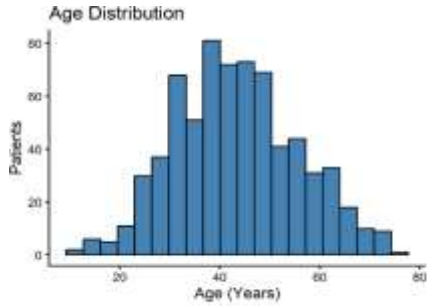
Parameter	**Value**
Total Patients	693
Mean Age	43.28
Median Age	43
Male	393
Female	290

Table 12: Gene Expression Summary

Statistic	Value
Genes	55523
Samples	693
Minimum	0
Maximum	24.437
Mean	3.144
Median	1
Standard Deviation	3.98

Table 13: Sample Quality Control Metrics

Parameter	Value
Total Genes	55523
Total Samples	693
Duplicate Genes	0
Duplicate Samples	0
Missing Values	0
Zero Count Genes	2333
Zero Count Samples	0



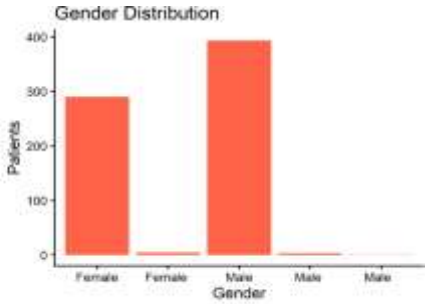


Figure 3A: Age Distribution Figure 3B: Gender Distribution

Figure 3. Clinical Characteristics of the CGGA Glioblastoma Cohort.

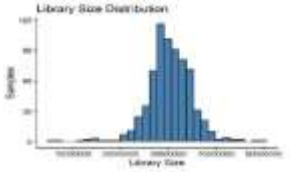


Figure 4A: Library size distribution

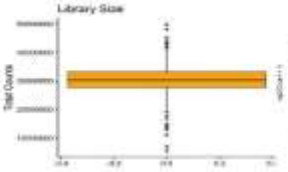


Figure 4B: Library size boxplot

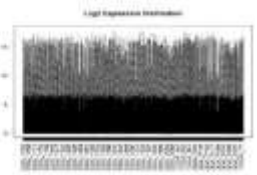


Figure 4C: Log2-transformed gene expression boxplot

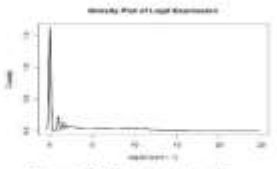


Figure 4D: Density plot of log2-transformed gene expression

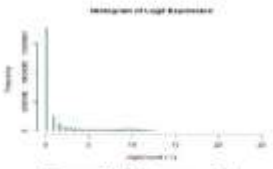


Figure 4E: Hfistogram of log2-transformed gene expression

Age Distribution of Patients Included in the CGGA Cohort. (B) Gender Distribution of the CGGA Cohort.

Figure 4. Quality Assessment and Transcriptomic Characteristics of the CGGA RNA Sequencing Dataset.

(A) Histogram showing the distribution of sequencing library sizes across CGGA samples. (B) Boxplot illustrating variation in library sizes among samples. (C) Boxplot of log_2-transformed gene expression values across all samples. (D) Density plot showing the distribution of normalized gene expression profiles. (E) Histogram illustrating the overall distribution of log_2-transformed gene expression values.

Differential Expression Analysis of the TCGA Glioblastoma Cohort Using limma

Using limma program, Differential gene expression analysis has been employed to find genes that substantial expression changes

between glioblastoma tumour and normal brain tissues. Study included 372 cancer samples and 5 normal samples, which served like a foundation to find differentially expressed genes (DEGs) (Table 14). To increase data quality and reduce background noise, lowly expressed genes were eliminated before statistical analysis began. There were initially 60,660 genes accessible, of which 41,864 remained after filtering, and 18,796 were removed from further research (Table 15). Consequently, 41,864 genes were subjected to differential expression analysis using the filtered expression matrix (Table 16).

The volcano map offered a summary of the differential expression results, displaying the distribution of highly elevated and downregulated genes in tumour and normal tissues (Figure 5A). Highlighted genes were those that satisfied the designated significance criteria ($|\log_2$ fold change$| > 1$ and adjusted $P < 0.05$). Red represented upregulated genes, blue represented downregulated genes, and grey represented non-significant genes. Numerous genes exhibited notable differential expression, indicating serious transcriptome alterations linked to glioblastoma.

Overall, 15,098 genes were identified as highly differently expressed, with 8,922 upregulated and 6,176 downregulated (Table 17). The differential expression statistics showed a maximum fold change of 9.4756, a minimum fold change of -8.5592, a mean fold change of 0.2962, and a median fold change of 1.449. Table 19 shows that the most differentially expressed genes had a minimum adjusted P-value of 7.28×10^{-45}, indicating exceptionally great statistical significance.

To delve deeper into the expression characteristics found within most relevant genes, a heatmap was created utilising the top 50 differentially expressed genes in all tumour as well as normal data. The hierarchical clustering analysis revealed a strong distinction between tumour and normal tissues, indicating unique gene expression profiles (Figure 5B). The identified genes' ability to differentiate glioma samples from normal brain tissues was shown by the clustered expression patterns, which also verified the differential expression analysis.

Finally, the discovered differentially expressed genes were exported for further functional enrichment, pathway analysis, gene set enrichment analysis (GSEA), and prognostic modelling. The exported datasets contained 14,957 differentially expressed genes, with 8,787 upregulated genes, 6,170 downregulated genes, and a ranked gene list generated for GSEA (Table 20). Overall, the differential expression analysis discovered a substantial number of significantly dysregulated genes, providing a solid foundation for future functional annotation and prognostic biomarker discovery in glioblastoma.

Table 14: Group Summary

Group	Samples
Tumor	372
Normal	5

Table 15: Gene Filtering Summary

Description	Count
Genes Before Filtering	60660
Genes After Filtering	41864
Genes Removed	18796

Table 16: limma Differential Expression Analysis Summary

Description	Count
Genes Analyzed	41864
Tumor Samples	372
Normal Samples	5

Table 17: Summary of Differentially Expressed Genes

Category	Count
Total Genes	41864
Significant DEGs	15098
Upregulated	8922
Downregulated	6176

Table 18: Heatmap Dataset Summary

Description	Count
Genes Displayed	50
Tumor Samples	372
Normal Samples	5

Table 19: Differential Expression Statistics

Parameter	Value
Total Genes Analyzed	41864
Significant DEGs	15098
Upregulated Genes	8922
Downregulated Genes	6176
Maximum log2FC	9.4756
Minimum log2FC	-8.5592
Mean log2FC	0.2962
Median log2FC	1.449
Minimum Adjusted P-value	7.28E-45

Table 20: Differential Expression Export Summary

Export	Count
All Genes	14957
Upregulated	8787
Downregulated	6170
Ranked Genes (GSEA)	14957

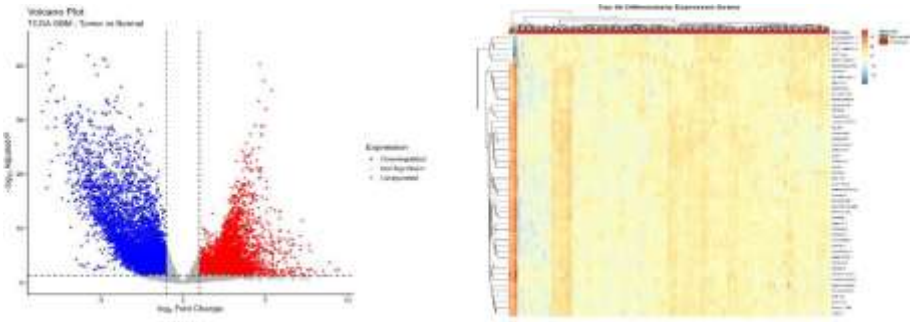


Figure 5A: Volcano Plot of DEGs Figure 5B: Heatmap of Top 50 DEGs

Figure 5. Differential Expression Analysis of the TCGA Glioblastoma Cohort Using the Limma Package

Volcano plot showing significantly upregulated (red) and downregulated (blue) genes between glioblastoma tumor and normal brain tissues based on the predefined thresholds of $|\log_2$ fold change$| > 1$ and adjusted $P < 0.05$. Grey dots represent non-significant genes. (B) Heatmap illustrating the expression patterns of the top 50 differentially expressed genes across tumor and normal samples, demonstrating clear separation between the two groups.

