Current Science
CS 5 (3), 2959-2979 (2025)
Received 24 March 2025 | **Revised 18 April** 2025 | **Accepted 22 May**. 2025 | **Online Available 13 June** 2025
https://doi.org/ 10.52845/CS/2025-5-3-68

Current Science

Current Science

OPEN ACCESS

ISSN (O) 2795-8639

Original Article

Epigenetic Remodeling in the Tumor Microenvironment: A Novel Prognostic Signature for Glioblastoma Immunotherapy and Immune Evasion

Shanshan Sun[1], Jia Wang[2]; Hao Jing[3], Peng Gao[1*]

[1]Department of Oncology, Jiaozhou Central Hospital of Qingdao, Qingdao, Shandong 266300, China

[2]Department of Oncology, Huantai County People's Hospital, Zibo, Shandong 256400, China

[3]Department of Vascular Surgery, Peking University People's Hospital, Qingdao Hospital, Qingdao, Shandong 266000, China

[*]Corresponding Author: Peng Gao

Abstract:

Background: Glioblastoma multiforme (GBM) is the most aggressive primary brain tumor, characterized by high heterogeneity and poor prognosis. Despite advances in understanding its genetic and epigenetic alterations, prognostic models based on epigenetic modifications remain underexplored. This study aimed to develop a robust prognostic signature for GBM by integrating multi-omics data to unravel the molecular mechanisms underlying tumor progression and therapeutic resistance.

Methods: We analyzed single-cell transcriptome data (GSE182109) from 44 glioma samples, identifying seven distinct cell populations and quantifying epigenetic scores via ssGSEA based on 801 epigenetics-related genes from the EpiFactors database. Weighted gene co-expression network analysis (WGCNA) was performed on TCGA-GBM data to identify epigenetic modification-related genes. A Random Survival Forest (RSF) model was developed and validated using TCGA (n=153) and two CGGA cohorts (n=693 and n=325). Functional enrichment, pathway analysis, and tumor microenvironment characterization were conducted using GSEA, IOBR, and CIBERSORTx. Drug sensitivity analysis was performed, and pan-cancer validation was conducted across 29 TCGA cancer types.

Results: Single-cell analysis revealed elevated epigenetic scores in recurrent GBM (rGBM) compared to newly diagnosed GBM (ndGBM), particularly in myeloid cells ($P<0.05$). WGCNA identified 533 epigenetic modification-related genes, and the RSF model demonstrated superior prognostic performance with C-indices of 0.918 (TCGA), 0.605 (CGGA693), and 0.652 (CGGA325). High-risk patients showed enrichment in KRAS signaling (HR=2.31) and reactive oxygen species pathways (HR=1.89). Tumor microenvironment analysis revealed higher T-cell infiltration ($P<0.001$) and tumor neo-antigen burden ($P=0.00497$) in low-risk groups. Drug sensitivity analysis identified differential responses to temozolomide ($P=0.0365$) and paclitaxel ($P=0.0262$) between risk groups. Pan-cancer validation showed prognostic significance in Kidney Renal Clear Cell Carcinoma (HR>1, $P<0.001$).

Conclusion: This study developed a novel epigenetic signature that can effectively stratifies GBM patients into distinct prognostic groups, providing insights into tumor progression and treatment response. Further validation studies are needed to translate these findings into clinical practice.

Keywords: Glioblastoma multiforme; Epigenetic modification; Random survival forest; Single-cell analysis; Tumor microenvironment; Drug sensitivity

1. Introduction

Glioma is the most common primary malignant tumor of the central nervous system in adults. Based on pathological grade by the World Health

Organization (WHO), gliomas can be divided into glioblastoma multiforme (GBM) and lower-grade glioma (LGG)(1). GBM, a grade IV malignancy, is a common malignant brain tumor found mainly in the elderly aged 75 to 84 years and is more prevalent in males(2). Besides, it's also the most aggressive primary malignant tumor of the central nervous system (CNS), characterized by its extreme biological aggressiveness and capacity to extensively infiltrate surrounding brain tissue(3). Due to the highly invasive nature of GBM, complete surgical resection remains challenging.

There are multiple hypotheses regarding the origin of GBM, with the prevailing theories suggesting that it arises from neural stem cells or gliocytes with stem-like properties, or from both, as proposed by the "double origin" hypothesis. It is widely recognized that the development of some GBMs involves mutations in genes such as IDH, EGFR, TP53, and PTEN(4-6), as well as alterations in signaling pathways including RTK , Notch, and WNT(7-9), leading to aberrant cellular behavior. In addition to traditional genetic alterations, epigenetic modifications, such as DNA methylation(10), have also been implicated in GBM pathogenesis. Furthermore, studies have shown that differences in histone modifications are associated with cellular and tissue atypia and differentiation status(11). Although epigenetic modifications play a significant role in the pathogenesis of GBM, prognostic models based on epigenetic genes remain unclear, and their clinical application and predictive value require further investigation.

Previous studies have attempted to predict the prognosis and therapeutic response of GBM patients using various biomarkers, including cfDNA(12) and protein molecules in blood, immune-related lncRNAs(13) in tumor tissue, and immune cell infiltration in the tumor microenvironment(14). However, optimizing treatment strategies and accurately predicting drug efficacy for different GBM subtypes remain significant challenges. In this study, we employed a multi-omics analysis based on pathological and omics data from GBM patients in different public databases. Using machine learning approaches, we aimed to identify novel therapeutic strategies and mechanisms of drug resistance, providing new insights into the treatment of GBM.

2. Materials and Methods

2.1 Data Source

The data utilized in this study were sourced from multiple publicly available databases to ensure comprehensive analysis and robust validation. Single-cell transcriptome data (GSE182109), comprising 44 samples including LGG, newly diagnosed glioblastoma (ndGBM), and recurrent glioblastoma (rGBM), were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) for in-depth single-cell analysis(15). For the identification of differentially expressed genes in glioblastoma, we integrated data from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/)(16) and the Genotype-Tissue Expression (GTEx) project (https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project)(17). Specifically, the TCGA-GBM dataset included 153 GBM samples and 5 adjacent normal tissue samples, while the GTEx dataset provided 2,642 normal control samples. These datasets were merged after applying batch effect correction to ensure data consistency. Furthermore, the TCGA-GBM dataset served as the primary training set for our analyses. To validate our findings, we employed two independent cohorts from the Chinese Glioma Genome Atlas (CGGA) database (https://www.cgga.org.cn/)(18), which included 693 and 325 GBM samples, respectively, along with detailed clinical and survival information. These external validation sets were instrumental in assessing the generalizability and robustness of our results.

2.2 Single-Cell Analysis

2.2.1 Preprocessing

Single-cell transcriptome data (GSE182109) were processed using the Seurat package (v5.1.0) (https://CRAN.R-project.org/package=Seurat) in R (v4.4.0)(19). Raw count matrices for each sample were loaded using the Read10X function, and Seurat objects were created with a minimum of 3 cells and 250 features per cell. Quality control metrics, including the percentage of mitochondrial and ribosomal genes, were calculated using the PercentageFeatureSet function. Cells were filtered based on stringent criteria: those with fewer than 200 or more than 6,000 genes, more than 20,000 unique molecular

identifiers (UMIs), or more than 10% mitochondrial gene content were excluded. Data normalization was performed using the NormalizeData function, and highly variable genes were identified with the FindVariableFeatures function. The data were then scaled using the ScaleData function. Principal component analysis (PCA) was conducted, and the top 30 principal components were selected for downstream analysis. Batch effects were corrected using Harmony (v1.2.0)(20), and uniform manifold approximation and projection (UMAP) was performed for dimensionality reduction(21).

2.2.2 Analysis

Clustering was performed using the FindNeighbors and FindClusters functions with a resolution of 0.5. Differentially expressed genes (DEGs) between clusters were identified using the FindAllMarkers function, applying a minimum expression fraction of 25% and a log-fold change threshold of 0.5. Cluster annotation was conducted using the scMayoMap package (v0.2.0) with the brain tissue database(22). Cell types were manually annotated based on marker gene expression(23), including *PTPRC*, *ITGAM*, *CD68* (myeloid cells), *SOX2*, *OLIG1*, *GFAP* (glioma cells), *CD3E*, *CD4*, *CD8A* (T cells), *CD79A*, *CD19* (B cells), *ACTA2*, *PDGFRB* (pericytes), *EMCN*, *MYCT1* (endothelial cells), and *OLIG2*, *MBP* (oligodendrocytes). Epigenetic scores were calculated using single-sample gene set enrichment analysis (ssGSEA) with the GSVA package (v1.42.0)(24), based on 801 epigenetics-related genes curated from the EpiFactors database (https://epifactors.autosome.org/)(25). Statistical comparisons of epigenetic scores across cell types and conditions were performed using the t-test with Benjamini-Hochberg correction for multiple testing. Differential gene expression analysis between high and low epigenetic score groups was conducted using the FindMarkers function, applying a minimum expression fraction of 25% and a log-fold change threshold of 0.25. Significant DEGs were identified and exported for further analysis.

2.3 Weighted Gene Co-expression Network Analysis (WGCNA)

To identify co-expression modules associated with epigenetic regulation, we performed a weighted gene co-expression network analysis (WGCNA) using the WGCNA package (v1.72.5) in R(26). The analysis was conducted on the TCGA-GBM transcriptomic dataset, normalized as transcripts per million (TPM) values. A total of 801 epigenetics-related genes were used as the input gene set also. The TPM data were log2-transformed and filtered to retain genes with a standard deviation greater than zero. Sample clustering was performed to detect and remove outliers using hierarchical clustering with a threshold of 60,000. The soft thresholding power was determined by evaluating the scale-free topology model fit and mean connectivity, and a fit power was selected for network construction. A signed network was built using the blockwiseModules function, with a minimum module size of 50 and a merge cut height of 0.15. Modules were identified and assigned colors for visualization.

Module-trait relationships were assessed by calculating the correlation between module eigengenes (MEs) and the epigenetic scores derived from ssGSEA. Significance was determined using Student's asymptotic p-value, and heatmaps were generated to visualize the associations. Gene significance (GS) for epigenetic scores and module membership (MM) were computed for each gene, and scatterplots were used to evaluate the relationship between MM and GS for key modules. The feature module, which exhibited the strongest association with epigenetic scores, was selected for further analysis. Genes within this module were extracted and subjected to functional enrichment analysis to identify potential biological pathways.

2.3 Machine Learning for Survival Analysis

To construct and validate prognostic models, we employed a comprehensive machine learning framework using the TCGA-GBM dataset as the training set and two independent cohorts from the Chinese Glioma Genome Atlas (CGGA) as validation sets. The analysis included 10 machine learning algorithms: Random Survival Forest (RSF), CoxBoost, Elastic Net (Enet), Gradient Boosting Machine (GBM), Lasso, Partial Least Squares Regression for Cox (plsRcox), Ridge Regression, Stepwise Cox Regression (StepCox), Supervised Principal Components (SuperPC), and Survival Support Vector Machine (survival-SVM)(23, 27). Each algorithm was implemented

with appropriate hyperparameter tuning and cross-validation to ensure robustness.

For RSF, the model was trained using 1,000 trees with a node size of 5 and the log-rank splitting rule. CoxBoost was optimized using a penalty parameter determined by 10-fold cross-validation. Elastic Net models were evaluated with alpha values ranging from 0.1 to 0.9 to balance L1 and L2 regularization. GBM was trained with 10,000 trees, a learning rate of 0.001, and 10-fold cross-validation to determine the optimal number of trees. Lasso and Ridge models were implemented using 10-fold cross-validation to select the best lambda value. plsRcox was applied with 10 components, and StepCox was evaluated in three directions: forward, backward, and both. SuperPC was used with 10-fold cross-validation and 20 thresholds, while survival-SVM was implemented with a gamma value of 1.

Model performance was assessed using the concordance index (C-index) across the training and validation datasets. The C-index reflects the model's ability to correctly rank survival times. The results were visualized using heatmaps and bar plots to compare the performance of different algorithms across datasets. The top-performing models were further analyzed to identify key prognostic features and their biological relevance.

2.4 External Validation

To validate the prognostic model, the trained best performance model was applied to two independent external validation cohorts from the CGGA: CGGA325 and CGGA693. Risk scores were calculated for each patient in the validation cohorts using the best performance model, and patients were stratified into high-risk and low-risk groups based on the median risk score. Kaplan-Meier survival analysis was performed to compare survival outcomes between the risk groups, with statistical significance assessed using the log-rank test. Time-dependent receiver operating characteristic (ROC) curves (28)were constructed to evaluate the model's predictive accuracy at 1, 3, and 5 years. The area under the curve (AUC) and 95% confidence intervals (CI) were calculated to quantify the model’s performance.

2.5 Functional Enrichment and Pathway Analysis

To explore the biological mechanisms underlying the prognostic model, we performed functional enrichment and pathway analysis. DEGs between high-risk and low-risk groups were identified using the limma package(29), with a significance threshold of adjusted p-value < 0.05. GSEA was conducted using the Hallmark gene sets from the Molecular Signatures Database (MSigDB)(30, 31) to identify enriched pathways. Kaplan-Meier survival analysis was performed for key pathways to assess their prognostic value, with patients stratified into high and low expression groups based on the median gene set variation analysis (GSVA) score. Hazard ratios (HR) and 95% CI were calculated using Cox proportional hazards models to evaluate the association between pathway activity and survival outcomes.

2.6 Nomogram Construction and Validation

To facilitate clinical application, a nomogram was constructed based on the multivariate Cox proportional hazards model incorporating age, gender, tumor type, and the standardized risk score. The nomogram was developed using the rms package in R, with survival probabilities predicted at 1, 3, and 5 years. Calibration curves were generated to assess the agreement between predicted and observed survival outcomes, using 500 bootstrap resamples to ensure reliability. The discriminative ability of the nomogram was evaluated using the C-index and time-dependent C-index analysis. Decision curve analysis (DCA) was performed to compare the clinical utility of the nomogram against individual predictors, including age, gender, tumor type, and the risk score. The C-index was calculated using the pec package(32), and DCA was conducted using the rmda package (v1.6) (https://CRAN.R-project.org/package=rmda).

2.7 Tumor Microenvironment and Immune Analysis

2.7.1 Immune Cell Infiltration Analysis

To characterize the tumor microenvironment (TME), we performed immune cell infiltration analysis using the IOBR package (v0.99.0) in R(33). Seven deconvolution methods were applied to estimate immune cell proportions: CIBERSORT (https://cibersortx.stanford.edu/)(34), EPIC (https://rdrr.io/github/ GfellerLab/ EPIC/man/EPIC.package.html)(35), MCP-counter(36), xCell(37), ESTIMATE(38), TIMER(39), and quanTIseq(40). Additionally,

immune cell infiltration scores were calculated using the ssGSEA approach based on predefined immune gene signatures. The results from these methods were integrated into a comprehensive TME dataset.