Kaplan–Meier Survival Analysis

Discover genes Typically considerably linked to overall survival in glioblastoma patients, expression and clinical information were combined prior to survival analysis. Initially, the TCGA expression dataset included 372 primary tumour samples. After matching with relevant clinical information, 283 patients with complete survival data were preserved for downstream analysis, together with 60,660 genes, resulting in a high-quality dataset for prognostic evaluation (Table 21). To increase statistical reliability and eliminate background noise, genes with low expression were deleted before to survival analysis. Gene filtering decreased the dataset from 60,660 to 41,864 genes, while all 283 matched patients were kept for further survival analysis (Table 22). All 41,864 filtered genes were subjected to univariate Cox proportional hazards regression analysis in order for finding genes associated with overall survival. All genes were successfully fitted with Cox regression models, resulting in the

identification of 76 statistically significant genes ($P < 0.05$). These genes continued to be important after false discovery rate (FDR) correction. Notably, all 76 genes were classed as risk-associated, with no protective genes found (Table 23).

The potential genes' prognostic value was evaluated more thoroughly utilizing the Kaplan-Meier survival analysis. Kaplan-Meier analysis were performed on 66 genes, with the median gene expression value serving as the cutoff point for dividing individuals across categories based on their expression levels. Among these, 35 genes had statistically significant changes in overall survival among two expression groups (Table 24). These prognostically significant genes were then chosen for downstream prognostic model building and validation, laying a solid platform for discovering clinically relevant biomarkers linked with glioblastoma patient survival.

Table 21. Data Preparation Summary

Parameter	Value
Primary Tumor Samples	372
Matched Patients	283
Genes	60660
Clinical Records	283

Table 22. Gene Expression Filtering Summary

Parameter	Value
Genes Before Filtering	60660
Genes After Filtering	41864
Patients	283

Table 23. Univariate Cox Regression Summary

Metric	Count
Total DEGs Tested	41864
Successful Cox Models	41864
Significant Genes (P<0.05)	76
Significant Genes (FDR<0.05)	76
Risk Genes	76
Protective Genes	0

Table 24. Kaplan–Meier Survival Analysis Summary

Metric	Value
Genes Tested	66
Kaplan-Meier Significant	35
Median Cutoff	Median

Cox Proportional Hazards Regression Analysis

To determine an independent prognosis markers connected to overall survival in glioblastoma patients, multivariable Cox proportional hazards regression analysis were carried out upon 35 candidate genes identified by the Kaplan-Meier survival study. The analysis comprised 283 patients who provided complete clinical information. The multivariable Cox model included all 35 genes, 5 of which were identified as statistically significant independent prognostic variables. The model had strong precision of a prediction, accompanied by a consistency value (C-index) of 0.7101 (SE = 0.0205), demonstrating adequate discriminating across patients with varying survival outcomes. The likelihood ratio test ($\chi^2 = 110.077$, $P = 1.09 \times 10^{-9}$), Wald test ($\chi^2 = 125.750$, $P = 3.67 \times 10^{-12}$), as well as score (log-rank) test ($\chi^2 = 253.724$, $P = 5.45 \times 10^{-35}$) confirmed the multivariable model's prognostic relevance (Table 25).

Schoenfeld residuals were used to evaluate the Cox regression model's proportional hazards assumption. The global proportional hazards test yielded a P-value of 0.0055, implying that presumption of equal hazards failed completely (Table 26).

The final Cox regression analysis showed 25 risk-associated genes (hazard ratio > 1) and 10 protective genes (hazard ratio < 1), with 5 genes having statistically significant relationships with overall survival. The final model contained all 35 potential genes and 283

patients, laying the groundwork for future prognostic model development and risk score development (Table 27). Figure 6 shows the predicted hazard ratios and 95% confidence intervals for all candidate genes.

Table 25: Cox Proportional Hazards Regression Summary

Parameter	Value
Total Patients	283
Total Genes	35
Significant Genes	5
Concordance Index	0.7101
Concordance SE	0.0205
Likelihood Ratio Statistic	110.077
Likelihood Ratio P	1.09E-09
Wald Statistic	125.75
Wald P	3.67E-12
Score Statistic	253.724
Score P	5.45E-35

Table 26: Proportional Hazards Assumption Test Summary

Total Variables	Global PH P Value	PH Assumption
35	0.005482553	Violated

Table 27: Final Cox Regression Analysis Summary

Parameter	Value
Total Patients	283
Total Genes Entered	35
Significant Genes	5
Risk Genes	25
Protective Genes	10
Concordance Index	0.7101
Concordance SE	0.0205
Global PH P-Value	0.005483
Ph-Assumption	Violated

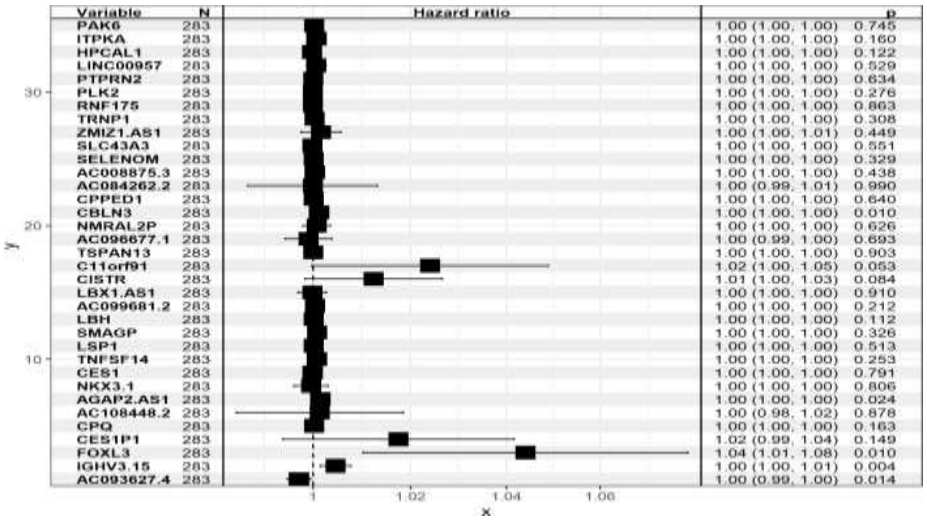


Figure 6: Forest Plot Showing the Hazard Ratios (Hrs) and 95% Confidence Intervals of the 35 Candidate Prognostic Genes Included in the Multivariable Cox Proportional Hazards Regression Model.

Squares Represent Hazard Ratios, Horizontal Lines Indicate The Corresponding 95% Confidence Intervals, And The Dashed Vertical Line Represents A Hazard Ratio Of 1. Genes With Hazard Ratios Greater Than 1 Were Associated With An Increased Risk Of Mortality, Whereas Genes With Hazard Ratios Less Than 1 Were Associated With A Protective Effect On

Overall Survival.

Lasso Cox Regression Analysis

To decrease model complexity and identify the most informative prognostic biomarkers, the 35 candidate genes determined based on Cox proportional hazards regression analysis had been subjected to least absolute shrinkage and selection operator (LASSO) Cox regression. The LASSO model consisted of 283 patients. Using the ideal penalty parameters of λ_{min} = 0.055606 and λ_{1se} = 0.204539, the analysis chose 18 genes with non-zero regression coefficients, therefore reducing the number of candidate genes and maintaining those with the highest prognostic contribution (Table 28).

The LASSO coefficient path showed that as the regularisation parameter (λ) grew, regression coefficients shrank, eliminating less useful variables and maintaining genes with increasing prognostic relevance (Figure 7A). The best λ values were obtained using 10-fold cross-validation, where λ_{min} related with least cross-validation error and λ_{1se} represented highly regularized designs within a single minimum error of the poorest (Figure 7B).

The final LASSO model included 18 prognostic genes based on their optimal λ value. FOXL3 had the highest regression coefficient (0.013765), followed by C11orf91 (0.009600), CISTR (0.008247), and CES1P1 (0.005500), while the remaining genes had smaller but non-zero coefficients, indicating their contribution to the prognostic model (Table 29).

The LASSO coefficient path showed that as the regularisation parameter (λ) grew, regression coefficients shrank, eliminating less useful variables and maintaining genes with increasing prognostic relevance (Figure 7A).