For visualization, correlation plots were generated to compare immune cell infiltration patterns between high-risk and low-risk groups. Kaplan-Meier survival analysis was performed for key immune cell types, and statistical significance was assessed using the log-rank test. Boxplots and scatterplots were used to visualize differences in immune cell infiltration and their correlation with risk scores.

2.7.2 Tumor Mutational Burden and Neoantigen Analysis

Tumor mutational burden (TMB) and tumor neoantigen burden (TNB) were calculated using the maftools package (v2.20.0)(41) and TCIA neoantigen data(42), respectively. TMB and TNB values were log2-transformed and compared between high-risk and low-risk groups using the Wilcoxon rank-sum test.

To evaluate the combined prognostic value of risk scores, TMB, and TNB, patients were stratified into subgroups based on risk score and TMB/TNB status. Kaplan-Meier survival curves were generated for these subgroups, and the log-rank test was used to assess differences in survival outcomes.

2.8 Drug Sensitivity Analysis

To evaluate the potential therapeutic implications of the prognostic model, drug sensitivity analysis was performed using the pRRophetic package (v0.5) in R(43). Gene expression data from the TCGA-GBM cohort were used to predict drug sensitivity scores (IC50 values) for a panel of 138 drugs from the Cancer Genome Project (CGP) dataset (http://www.sanger.ac.uk/genetics/CGP/).

Patients were stratified into high-risk and low-risk groups based on their risk scores, and differences in drug sensitivity between the groups were assessed using the Wilcoxon rank-sum test. Statistical significance was defined as $p < 0.05$. Box plots were generated to visualize the distribution of IC50 values for each drug, with p-values annotated on the plots.

2.9 Pan-Cancer Analysis

To evaluate the generalizability of the prognostic model across different cancer types, we performed a pan-cancer analysis using data from TCGA. The analysis included 29 cancer types, covering major organ systems such as breast and gynecologic cancers, respiratory cancers, digestive cancers, urinary cancers, and others. Gene expression data and clinical survival data were downloaded for each cancer type using the UCSCXenaTools package (v1.6.0)(44).

The best performance model trained on the TCGA-GBM dataset was applied to each cancer type to predict risk scores. Patients were stratified into high-risk and low-risk groups based on the median risk score, and survival differences were assessed using Cox proportional hazards models and Kaplan-Meier analysis. HR and 95% CI were calculated for each cancer type, and statistical significance was defined as $p < 0.05$.

Results were visualized using a forest plot, with cancer types grouped by organ system. The plot included HR, p-values, and sample sizes for each cancer type. Kaplan-Meier survival curves were generated for significant associations.

3. Results

The research process is shown in Figure 1.

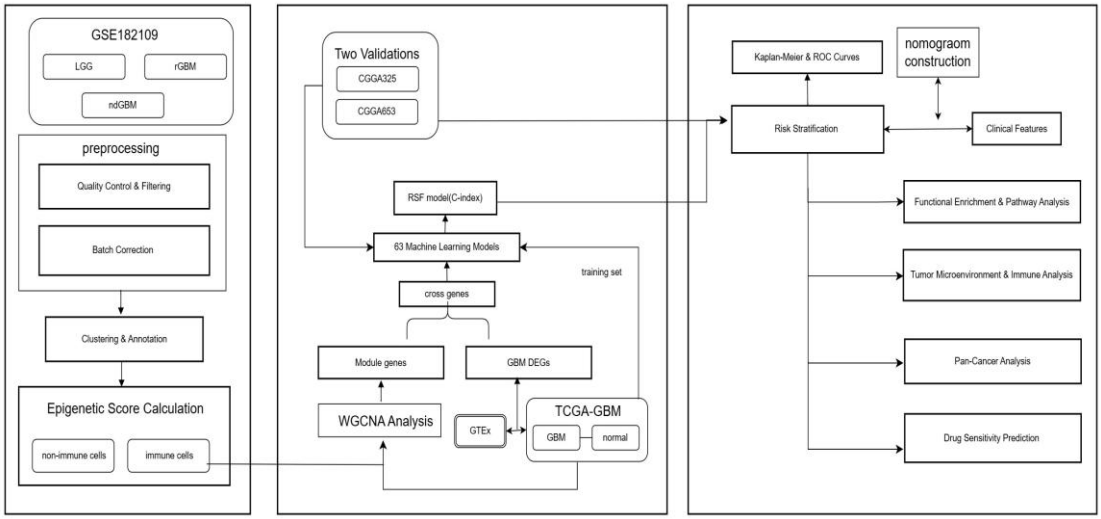

Figure 1. Study Flowchart

3.1 Single-Cell Analysis Uncovers Cellular Diversity and Epigenetic Landscapes in the Glioma Microenvironment

Our comprehensive single-cell analysis revealed intricate cellular heterogeneity within the glioma microenvironment. Through t-SNE dimensional reduction analysis (Figure 2A), we identified and visualized seven distinct cellular populations: Glioma cells, Myeloid cells, T cells, Oligodendrocytes, B cells, Pericytes, and Endothelial cells. These populations exhibited clear spatial segregation in the t-SNE plot, reflecting their unique transcriptional signatures. The characteristic gene expression profiles of each cell type were further illustrated through a dot plot (Figure 2B), providing molecular evidence for cell type identification.

To investigate the epigenetic landscape across these cellular populations, we implemented ssGSEA to quantify epigenetic modification levels. The resulting epigenetic scores, visualized through t-SNE dimensionality reduction, revealed that Glioma cells consistently exhibited the highest epigenetic scores among all cell types (Figure 2C). This observation prompted us to explore potential functional differences in epigenetic regulation between immune and non-immune cell populations.

Our comparative analysis of epigenetic scores across different glioma grades (Figure 2D) revealed a progressive increase from ndGBM to LGG and rGBM. This gradient pattern suggests a potential correlation between epigenetic modification levels and glioma progression. Notably, LGG consistently displayed intermediate epigenetic scores between ndGBM and rGBM, indicating a distinct molecular landscape that differentiates LGG from its more malignant counterparts.

Given the predominance of myeloid cells in the immune cell population, we conducted focused analysis on their epigenetic profiles. Statistical comparisons revealed significant differences in epigenetic scores between ndGBM and rGBM samples (Figure 2E), highlighting the potential role of myeloid cell epigenetic regulation in glioma progression. Subsequent subgroup analysis of myeloid cells stratified by epigenetic scores (Figure 2F-G) unveiled substantial heterogeneity in cellular composition between high and low epigenetic score groups. While the C1 subgroup maintained relatively consistent proportions across groups, other subgroups displayed distinct distribution patterns, suggesting differential responses to epigenetic modifications.

This observed heterogeneity motivated us to perform differential expression analysis between high and low epigenetic score groups, aiming to identify potential molecular drivers underlying the observed phenotypic differences.