Table 28: Lasso Cox Regression Summary

Parameter	Value
Total Patients	283
Candidate Genes	35
Selected Genes	18
Lambda Min	0.055606
Lambda 1SE	0.204539

Table 29: Genes Selected by LASSO Cox Regression

Gene	Coefficient
FOXL3	0.013765
C11orf91	0.0096
CISTR	0.008247
CES1P1	0.0055
ZMIZ1-AS1	0.001901
IGHV3-15	0.001705
AGAP2-AS1	0.001083
AC008875.3	0.000398
CBLN3	0.000316
ITPKA	0.000284
TNFSF14	0.000245
TRNP1	0.000213
CPQ	0.000107
SELENOM	6.48E-05
LBH	2.20E-05
SLC43A3	2.16E-05
RNF175	1.15E-05

Table 30: Final Lasso Cox Regression Summary

Parameter	Value
Total Patients	283
Candidate Genes	35
LASSO Selected Genes	18
Lambda Min	0.055606

Lambda 1SE	0.204539
Analysis	LASSO Cox Regression

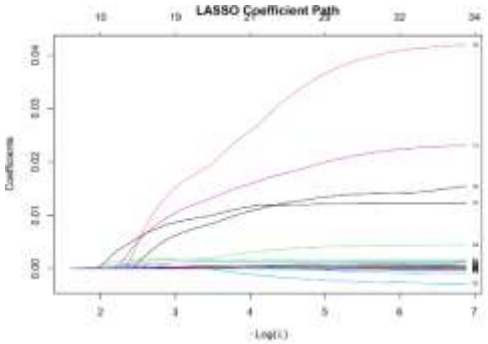


Figure 7A: LASSO coefficient path

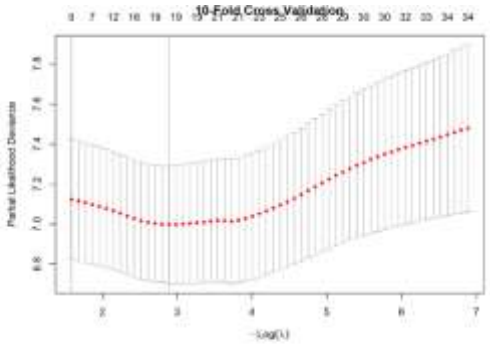


Figure 7B: Ten-fold cross-validation for optimal λ selection.

Figure 7. LASSO Cox Regression Analysis for Prognostic Gene Selection.

(C) LASSO coefficient profiles of the 35 candidate prognostic genes showing the shrinkage of
(D) regression coefficients with increasing regularization parameter (−log λ).
(E) Ten-fold cross-validation curve used to determine the optimal penalty parameter (λ). The two dashed vertical lines indicate λ_{min}, corresponding to the minimum cross-validation error, and λ_{1se}, representing the most regularized model within one standard error of the minimum. Based on the optimal λ value, 18 prognostic genes with non-zero coefficients were selected for subsequent prognostic model construction.

Time-dependent Receiver Operating Characteristic (ROC) Analysis

A median score for risk of 1.135096 was used to categorise patients into high-risk as well as low to assess prognostic performance based on LASSO-derived gene profile. Among the 283 patients included in the analysis, 142 were classed as high-risk, while 141 were classified as low-risk, yielding two virtually equal cohorts for evaluating the prognostic model's predictive ability.

The prognostic model's Statistical prediction precision had been then assessed with time-dependent receiver operating characteristic (ROC) analysis at one, three, and five years of overall survival. The model had an area under the curve (AUC) of 0.7711 at one year, 0.8560 at three years, and 0.8864 at five years, suggesting good to excellent discrimination in predicting patient survival throughout different follow-up periods (Table 32).

The time-dependent ROC curves revealed the prognostic model's predictive capabilities at all three evaluation time periods (Figure 8). The gradual improvement in AUC values from 0.7711 at one year to 0.8864 at five years suggested better prediction performance over longer follow-up periods. The final ROC analysis, which included all 283 patients, indicated that the LASSO-derived prognostic model had strong and consistent predictive accuracy for overall survival in glioblastoma patients (Table 33).

Table 31: Risk Group Summary

Parameter	Value
Total Patients	283
High Risk	142
Low Risk	141
Median Cutoff	1.135096

Table 32: Time-Dependent Auc Summary

Time	Days	AUC
1 Year	365	0.7711
3 Years	1095	0.856
5 Years	1825	0.8864

Table 33: Final Time-Dependent Roc Analysis Summary

Analysis	Time-dependent ROC
Patients	283
One Year AUC	0.7711
Three Year AUC	0.856
Five Year AUC	0.8864

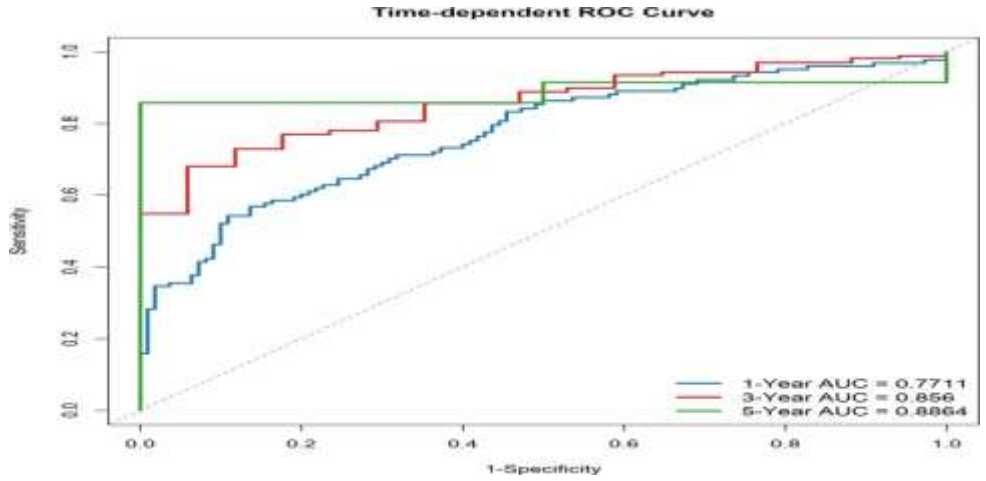


Figure 8: Time-Dependent Receiver Operating Characteristic (ROC) Curves Showing the Predictive Performance of the LASSO-Derived Prognostic Model for Overall Survival in the TCGA Glioblastoma Cohort. ROC Curves Are Presented For 1-Year (Blue), 3-Year (Red), And 5-Year (Green) Survival. The Corresponding Areas under the Curve (Aucs) Were 0.7711, 0.8560, And 0.8864, Respectively, Demonstrating Good Predictive Performance of the Prognostic Model across Different Follow-Up Periods. The Diagonal Dashed Line Represents the Performance of A Random Classifier (AUC = 0.5).

Nomogram Construction

The independent prognostic variables found in the multivariable Cox regression model had been applied for the creation of a nomogram in order to develop an individualised prognostic prediction model for glioblastoma patients. The nomogram used age and risk score to estimate overall survival. The regression coefficients were 0.027826 for age and 1.401713 for the risk score, showing that the risk score had a greater impact on the prognostic model than age (Table 34).

The created nomogram includes a graphical grading system for assessing the likelihood of 1-year, 3-year, and 5-year overall survival. Each predictor is given a score according to its relative contribution, and the total score is converted into the pertinent survival probability (Figure 9).

The final nomogram was created using data from 283 individuals and included age and the LASSO-derived risk score in a multivariable Cox regression model, resulting in an individualised tool for predicting survival outcomes in patients with glioblastoma (Table 35).

Table 34. Nomogram Model Variables

Variable	Coefficient
Age years	0.027826
Risk Score	1.401713

Table 35. Final Nomogram Model Summary

Item	Value
Total Patients	283
Predictors	Age + Risk Score
Model	Multivariable Cox
Nomogram	Generated

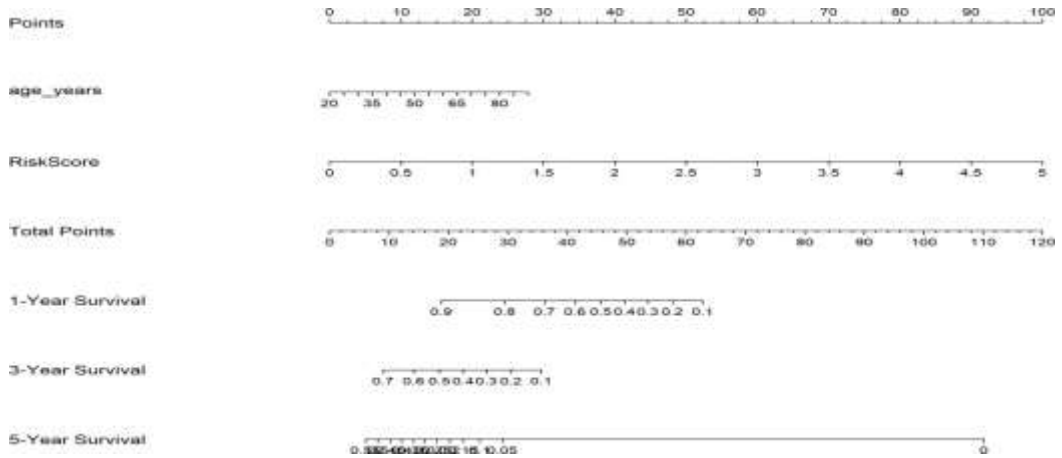


Figure 9: Nomogram Developed Using A Multivariable Cox Regression Model Incorporating Age and the LASSO-Derived Risk Score To Predict 1-Year, 3-Year, and 5-Year Overall Survival in Patients with Glioblastoma. Individual Values for Each Predictor are Assigned Corresponding Points, Which are summed to generate a Total Point Score. The Total Score is then used To Estimate the Probability of Survival at the Three Specified Time Points.

Calibration Analysis

The nomogram developed for the TCGA glioblastoma cohort's anticipated and observed overall survival probability were compared using calibration analysis. Calibration curves were created for one- and three-year overall survival (Figure 10).

The 1-year The calibration curve demonstrated a strong correlation among anticipated also noted survival chances, using the calibration plot close to the ideal 45° reference line throughout most probability ranges (Figure 10A). Similarly, the 3-year calibration curve demonstrated satisfactory agreement between anticipated and actual survival probability, with somewhat more variability at higher expected survival probabilities, indicating decreasing sample availability during extended follow-up (Figure 10B).

Overall, the calibration study revealed satisfactory consensus within projected as well as noted survival outcomes, implying that the nomogram gave accurate survival forecasts for glioblastoma patients.

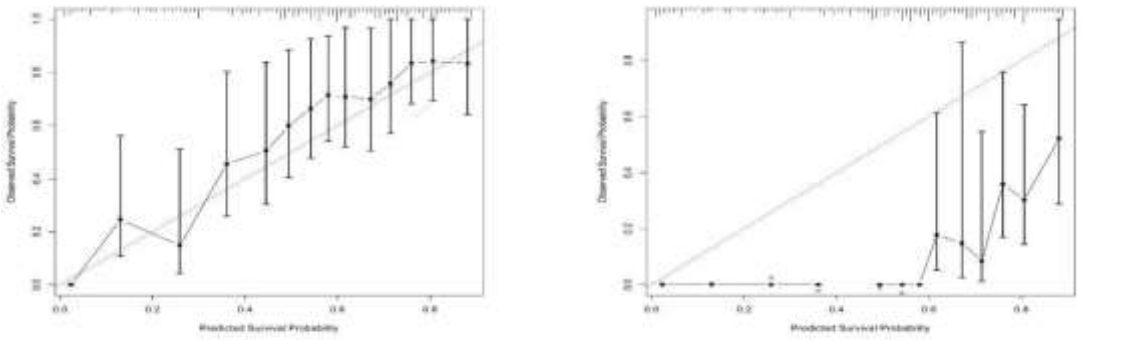


Figure 10A: One-year calibration curve Figure 10B: Three-year calibration curve

Figure 10: Calibration Curves Evaluating the Agreement between the Predicted and Observed Overall Survival Probabilities of the Nomogram

One-year calibration curve. (B) Three-year calibration curve. The red dashed diagonal line represents the ideal prediction, where the predicted survival probabilities

Perfectly match the observed outcomes. Points closer to the reference line indicate better calibration of the prognostic model.

Decision Curve Analysis (DCA)

Decision curve analysis (DCA) was used For assessing a medical effectiveness based on LASSO-derived prognostic model by measuring the net benefit across a variety of threshold probabilities. DCA curves were created for one-, three-, and five-year overall survival (Figure 11).

The prognostic model fared better than the treat-all and treat-none approaches throughout a broad range of clinically significant threshold probabilities, according to the one-year DCA

(Figure 11A). Similarly, the 3-year DCA demonstrated sustained clinical improvement over a broad spectrum of threshold probabilities, demonstrating enhanced choice-making performance than conventional treatment techniques (Figure 11B).

The 5-year DCA revealed that the prognostic model maintained good clinical utility at higher threshold probabilities, with only minor changes in net benefit found at the top end of the threshold range (Figure 11C). Overall, the DCA results showed that the LASSO-derived prognostic model provided a greater clinical overall advantage above the standard approaches of treatment every patient or none at all, indicating its potential for individualised prognostic assessment in glioblastoma patients.

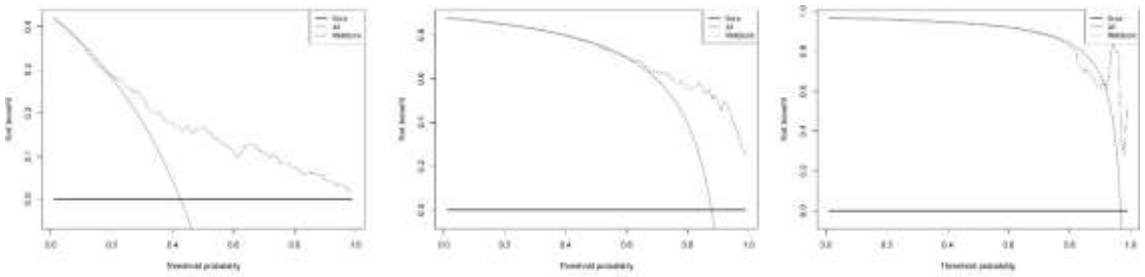


Figure 11A: One-year DCA Figure 11B: Three-year DCA Figure 11B: Five-year DCA

Figure 11. Decision Curve Analysis (DCA) Evaluating the Clinical Utility of the LASSO-Derived Prognostic Model for Predicting Overall Survival in the TCGA Glioblastoma Cohort

(A) One-year DCA. (B) Three-year DCA. (C) Five-year DCA. The dashed black line represents the prognostic model (Risk Score), the grey line represents the strategy of treating all patients (All), and the solid black horizontal line represents the strategy of treating no patients (None). Higher net benefit across a range of threshold probabilities indicates superior clinical usefulness of the prognostic model.

GO and KEGG Enrichment Analysis

Enrichment studies of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) was used for study biological functions as well as signaling pathways linked to prognostic genes. Enrichment study includes 10,357 genes, with 718 significantly enriched Biological Process (BP) terms, 177 Cellular Component (CC) terms, 127 Molecular Function (MF) terms, also 80 highly enriched KEGG pathways.

GO Biological Process enrichment revealed that prognostic genes were primarily involved in trans-synaptic signaling regulation, modulation of chemical synaptic transmission, neurone differentiation-related cell morphogenesis, membrane potential regulation, signal release, axon development, axonogenesis, synapse assembly, neurotransmitter transport, potassium ion transport, and synaptic plasticity regulation (Figure 12A-B). Such outcomes imply how discovered genes are largely involved in neural transmission and synaptic architecture.

Significant enrichment was found in synaptic membrane, neuronal cell body, neuron-to-neuron synapse, transmembrane transporter complex, postsynaptic specialisation, postsynaptic density, asymmetric synapse, monoatomic ion channel complex, dendritic spine, presynaptic membrane, exocytic vesicle, and GABAergic synapse (Figure 12C-D), indicating that prognostic genes are primarily localised to neuronal also synaptic structures.

The significance of ion transport and channel regulation in glioblastoma progression was highlighted by molecular function enrichment, which revealed that the significantly enriched genes were primarily linked to monoatomic ion channel activity, metal ion transmembrane transporter activity, gated channel activity, voltage-gated channel activity, ligand-gated ion channel activity, potassium channel activity, calcium-dependent phospholipid binding, calmodulin binding, neurotransmitter receptor activity, and transporter regulator activity (Figure 12E–F).

Neuroactive ligand–receptor interaction, calcium signaling, MAPK signaling, cAMP signaling, glutamatergic synapse, GABAergic synapse, cholinergic synapse, dopaminergic synapse, serotonergic synapse, oxytocin signaling pathway, retrograde endocannabinoid signaling, and IgSF CAM signaling were all significantly enriched, according to KEGG pathway enrichment analysis (Figure 12G–H). These findings collectively imply that the prognostic genes are intimately linked to intracellular signaling pathways, ion channel

modulation, neuronal signaling, and synaptic transmission, this could contribute to the development and metastasis of glioblastoma.