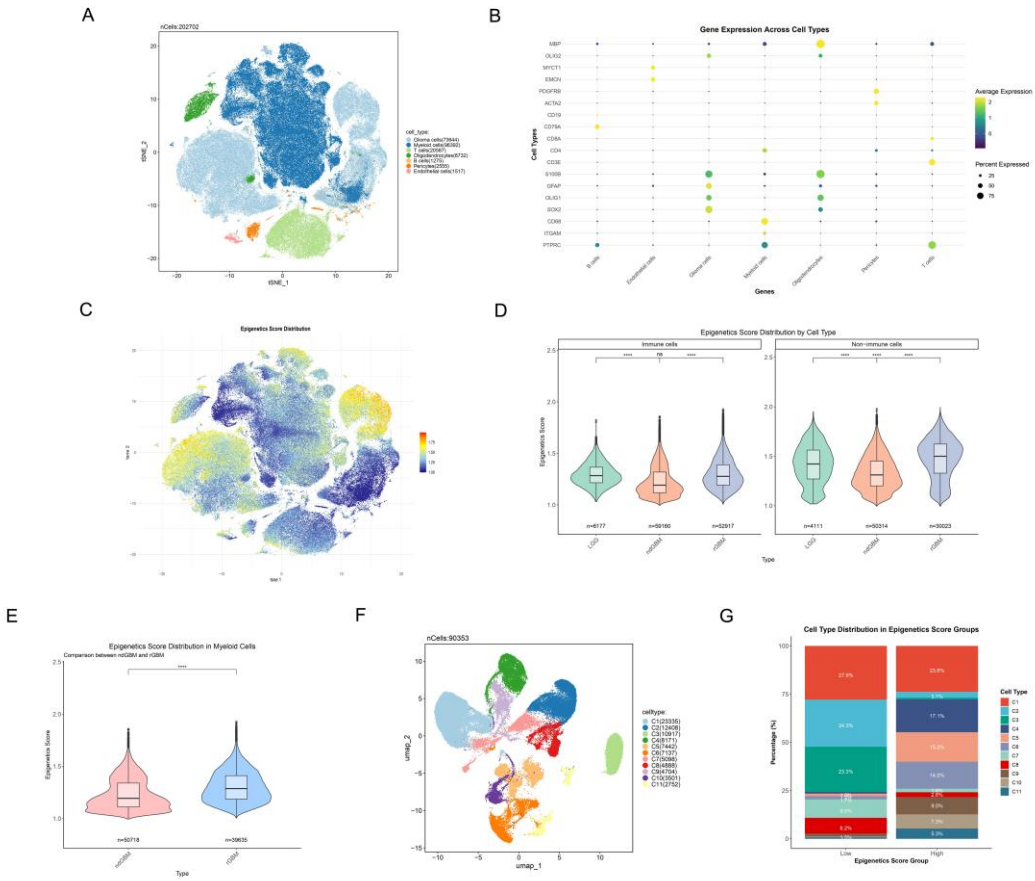

Figure 2. Single-cell analysis of cellular heterogeneity and epigenetic patterns in the glioma microenvironment.

(A) t-SNE visualization of cellular clusters in the glioma microenvironment. **(B)** Dot plot showing marker gene expression across cell types. **(C)** t-SNE plot of epigenetic score distribution across cell populations. **(D)** Epigenetic score comparison in immune and non-immune cells across LGG, ndGBM, and rGBM samples. **(E)** Boxplot of epigenetic scores in myeloid cells from ndGBM and rGBM samples. **(F)** UMAP dimensionality reduction plot showing spatial distribution and heterogeneity of myeloid cell subtypes. **(G)** Cell abundance plots comparing myeloid cell subtypes between high and low epigenetic score groups.

3.2 Systematic Identification of Epigenetic Modification-Related Genes in GBM

To systematically identify genes associated with epigenetic modifications in GBM, we performed a comprehensive WGCNA by using the differential expression analysis outcomes of epigenetic score groups. The heatmap of WGCNA analysis (Figure 3A) illustrates the distribution of epigenetic scores across samples, highlighting the molecular heterogeneity and characteristic variations in GBM. Sample clustering dendrogram (Figure 3B) and module-trait heatmap (Figure 3C) revealed that the turquoise module exhibited the strongest positive correlation with epigenetic modification (cor = 0.86, P = 3e-45). Further analysis of module membership demonstrated a robust positive correlation between GS and MM within the turquoise module (cor = 0.82, P < 1e-200, Figure 3D), indicating its central role in the epigenetic regulatory network.

To identify DEGs specific to GBM, we conducted EnhancedVolcano plot analysis, which highlighted significant up- and downregulated genes between normal and GBM samples (Figure 3E). The top 50 most significantly altered genes were visualized in a circular plot (Figure 3F), providing a comprehensive overview of their expression changes and statistical significance. Finally, we cross-referenced 7,785 GBM DEGs with the 1,061 genes from the turquoise module, identifying 533 genes specifically associated with epigenetic modifications in GBM (Figure 3G).

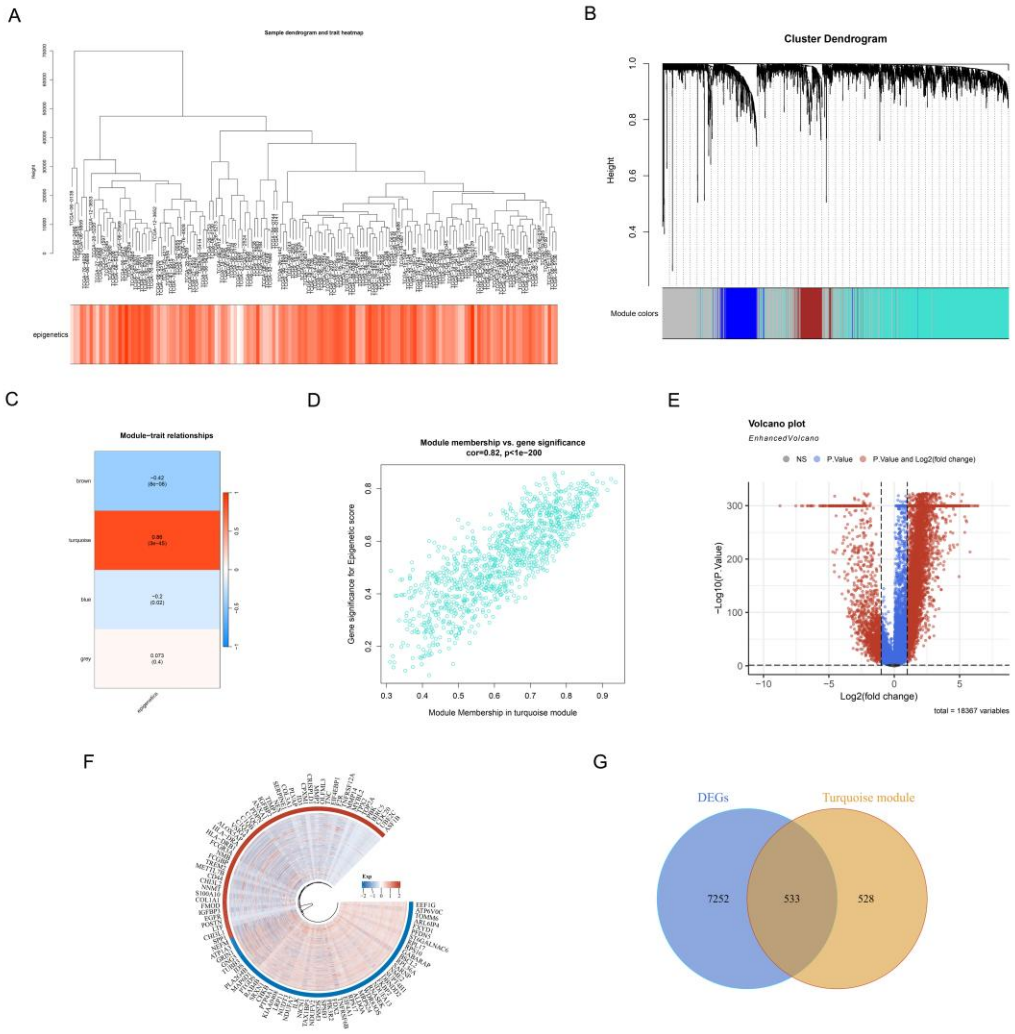

Figure 3. Systematic identification of epigenetic modification-related genes in GBM.
(A) Heatmap showing hierarchical clustering of TCGA-GBM samples, with epigenetic scores represented by color intensity. (B) Cluster dendrogram from WGCNA analysis, with colored bands indicating functional modules identified by dynamic tree-cutting. (C) Module-trait heatmap displaying Pearson correlation coefficients between co-expression modules and epigenetic scores. (D) Scatter plot of GS versus MM for the turquoise module. (E) EnhancedVolcano plot highlighting DEGs between normal and GBM samples. (F) Circular plot visualizing the top 50 most significantly up- and downregulated genes. (G) Venn diagram showing the overlap between GBM DEGs and turquoise module genes.

3.3 Development and Multi-center Validation of Prognostic Models

To establish a robust prognostic signature based on epigenetic modification-related genes, we systematically evaluated 63 machine learning algorithms using these genes as feature inputs. The RSF model emerged as the top-performing algorithm, demonstrating superior predictive accuracy with C-indices of 0.918 in the TCGA dataset, 0.605 in the CGGA693 cohort, and 0.652 in the CGGA325 cohort, achieving an average C-index of 0.725 (Figure 4A). Feature importance analysis of the RSF model identified the top 20 genes with the most significant contributions to prognostic prediction (Figure 4B).

Using these key genes, patients were stratified into high- and low-risk groups based on risk scores calculated by the RSF model. Kaplan-Meier survival analysis revealed significant differences in survival outcomes between these groups (log-rank test, P = 2.15e-36, Figure 4C), validating the model's prognostic discrimination ability. Time-dependent ROC analysis further demonstrated the model's robust predictive performance, with AUC values of 0.981, 0.953, and 0.855 for 1-, 3-, and 5-year survival predictions, respectively (Figure 4D).

To assess the generalizability of the model, we validated it in two independent cohorts. In the CGGA325 cohort, the model significantly stratified patients into high- and low-risk groups (log-rank test, P = 1.24e-8, Figure 4E), with AUC values of 0.660, 0.739, and 0.721 for 1-, 3-, and 5-year predictions, respectively (Figure 4F). Similarly, in the CGGA693 cohort, the model effectively distinguished survival outcomes (log-rank test, P = 1.87e-3, Figure 4G), with AUC values of 0.635, 0.614, and 0.574 for the corresponding time points (Figure 4H). These multi-center validation results underscore the model's robust prognostic capability and broad clinical applicability.

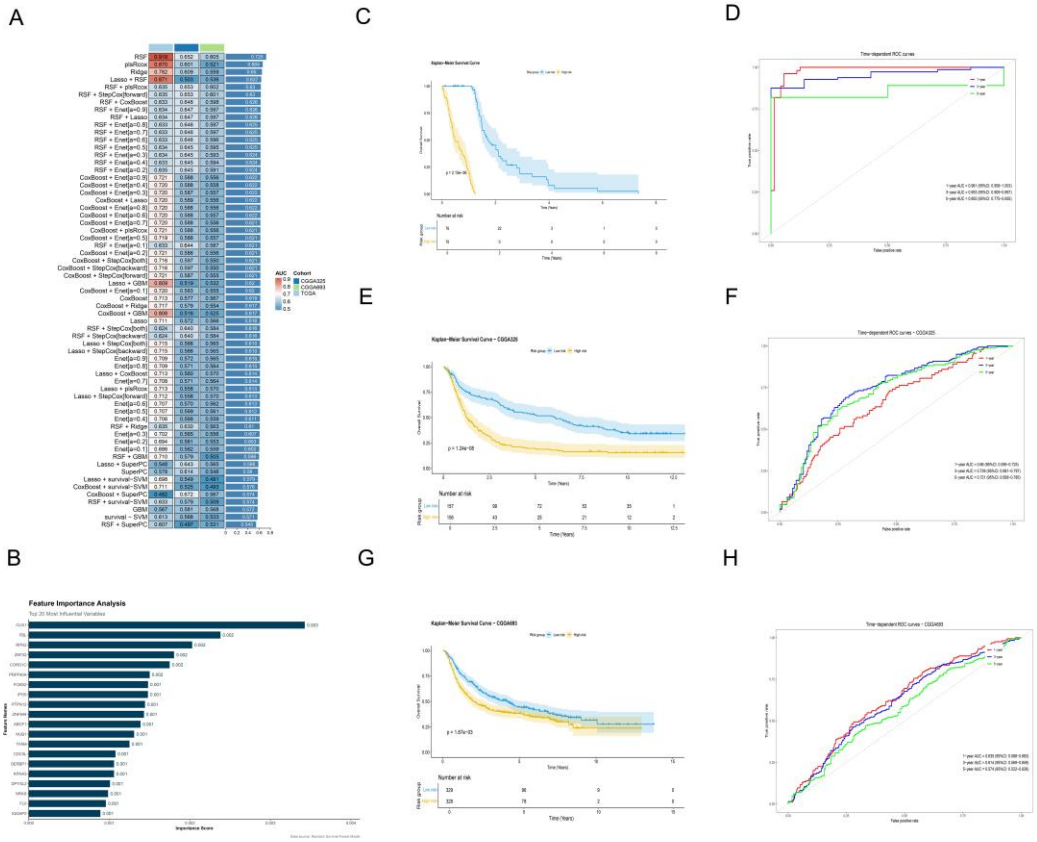

Figure 4. Development and validation of the RSF prognostic prediction model.
(A) Heatmap and bar plots comparing the performance of 63 machine learning algorithms across the TCGA training set and two independent validation sets (CGGA693 and CGGA325). Color intensity represents C-index values. (B) Feature importance ranking plot of the top 20 genes identified by the RSF model. (C) Kaplan-Meier survival curves for the TCGA cohort, stratified by high- and low-risk groups. (D) Time-dependent ROC curves for the TCGA cohort, showing predictive performance at 1-, 3-, and 5-year time points. (E, G) Kaplan-Meier survival curves for the CGGA325 and CGGA693 validation cohorts, respectively. (F, H) Time-dependent ROC curves for the CGGA325 and CGGA693 validation cohorts, respectively.

3.4 Functional Enrichment and Pathway Analysis of RSF Risk Stratification

To elucidate the molecular mechanisms underlying the prognostic stratification by the RSF model, we conducted a comprehensive functional enrichment analysis comparing high-risk and low-risk groups. GSEA revealed significant enrichment of several cancer-related Hallmark pathways in the high-risk group, including angiogenesis, coagulation, epithelial-mesenchymal transition, IL2 STAT5 signaling, KRAS signaling down, and reactive oxygen species pathways (Figure 5A).

To assess the clinical relevance of these pathways, we focused on the top three significantly enriched pathways based on their GSVA scores: KRAS signaling down, UV response up, and reactive

oxygen species pathways. Kaplan-Meier survival analysis demonstrated that patients with high expression levels of these pathways (stratified by median GSVA scores) exhibited significantly poorer overall survival (OS) compared to those with low expression levels (all P < 0.005; Figures 5B-D).

HR and 95% CI were calculated to quantify the associations between these pathways and clinical outcomes. The KRAS signaling down pathway, categorized as a signaling pathway, showed a strong association with poor prognosis (Figure 5E). Similarly, the UV response up and reactive oxygen species pathways, classified as stress responses, also demonstrated significant links to adverse clinical outcomes. These findings highlight the potential role of these pathways in driving the poor prognosis observed in high-risk patients.