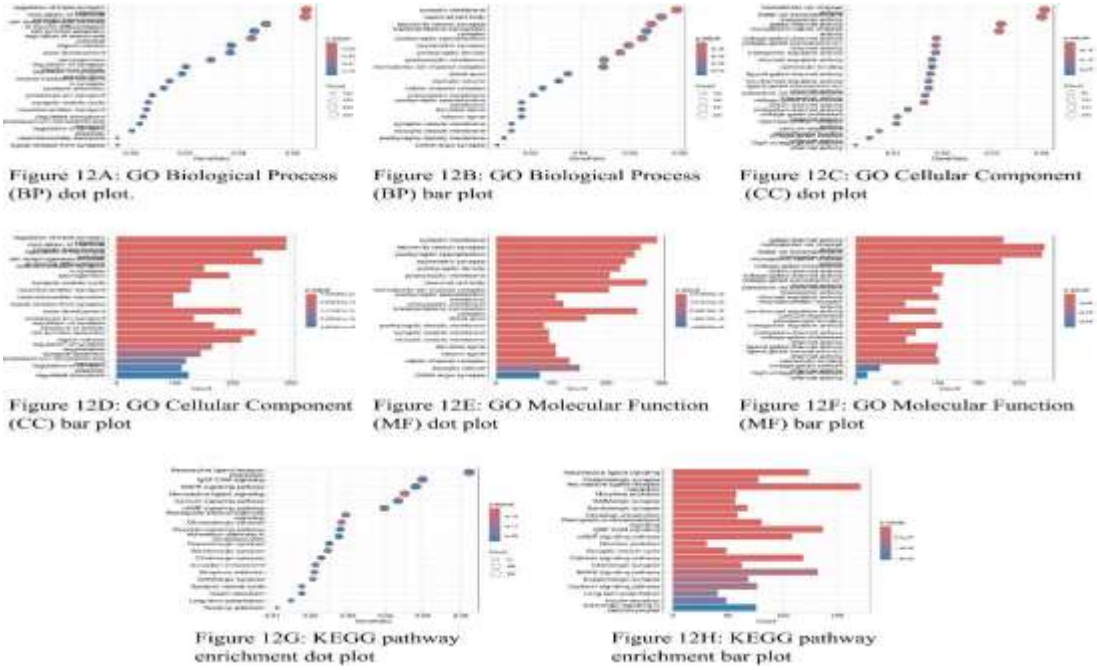


Figure 12: Functional Enrichment Analysis of the Prognostic Genes

(A) GO Biological Process (BP) dot plot. (B) GO Biological Process (BP) bar plot. (C) GO Cellular Component (CC) dot plot. (D) GO Cellular Component (CC) bar plot. (E) GO Molecular Function (MF) dot plot. (F) GO Molecular Function (MF) bar plot. (G) KEGG pathway enrichment dot plot. (H) KEGG pathway enrichment bar plot. Bubble size represents the number of enriched genes, colour indicates the adjusted P-value, and bar length represents the number of genes associated with each significantly enriched GO term or KEGG pathway.

Gene Set Enrichment Analysis (GSEA)

Ranking differential gene expression profile was used in Gene Set Enrichment Analysis (GSEA) to find biological pathways linked to glioblastoma progression. To describe the functional changes that underlie the predictive gene signature, Hallmark and Canonical (Reactome/WikiPathways) gene sets were examined.

Several biological processes were shown to be considerably enhanced by Hallmark GSEA (Figure 13A). G2M checkpoint, E2F targets, interferon-α response, interferon-γ response, angiogenesis, as well as epithelial–mesenchymal transition (EMT) were among most positively enriched Hallmark gene sets, suggesting increased cell-cycle progression, proliferative activity, immune-related signaling, tumour vascularization, and invasive potential in glioblastoma. On the other hand, peroxisome, fatty acid metabolism, bile acid metabolism, apical junction, also spermatogenesis were among negatively enriched Hallmark gene sets, indicating inhibition of multiple metabolic and cellular homeostasis pathways. The activation of cell-cycle regulatory systems in the cancer samples was confirmed by the enrichment plot of the Hallmark G2M checkpoint gene set, which showed considerable positive enrichment toward the highly expressed genes (Figure 13B).

Significant enrichment of pathways linked to transcriptional control and neuronal signaling was also shown by canonical pathway GSEA (Figure 13C). The dysregulation of epigenetic regulation, DNA replication, and transcriptional control in glioblastoma was highlighted by positively enriched pathways such as activation of anterior HOX genes during hindbrain development, HDAC-mediated histone deacetylation, histone acetylation, transcriptional regulation by small RNAs, and assembly of the ORC complex at the origin of DNA replication. On the other hand, a number of neuronal signaling pathways, like the neuronal system, transmission across chemical synapses, neurotransmitter receptors and

postsynaptic signal transmission, potassium channels, and GABA and glutamate signaling, demonstrated significant negative enrichment, suggesting that normal neuronal communication is disrupted during tumour progression. The Reactome Neuronal System pathway's enrichment plot showed a significant negative enrichment score, indicating that genes involved in typical neuronal. The Reactome Neuronal System pathway enrichment plot showed a significant negative enrichment score, indicating that GBM primarily downregulated genes involved in normal neuronal function (Figure 13D).

Overall, GSEA showed that glioblastoma is characterised by suppression of neuronal signaling, synaptic transmission, ion channel activity, and metabolic processes, as well as activation of pathways controlling cell-cycle progression, proliferation, immune response, angiogenesis, and epigenetic regulation. These findings support the aggressive biological phenotype identified by the prognostic gene signature.

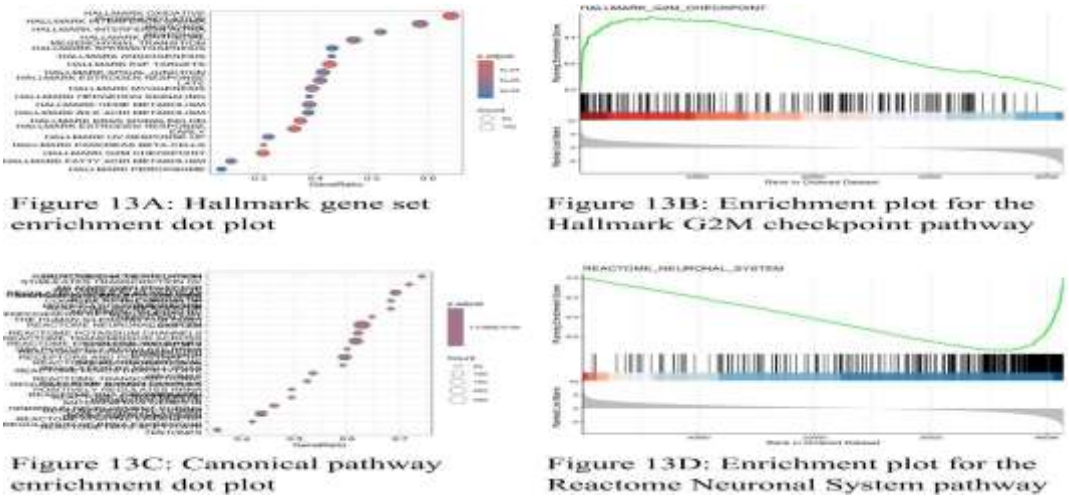


Figure 13A: Hallmark gene set enrichment dot plot

Figure 13B: Enrichment plot for the Hallmark G2M checkpoint pathway

Figure 13C: Canonical pathway enrichment dot plot

Figure 13D: Enrichment plot for the Reactome Neuronal System pathway

Figure 13: Gene Set Enrichment Analysis (GSEA) of the Ranked Gene Expression Profile.

(A) Hallmark gene set enrichment dot plot showing significantly enriched Hallmark pathways. (B) Enrichment plot for the Hallmark G2M checkpoint pathway demonstrating positive enrichment in glioblastoma samples. (C) Canonical (Reactome/WikiPathways) pathway enrichment dot plot showing significantly enriched biological pathways. (D) Enrichment plot for the Reactome Neuronal System pathway demonstrating negative enrichment in glioblastoma samples. Bubble size represents the number of enriched genes, colour indicates the adjusted P-value, and the enrichment plots display the running enrichment score across the ranked gene list.

STRING Protein–Protein Interaction (PPI) Network Analysis

Examine functional relationships among the glioblastoma-associated proteins, a STRING protein–protein interaction (PPI) network was built. Top 15 highly linked proteins were chosen from the STRING network and shown in Cytoscape for easy visualisation (Figure 14). With an average node degree of 11.2, final network consisted with 15 proteins linked by 84 interaction connections. A highly integrated protein interaction network was indicated by the maximum and minimum node degrees of 14 and 2, respectively.

Significant relationships between proteins involved in angiogenesis, apoptosis, signal transduction, cell proliferation, and cancer growth were shown by the visualised network. The interaction network's essential roles were indicated by ALB's highest degree of connection, which was followed by KRAS, HSP90AA1, HRAS, MYC, EGFR, TP53, VEGFA, MAPK3, and TNF. The overall network architecture and functional connections were also influenced by other proteins, such as FN1, NOTCH1, CDK1, DLG4, and SNAP25. When taken as a whole, these connections imply that the chosen proteins take part in coordinated biological processes linked to the development of glioblastoma.

Overall, the STRING PPI analysis showed that the top 15 highly linked proteins constituted a dense interaction network, confirming the potential biological significance of these proteins and highlighting important proteins that may contribute to glioblastoma

pathogenesis.