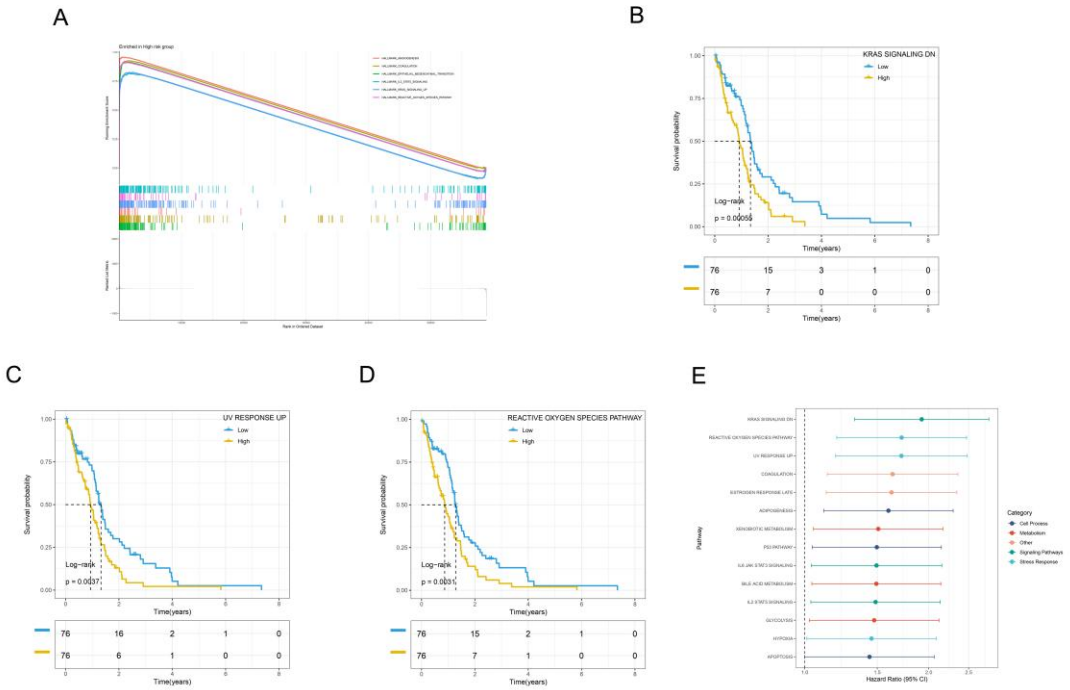

Figure 5. Functional enrichment and pathway analysis of RSF risk stratification.
(A) GSEA waterfall plot showing six critical pathways significantly enriched in the high-risk group. Enrichment profiles are displayed in the upper panel, with a corresponding gene expression heatmap in the lower section. (B) Kaplan-Meier survival curve for the KRAS signaling down pathway, stratified by high and low expression groups based on median GSVA scores. (C) Kaplan-Meier survival curve for the UV response up pathway. (D) Kaplan-Meier survival curve for the reactive oxygen species pathway. (E) Forest plot of pathway HRs, illustrating the prognostic influence of pathways through HR and 95% CI.

3.5 Development and Validation of a Nomogram Integrating Clinical Features

To enhance clinical utility, we systematically evaluated the impact of age, gender, tumor type, and riskscore (derived from RSF risk stratification) on OS. Univariate Cox regression analysis identified age and riskscore as significant prognostic factors for OS (p < 0.001, Figure 6A). Multivariate Cox regression analysis further confirmed that riskscore remained the only statistically significant predictor (p < 0.001, Figure 6B).

Based on these findings and clinical relevance, we integrated age, gender, tumor type, and riskscore into a comprehensive prognostic nomogram (Figure 6E). The nomogram's predictive accuracy was validated using calibration curves, which demonstrated close alignment between predicted and observed 1-year OS (Figure 6C), confirming the model's reliability. Decision curve analysis revealed that within a specific high-risk threshold range, the nomogram-based decision strategy, which aligned with riskscore-based decisions, achieved higher standardized net benefits compared to using individual clinical features

alone (Figure 6D). These results suggest that riskscore can effectively substitute for the nomogram, a conclusion further supported by the time-dependent C-index analysis (Figure 6F).

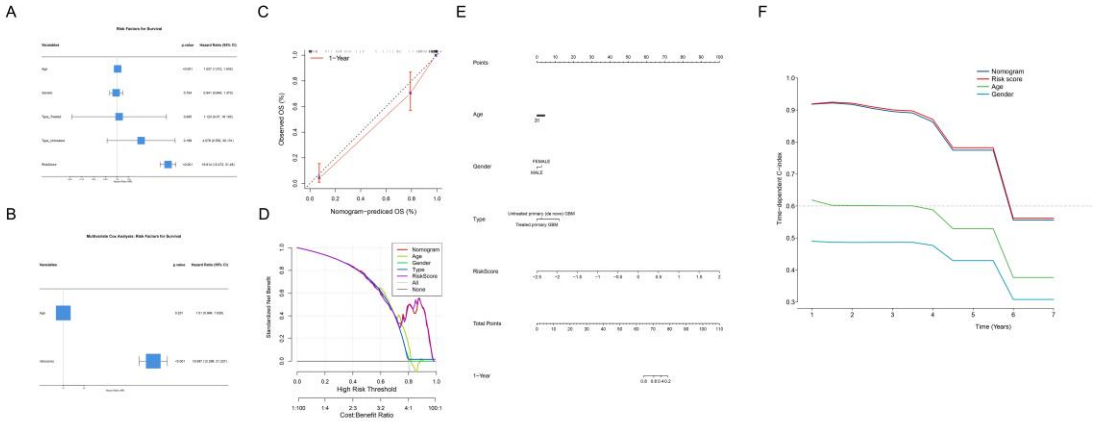

Figure 6. Development and performance evaluation of the integrated prognostic prediction model. (A) Forest plot of univariate Cox regression analysis showing HR and 95% CI for clinical features and riskscore. (B) Forest plot of multivariate Cox regression analysis identifying independent prognostic factors. (C) Calibration curve of the nomogram for 1-year OS. (D) Decision curve analysis comparing the standardized net benefit of the nomogram and individual clinical characteristics. (E) Nomogram integrating riskscore and clinical characteristics. (F) Time-dependent C-index curve illustrating the predictive performance of different prognostic factors across follow-up years.

3.6 Tumor Microenvironment Characteristic Analysis

We conducted a detailed analysis of the tumor microenvironment to compare immune cell infiltration between high-risk and low-risk groups. Several immune cell populations, including T_cell_probe_Jeschke_et_al, CD8+_Tem_xCell, CD4_c10_Tn_LEF1_ANKRD55, CD4_c4_Tstr, CD4_c6_Tn_FHIT, CD8_c4_Tstr, Macrophages_M2_xCell, Macrophages_Rooney_et_al, and Neutrophils_Rooney_et_al, exhibited significantly higher expression in the high-risk group compared to the low-risk group (all $P < 0.05$, Figure 7A). Conversely, CD4+_naive_T-cells_xCell and CD8+_naive_T-cells_xCell were significantly more abundant in the low-risk group (all $P < 0.05$).

Further evaluation of immune suppression (Figure 7B) and immune exclusion (Figure 7C) revealed subtle differences in molecular marker distribution between the two risk groups, although these differences were not statistically significant. Analysis of immunotherapy biomarkers (io biomarkers) showed that most markers did not differ significantly between the groups (Figure 7D). However, IPS_IPS and APM were significantly upregulated in the low-risk group (both $P < 0.05$), suggesting potential immunological distinctions associated with risk stratification.