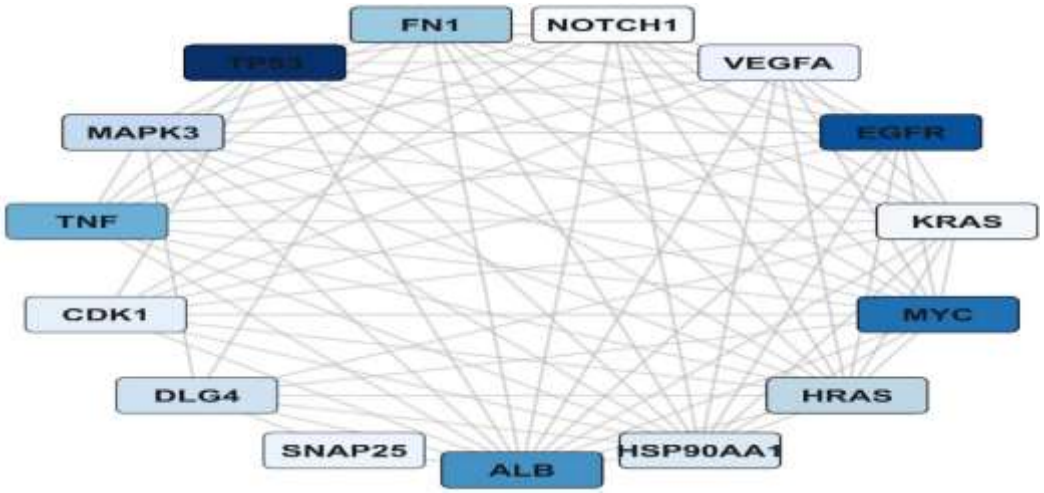


Figure 14. STRING Protein–Protein Interaction (PPI) Network Illustrating the Top 15 Highly Connected Proteins Identified From the STRING Interaction Network. Nodes Represent Proteins, while Edges Represent Predicted or Experimentally Validated Protein–Protein Interactions. Node Colour Intensity and Node Size Correspond To the Degree of Connectivity, with Darker and Larger Nodes Indicating Proteins With Higher Interaction Degrees. The Network Highlights the Complex Interactions among Key Proteins Involved in Glioblastoma-Associated Biological Processes.

Immune Cell Infiltration Analysis

Using MCPcounter method, immune cell infiltration analysis was carried out to assess the immune and stromal cell populations' abundance between the prognostic signature-identified low- and high-risk groups. The infiltration scores of eight immunological and stromal cell groups (B-lineage cells, $CD8^+$ T cells, cytotoxic lymphocytes, fibroblasts, monocytic lineage cells, myeloid dendritic cells, neutrophils, and T cells) varied statistically significantly (Figure 15).

The high-risk group had substantially higher B-lineage cell infiltration score than the low-risk group (Figure 15A). Similarly, high-risk patients showed greater infiltration of $CD8^+$ T cells (Figure 15B). Additionally, the high-risk group's cytotoxic lymphocyte abundance was much higher than that of the low-risk group (Figure 15C).

Fibroblasts showed noticeably higher infiltration scores among the stromal cell types in the high-risk group (Figure 15D). Similarly, high-risk patients had a markedly higher infiltration of monocytic lineage cells (Figure 15E). In the high-risk group, myeloid dendritic cell infiltration was also considerably greater (Figure 15F).

Additionally, compared to the low-risk group, neutrophils in the high-risk group showed noticeably higher infiltration scores (Figure 15G). In a similar vein, T cells showed noticeably more infiltration in high-risk individuals than in low-risk patients (Figure 15H). The tumour immune microenvironment differed significantly between the two prognostic groups, as evidenced by the increased infiltration of several immune and stromal cell types in the high-risk glioblastoma grouping.

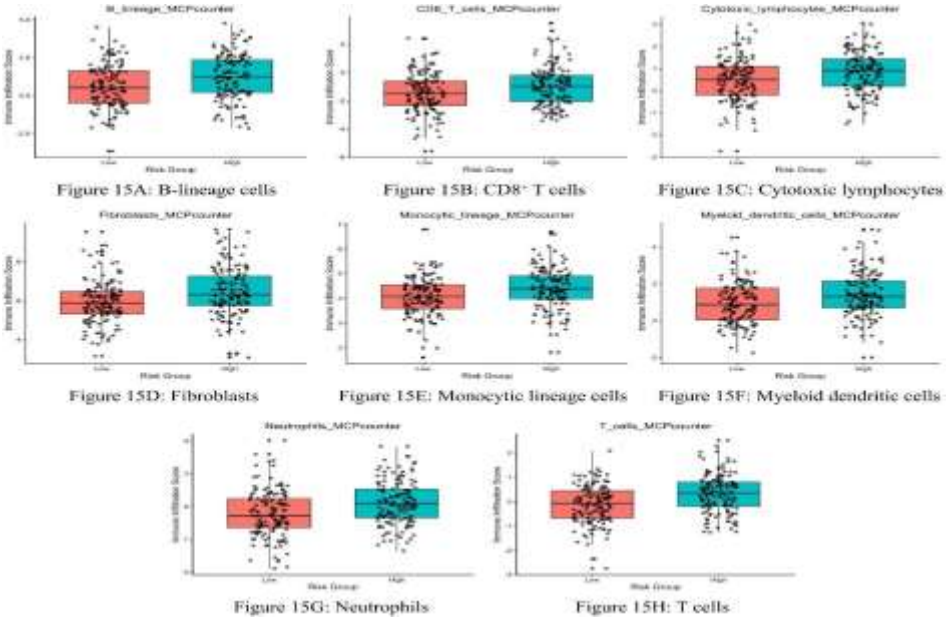


Figure 15: Comparison of Significant Immune Cell Infiltration between Low- And High-Risk Glioblastoma Patients Using Mcpcounter. Boxplots Illustrate the Distribution of Infiltration

Scores for (A) B-Lineage Cells, (B) Cd8⁺ T Cells, (C) Cytotoxic Lymphocytes (D) fibroblasts, (E) monocytic lineage cells, (F) myeloid dendritic cells, (G) neutrophils, and (H) T cells. Each point represents an individual patient sample, the horizontal line within each box indicates the median, the box represents the interquartile range (IQR), and the whiskers denote the spread of the data. Only immune and stromal cell populations showing statistically significant differences between the low- and high-risk groups (FDR < 0.05) are presented.

CONCLUSION

This work developed and externally assessed a comprehensive transcriptome-based glioma prediction model using integrated bioinformatics analysis of the TCGA and CGGA cohorts. A comprehensive quality evaluation determined that both datasets were suitable for further transcriptome and survival research. Differential expression analysis revealed 15,098 significantly dysregulated genes, including 6,176 downregulated and 8,922 upregulated genes, suggesting extensive transcriptional alterations associated with glioblastoma. Following survival analyses using Kaplan-Meier analysis to identify prognostically relevant genes, multivariable Cox proportional hazards regression was used to further refine these candidates into independent prognostic biomarkers. LASSO Cox regression reduced model overfitting and generated an ideal model for risk stratification, resulting in an 18-gene prognostic signature.

The prognostic signature demonstrated strong predictive performance using time-dependent ROC analysis, and it was transformed into a personalised prognostic nomogram. The calibration investigation revealed good agreement between predicted and observed survival rates, while decision curve analysis indicated potential therapeutic efficacy over clinically relevant threshold probabilities. The independent CGGA cohort was used to externally validate the robustness, repeatability, and generalisability of the prognostic model.

Gene Ontology, KEGG pathway, and Gene Set Enrichment analyses showed that the discovered prognostic genes were considerably enriched in several biological processes and signalling pathways linked to glioblastoma progression. Highly interconnected molecular networks were also found by protein–protein interaction network analysis, indicating coordinated regulation of genes implicated in cancer growth. Furthermore, immune infiltration analysis revealed different immune microenvironment features between high- and low-risk groups, demonstrating the role of immune-related mechanisms in glioblastoma progression and bolstering the biological significance of the suggested prognostic signature.

In order to enhance prognostic stratification in glioblastoma, this work offers a repeatable computational method that combines transcriptome profiling, survival modelling, pathway enrichment, protein interaction

analysis, immune landscape evaluation, and independent external validation. The suggested predictive signature could help with individualised risk assessment and offer potential biomarkers for upcoming pharmacological and mechanistic studies. However, prior to standard clinical application, experimental validation and prospective multicenter clinical study are necessary because the results were obtained using retrospective public transcriptome data.