Assessment of TMB and tumor neoantigen burden (TNB) revealed that TMB peaked at Log2(TMB+1) ≈ 5 in both groups, with no significant difference ($P = 0.177$, Figure 7E). In contrast, TNB peaked at Log2(TNB+1) ≈ 5.5 and was significantly higher in the low-risk group ($P = 0.00497$, Figure 7F), indicating that low-risk patients may exhibit a more active immune response or greater potential for immunotherapy efficacy.

Further TME analysis revealed among different immune cells (Figure 8) that Macrophages_Rooney_et_al exhibited the strongest positive correlation with the risk score ($R = 0.32$, Figure 7H and Figure 9), implicating their potential role in tumor progression and prognosis in the high-risk group. Survival analysis further demonstrated that GBM patients with low consensus machine learning-driven signature (CMLS) combined with high TMB, high TNB, or high macrophage infiltration exhibited improved prognosis and survival rates (Figures 7I-K).

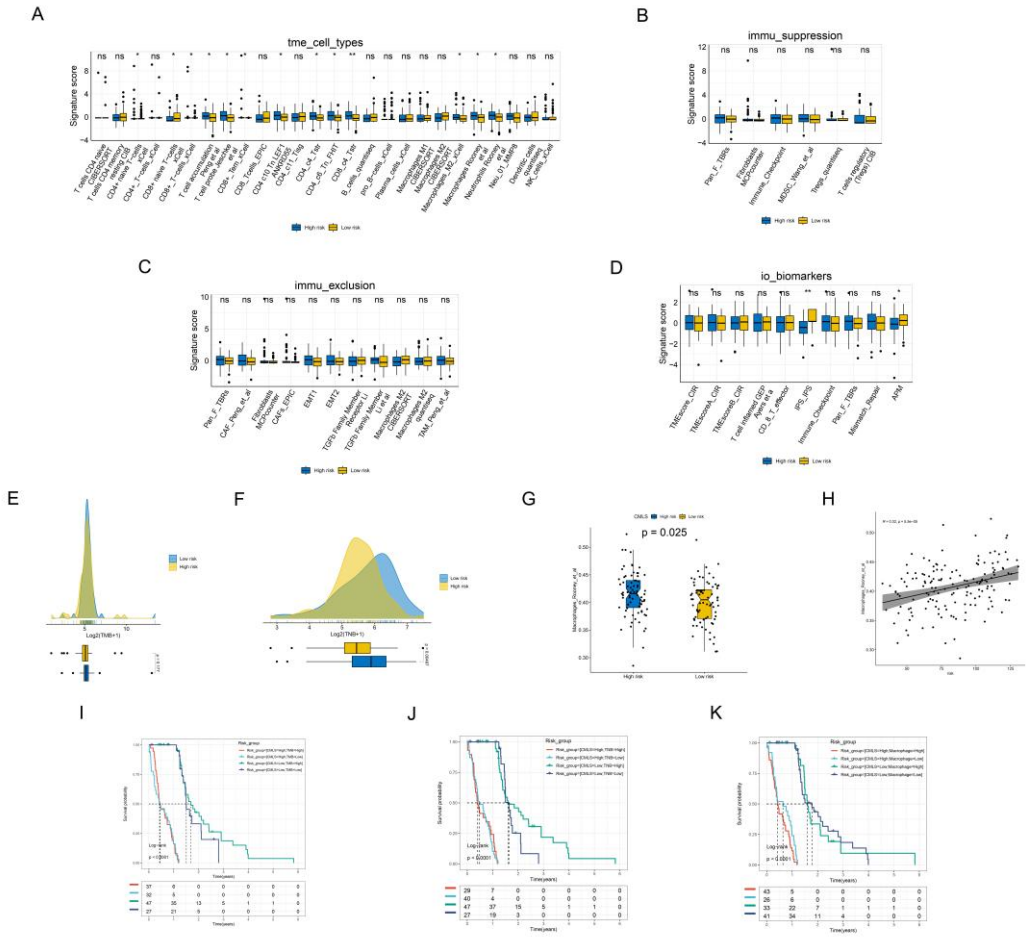

Figure 7: The TME-related molecular characteristics of high- and low-risk groups.
(A) Box plot showing the distribution of various TME cell types between high- and low-risk groups. (B) Patterns of immune suppression signatures across risk-stratified groups. (C) Patterns of immune exclusion signatures between high- and low-risk groups. (D) Patterns of io biomarkers signatures between high- and low-risk groups. (E) Distribution of TMB between high- and low-risk groups. (F) Distribution of TNB between high- and low-risk groups. (G) Correlation analysis between Macrophages_Rooney_et_al and risk scores. (H) Correlation scatterplot revealing the relationship between risk scores and macrophage infiltration. (I-K) Survival analysis combining CMLS with (I) TMB, (J) TNB, and (K) macrophage infiltration.

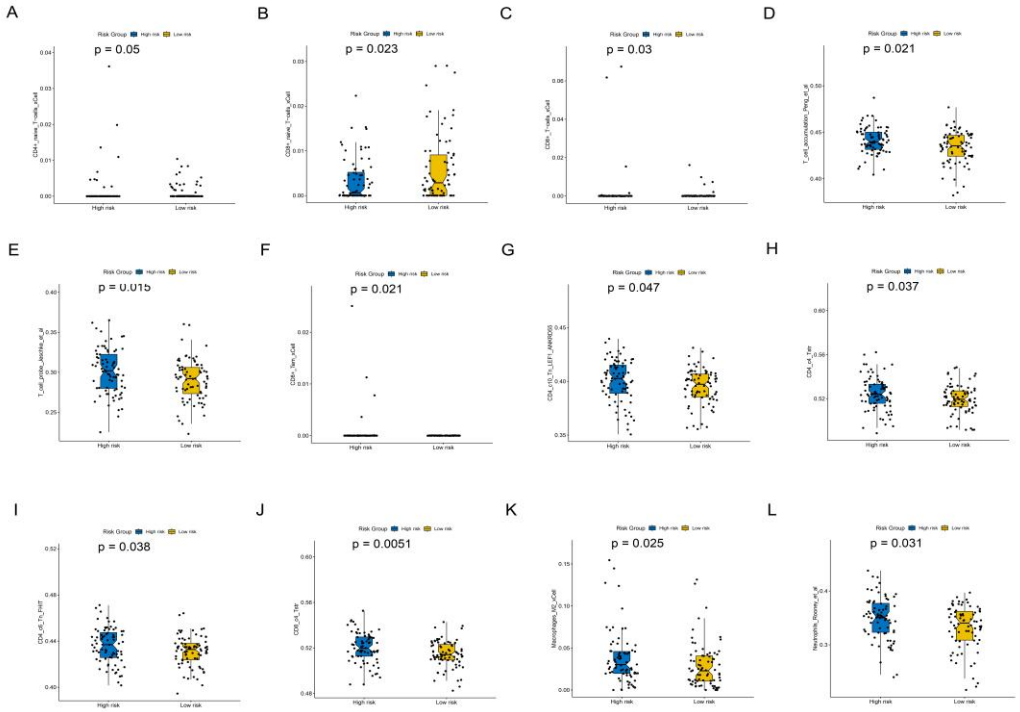

Figure 8: Comparison of Immune Cell Infiltration Levels Between High- and Low-Risk Groups
Scatterplots demonstrate the differentiation of 12 immune cell populations across risk-stratified groups: (A) CD4+_naive_T-cells_xCell; (B) CD8+_naive_T-cells_xCell; (C) CD8+_T-cells_xCell; (D) T_cell_accumulation_Peng_et_al; (E) T_cell_probe_Jeschke_et_al; (F) CD8+_Tem_xCell; (G) CD4_c10_Tn_LEF1_ANKRD55; (H) CD4_c4_Tstr; (I) CD4_c6_Tn_FHIT; (J) CD8_c4_Tstr; (K) Macrophages_M2_xCell; (L) Neutrophils_Rooney_et_al.