REFERENCES

1. Hill CM, Nwagwu CD, Odukoya AO, Hsueh B, Wang AZ, Dunn GP. Tumor infiltrating lymphocytes in glioblastoma: immunobiology and translational implications. NPJ Precis Oncol. 2026 Jul 13;10(1):274. doi:10.1038/s41698-026-01330-3 PubMed PMID: 42437767; PubMed Central PMCID: PMC13357577.
2. Guterres A. Circadian-guided oncolytic virotherapy for glioblastoma. J Neurovirol. 2026 Jul 6;32(4):24. doi:10.1007/s13365-026-01325-7 PubMed PMID: 42406201; PubMed Central PMCID: PMC13337691.
3. Javaid S, Song W, Ali S, Hou W, Su X, Wang H, et al. BBB-aware stimuli-responsive and biomimetic nanomedicines for glioblastoma. Chin Neurosurg J. 2026 Jul 1;12(1):20. doi:10.1186/s41016-026-00438-6 PubMed PMID: 42380984; PubMed Central PMCID: PMC13321699.
4. Wu WYY, Numan Hellquist B, Melin B, Björkblom B, Sjöberg RL. Intratumoral serotonin and antidepressants in glioblastoma patients: narrowing the uncertainties. J Neurooncol. 2026 Jun 30;178(3):84. doi:10.1007/s11060-026-05694-1 PubMed PMID: 42377764; PubMed Central PMCID: PMC13319255.
5. Mukherjee A. Immune checkpoint crosstalk between LAG-3 and CD39/CD73 in glioblastoma: dual-pathway regulation of metabolic exhaustion and therapeutic reversal strategies. J Egypt Natl Cancer Inst. 2026 Jun 15;38(1):21. doi:10.1186/s43046-026-00361-y PubMed PMID: 42289542; PubMed Central PMCID: PMC13313299.
6. Woodson K. Molecular and Treatment-Associated Correlates of Survival in Glioblastoma: Complementary Surveillance, Epidemiology, and End Results (SEER) and The Cancer Genome Atlas-Glioblastoma (TCGA-GBM) Cohort Analyses. Cureus. 2026 Jul;18(7):e112222. doi:10.7759/cureus.112222 PubMed PMID: 42421802; PubMed Central PMCID: PMC13343246.
7. Putra MAR, Ghozali SAS, Annisa MF. Focused ultrasound-induced blood-brain barrier modulation for drug delivery in recurrent glioblastoma: A systematic review. Indian J Med Res. 2026 Jul;164(1):19–28. doi:10.25259/IJMR_2529_2025 PubMed PMID: 42397830; PubMed Central PMCID: PMC13363183.
8. Pouyan A, Ghorbanlo M, Eslami M, Jahanshahi M, Ziaei E, Salami A, et al. Correction: Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. Mol Cancer. 2026 Jun 26;25(1):156. doi:10.1186/s12943-026-02717-3 PubMed PMID: 42363151; PubMed Central PMCID: PMC13307701.
9. Mahmoud O, Mounir M, Gad W. Graph informed biomarker discovery framework using transcriptomic machine learning for glioblastoma prognosis. Sci Rep. 2026 Jun 23;16(1):19530. doi:10.1038/s41598-026-58062-4 PubMed PMID: 42336875; PubMed Central PMCID: PMC13291226.
10. Barbe-Richaud JB, Cebula H, Todeschi J, Santin M des N, Baloglu S, Schott R, et al. From relapse to survival: real-world clinical impact of a multimodal approach in first-time recurrent glioblastoma : FROST: multimodal treatment for GBM recurrence. J Neurooncol. 2026 Jun 23;178(3):73. doi:10.1007/s11060-026-05635-y PubMed PMID: 42334687; PubMed Central PMCID: PMC13290778.
11. Sharpe MA, Baskin AM, Teh BS, Butler EB, Baskin DS. Optimizing the Stupp protocol for treatment of glioblastoma: eliminating age bias, enhancing treatment timing, use of stereotactically-guided sequential boost, and dexamethasone dosing. J Neurooncol. 2026 Jul 3;178(3):88. doi:10.1007/s11060-026-05685-2 PubMed PMID: 42397615; PubMed Central PMCID: PMC13331880.
12. Williams ST, Kingdon S, Rominiyi O, Lorimer C, Chandy E, Thompson M, et al. Report of treatment intensity and survival outcomes in older patients with glioblastoma diagnosed according to WHO CNS 5 classification. J Neurooncol.

2026 Jun 4;178(2):41. doi:10.1007/s11060-026-05628-x PubMed PMID: 42240773; PubMed Central PMCID: PMC13236755.

13. Ivanidze J, Vu KV, Fu R, Brandmaier A, Szidonya L, Choudhary G, et al. Clinical Utility of [18F]-Fluciclovine PET/MRI for Differentiating True Progression from Treatment-Related Changes in Patients with Glioblastoma. AJNR Am J Neuroradiol. 2026 Jun 3;47(6):1578–86. doi:10.3174/ajnr.A9120 PubMed PMID: 41326230; PubMed Central PMCID: PMC13020911.
14. Mishchenko TA, Olajide OJ, Gorshkova EN, Vedunova MV, Krysko DV. Reprogramming temozolomide response in glioblastoma through regulated and immunogenic cell death modalities. Cell Death Discov. 2026 May 26. doi:10.1038/s41420-026-03151-6 PubMed PMID: 42191676.
15. Zeng L, Li D, Du M, Wu T, Liao Y, Huang Y, et al. Integrative bioinformatics analysis unveils neuro-cancer crosstalk-related genes and establishes prognostic risk model in Glioblastoma. Sci Rep. 2026 May 12;16(1):21699. doi:10.1038/s41598-026-51730-5 PubMed PMID: 42120472; PubMed Central PMCID: PMC13357736.
16. Calcaterra E, Labellarte D, Hasan A, Mannino D, Scuderi SA, De Luca F, et al. The promising role of cathepsins as biomarkers and therapeutic targets in glioblastoma. Discov Oncol. 2026 May 11;17(1):983. doi:10.1007/s12672-026-05192-w PubMed PMID: 42115563; PubMed Central PMCID: PMC13333578.
17. Zhang Z, Wang Z, Li R, Pei D, Liu J, Qiu Y, et al. Machine learning model on multi-omics data enables risk stratification and identifies molecular heterogeneity and therapeutic targets in glioblastoma. Mol Cancer. 2026 Mar 7;25(1):104. doi:10.1186/s12943-026-02637-2 PubMed PMID: 41795092; PubMed Central PMCID: PMC13081562.
18. Hickman-Chow DA, Luckett PH, Olufawo M, Dierker D, Shimony JS, Leuthardt EC. Predicting progression-free survival in glioblastoma with neuroimaging and machine learning. J Neurooncol. 2026 May 28;178(1):29. doi:10.1007/s11060-026-05650-z PubMed PMID: 42209729; PubMed Central PMCID: PMC13219129.
19. Wang T, Alfaro J, Brennan PM, Rajan A. Multimodal survival analysis of glioblastoma using whole-slide histopathology, gene expression, clinical variables and language-model-derived mutation features. Sci Rep. 2026 Apr 21;16(1):18546. doi:10.1038/s41598-026-48666-1 PubMed PMID: 42014603; PubMed Central PMCID: PMC13269476.
20. Zhao H, Liu K, Guan M, Tang X, Zheng Y, Yu H, et al. Deciphering the glioblastoma microenvironmental landscape with multi-modal radiogenomics to guide prognosis and personalized therapy. J Transl Med. 2026 May 5;24(1):718. doi:10.1186/s12967-026-08226-7 PubMed PMID: 42087207; PubMed Central PMCID: PMC13221763.
21. Man Y, Chen W, Shen G, Zhao X, Lu J, Zhang X. A manganese metabolism-related gene signature for prognosis prediction and immune microenvironment description of glioblastoma. Discov Oncol. 2026 Jan 31;17(1):375. doi:10.1007/s12672-026-04510-6 PubMed PMID: 41619094; PubMed Central PMCID: PMC12949155.
22. Li Y, Ouyang P, Zheng Z, Deng J, Guo A, Wang W, et al. Construction and validation of a prognostic nomogram model integrating machine learning-pathomics and clinical features in IDH-wildtype glioblastoma. J Transl Med. 2025 Jun 3;23(1):618. doi:10.1186/s12967-025-06482-7 PubMed PMID: 40462164; PubMed Central PMCID: PMC12131361.
23. Mu Y, Luo J, Xiong T, Zhang J, Lan J, Zhang J, et al. Development and validation of nomogram model predicting overall survival and cancer specific survival in glioblastoma patients. Discov Oncol. 2025 Apr 18;16(1):562. doi:10.1007/s12672-025-02331-7 PubMed PMID: 40249416; PubMed Central PMCID: PMC12008090.
24. Liu D, Liu Y, Ci Y, Wang C, Ma W. Efficacy and safety of targeted therapies for glioblastoma: a systematic review and network meta-analysis. BMC Cancer. 2026 May 18;26(1):806. doi:10.1186/s12885-026-16190-0 PubMed PMID: 42151869; PubMed Central PMCID: PMC13326477.
25. Sanikommu S, Santos AN, Dawoud B, Komotar RJ, Lu VM. Fluorescence-guided versus non-fluorescence-guided resection in high-grade glioma: a systematic review and meta-analysis of survival outcomes. Neurosurg Rev. 2026 Apr 23;49(1):370. doi:10.1007/s10143-

026-04297-8 PubMed PMID: 42020844; PubMed Central PMCID: PMC13102904.

26. Wang F, Huo A, Li M, Zhao C, Guo Y, Jing Z, et al. Comprehensive analysis of mitochondria-associated genes in glioblastoma via single-cell and bulk RNA sequencing. Discov Oncol. 2025 Oct 27;16(1):1974. doi:10.1007/s12672-025-03795-3 PubMed PMID: 41144088; PubMed Central PMCID: PMC12559552.
27. Du Y, Long X, Li X, Guan F, Gao W, Deng K, et al. Spatial-reprogramming derived GPNMB+ macrophages interact with COL6A3+ fibroblasts to enhance vascular fibrosis in glioblastoma. Genome Med. 2025 Oct 31;17(1):136. doi:10.1186/s13073-025-01553-2 PubMed PMID: 41174767; PubMed Central PMCID: PMC12577258.
28. Huang J, Wang Q, Li Z, Huang X, Wang W, Liu J. Identification and validation of LDHA and SLC16A1 for predicting prognosis and diagnosis in lower-grade glioma. Discov Oncol. 2025 Aug 9;16(1):1511. doi:10.1007/s12672-025-03297-2 PubMed PMID: 40782248; PubMed Central PMCID: PMC12335421.
29. Yang C, Cai L, Xiang X, Zhu X, Liu C, Liao J, et al. PLX3397 attenuated tumor growth and remodeled tumor microenvironment of recurrent glioblastoma. Sci Rep. 2025 Dec 24;16(1):3153. doi:10.1038/s41598-025-32943-6 PubMed PMID: 41444392; PubMed Central PMCID: PMC12830872.
30. Zhang L, Fang X, Yang Z, Zhang M, Ma Y, Li P. KNTC1 and PRC1 define an immunosuppressive microenvironment and poor prognosis in liposarcoma. Eur J Med Res. 2026 Jan 9;31(1):242. doi:10.1186/s40001-025-03703-z PubMed PMID: 41514406; PubMed Central PMCID: PMC12882619.
31. Guo LF, Zhu LC, Yu YF, Lu ZZ, Lin Q, Wu SG. Risk Stratification in T4 or N3 Nasopharyngeal Carcinoma. Cancer Res Treat. 2026 Jul;58(3):728–37. doi:10.4143/crt.2025.573 PubMed PMID: 40808534; PubMed Central PMCID: PMC13382564.
32. Cavaliere S, Fogolari M, Iuliani M, Foderaro S, Cortellini A, Simonetti S, et al. Salivary microbial signature highlighting actinomyces as a predictor of immune-checkpoint inhibitor monotherapy response in advanced non-small cell lung cancer. J Transl Med. 2026 Jan 17;24(1):180. doi:10.1186/s12967-025-07570-4 PubMed PMID: 41547827; PubMed Central PMCID: PMC12895580.
33. Luo T, Xiong J, Yang H, Xiang J, Yi M, Xu B, et al. The prognostic value of neutrophil-to-albumin ratio for dual outcomes in locally advanced esophageal cancer: an integrated analysis based on propensity score matching and Lasso-Cox regression. BMC Cancer. 2026 Apr 29;26(1):681. doi:10.1186/s12885-026-16094-z PubMed PMID: 42056943; PubMed Central PMCID: PMC13217722.
34. Tang XL, He Y, Luan WY, Xu KZ, Zhang Z, Xu ZX, et al. Integrative single-cell and spatial multi-Omics analyses identify NETs-driven colon cancer subtypes with distinct metabolic features and prognostic implications. J Transl Med. 2026 May 9;24(1):849. doi:10.1186/s12967-026-08227-6 PubMed PMID: 42106773; PubMed Central PMCID: PMC13330288.
35. Wei S, Wang T, Li C, Zhu Y, Fan W, Ren H, et al. SOFA-2 score predicts mortality in pneumonia-associated sepsis: a retrospective cohort study. Crit Care. 2026 Apr 27;30(1):302. doi:10.1186/s13054-026-06027-4 PubMed PMID: 42045961; PubMed Central PMCID: PMC13255325.
36. Wang Z, Ma T, Li Y, Sun X, Wang Y, Mao Y, et al. Multidimensional MRI radiomics-based model predicts recurrence risk in triple-negative breast cancer. NPJ Precis Oncol. 2026 May 26. doi:10.1038/s41698-026-01511-0 PubMed PMID: 42191845.
37. Vanichbuncha T, Leetanaporn K, Hanprasertpong J. Impact of the diagnosis-surgery interval on survival outcomes and development of a nomogram for predicting survival in early-stage cervical cancer: a Surveillance, Epidemiology, and End Results database analysis. J Gynecol Oncol. 2026 May 14. doi:10.3802/jgo.2026.37.e108 PubMed PMID: 42172492.
38. Lai JM, Zheng Y, Liang TY, Huang J, Quan YY. Risk factors and nomogram prediction model for prognosis in sepsis with acute kidney injury. BMC Nephrol. 2026 May 23;27(1):437. doi:10.1186/s12882-026-05039-0 PubMed PMID: 42177426.
39. Jiao J, Li F, Liu Q, Zhang H, Yu G, Yu W, et al. Development of a prognostic model for postoperative survival in

appendiceal pseudomyxoma peritonei based on clinical indicators. World J Surg Oncol. 2026 Jun 11;24(1):249. doi:10.1186/s12957-026-04363-x PubMed PMID: 42277799; PubMed Central PMCID: PMC13262463.

40. Shuai Y, Liu Y, Yang X, Chen Z, Wan Q, Zhao J, et al. Prognostic value of stress hyperglycemia ratio in critically Ill patients with acute kidney injury: a machine learning-driven retrospective cohort analysis. Health Inf Sci Syst. 2026 Dec;14(1):46. doi:10.1007/s13755-026-00443-0 PubMed PMID: 41867166; PubMed Central PMCID: PMC13000052.
41. Zhang W, Li G, Gao Y, Liu Z, Zhang B, Wang J. Construction and verification of multimodal model for prognosis in elderly patients with aneurysmal subarachnoid hemorrhage. BMC Neurol. 2026 May 8;26(1):423. doi:10.1186/s12883-026-04940-3 PubMed PMID: 42104264; PubMed Central PMCID: PMC13321583.
42. Kato E, Shigematsu H, Amioka A, Sasada S, Kadoya T, Okada M. Interim Breast MRI for Predicting Pathologic Complete Response After Neoadjuvant Chemotherapy in Early Breast Cancer. Eur J Breast Health. 2026 Jun 17;22(3):350–7. doi:10.4274/ejbh.galenos.2026.2026-1-9 PubMed PMID: 42317098; PubMed Central PMCID: PMC13280616.
43. Yu T, Fang Z, Cheng Y, Zhou Y, Ji Y, Liu C. A novel palmitoylation-based molecular signature reveals COX6A1 as a key regulator in metabolic dysfunction-associated steatotic liver disease. J Transl Med. 2025 Nov 3;23(1):1212. doi:10.1186/s12967-025-07253-0 PubMed PMID: 41185013; PubMed Central PMCID: PMC12581582.
44. Milton R, McLaughlin AP, Mariani N, Kose M, Zajkowska Z, Lombardo G, et al. Peripheral blood transcriptomic differences predict depression status in individuals undergoing bariatric surgery. Brain Behav Immun - Health. 2025 Aug;47:101029. doi:10.1016/j.bbih.2025.101029 PubMed PMID: 40585350; PubMed Central PMCID: PMC12205765.
45. Tabassum N, Khurshid Y, Syed B, Ahmed A, Muhammad S, Imad R, et al. Proteome-based identification and validation of NXPE3 in childhood acute lymphoblastic leukaemia. Pak J Pharm Sci. 2026 Jun;39(6):1854–61. doi:10.36721/PJPS.2026.39.6.176.1 PubMed PMID: 42001290.
46. Gan X, Liang M, Shadekejiang H, Hong H, Lu J, Li X, et al. Integrated network pharmacology and bioinformatics analysis reveals MME as key target of Notoginsenoside R1 in diabetic nephropathy. BMC Complement Med Ther. 2026 Feb 1;26(1):80. doi:10.1186/s12906-026-05272-y PubMed PMID: 41622156; PubMed Central PMCID: PMC12955068.
47. Hair JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S. Overview of R and RStudio. In: Hair Jr. JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S, editors. Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook [Internet]. Cham: Springer International Publishing; 2021 [cited 2026 Jul 28]. p. 31–47. Available from: https://doi.org/10.1007/978-3-030-80519-7_2 doi:10.1007/978-3-030-80519-7_2