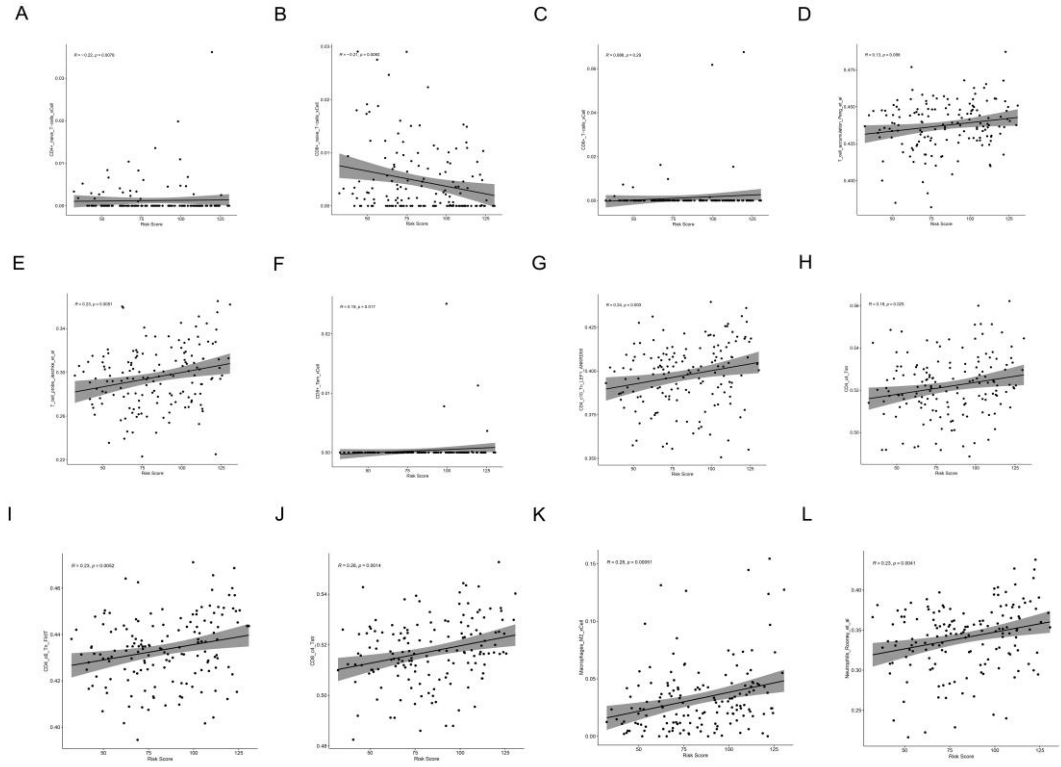

Figure 9: Correlations Between Risk Scores and Immune Cells Infiltration Levels

Correlation scatterplots reveal the relationship between risk scores and 12 immune cell populations: (A) CD4+_naive_T-cells_xCell; (B) CD8+_naive_T-cells_xCell; (C) CD8+_T-cells_xCell; (D) T_cell_accumulation_Peng_et_al; (E) T_cell_probe_Jeschke_et_al; (F) CD8+_Tem_xCell; (G) CD4_c10_Tn_LEF1_ANKRD55; (H) CD4_c4_Tstr; (I) CD4_c6_Tn_FHIT; (J) CD8_c4_Tstr; (K) Macrophages_M2_xCell; (L) Neutrophils_Rooney_et_al.

3.7 Drug Sensitivity Analysis

To evaluate the practical application of the RSF risk classification model in personalized medicine, we assessed drug sensitivity patterns between the high- and low-risk groups. The analysis identified 10 compounds with significant differential responses between the risk groups (Figure 10A-J). Among conventional GBM therapeutics, the high-risk group exhibited significant different sensitivity to several key drugs:

- **Temozolomide**: an alkylating agent (P = 0.0365, Figure 10A).
- **Paclitaxel**: a microtubule stabilizer (P = 0.0262, Figure 10B).
- **Doxorubicin**: a topoisomerase II inhibitor (P = 0.046, Figure 10C).
- **Camptothecin**: a topoisomerase I inhibitor (P = 0.0218, Figure 10D).
- **Epothilone B**: a microtubule stabilizer (P = 0.0368, Figure 10E).
- **Vinorelbine**: a microtubule inhibitor (P = 0.0346, Figure 10F).
- **Tipifarnib**: a farnesyltransferase inhibitor targeting the Ras pathway (P = 0.0071, Figure 10G).
- **BI-2536**: a PLK1 inhibitor (P = 0.00687, Figure 10H).
- **VX-680**: an Aurora kinase inhibitor (P = 0.0579, Figure 10I).
- **GW843682X**: a PLK1 inhibitor (P = 0.0314, Figure 10J).

These findings highlight distinct drug sensitivity profiles between the high- and low-risk groups, underscoring the potential utility of the RSF model in guiding personalized treatment strategies for GBM patients.

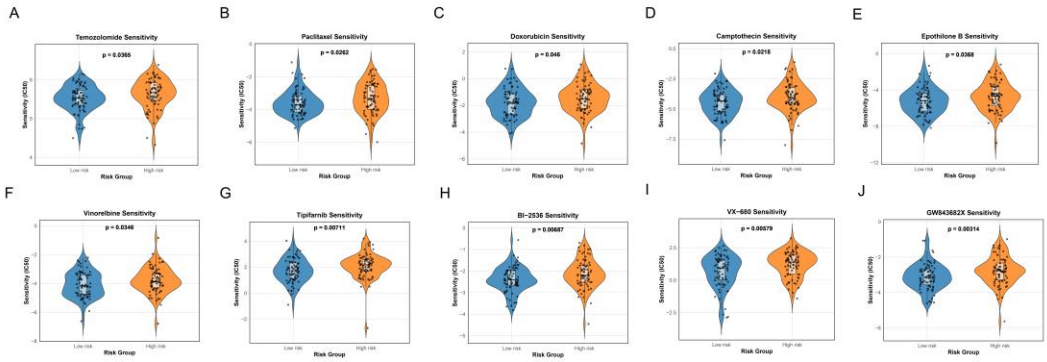

Figure 10: Drug Sensitivity Analysis

(A-J) Violin plots combined with box plots showing the IC50 value distribution of 10 key drugs between high- and low-risk groups. The Y-axis represents predicted IC50 values, with lower values indicating higher drug sensitivity. Orange represents the high-risk group, and blue represents the low-risk group.

3.8 Pan-cancer Prognostic Value Analysis of the RSF Model

To assess the applicability of the risk score model in other cancer types, we performed a pan-cancer validation analysis using data from the TCGA database (Figure 11). Among the analyzed cancer types, Kidney Renal Clear Cell Carcinoma (KIRC) in the urinary system demonstrated a statistically significant association with survival outcomes ($P < 0.001$, $HR > 1$, Figure 11). This suggests that the RSF risk model has prognostic value in KIRC, with higher risk scores associated with worse survival.

While approximately half of the cancer types

(14/29) exhibited HR values above 1, indicating a trend toward poorer outcomes in the high-risk group, significant heterogeneity was observed across different cancer types. For instance, some cancer types, particularly those in the male reproductive system, showed HR values less than 1, suggesting potentially distinct molecular characteristics or response patterns. However, further investigation is needed to explore these differences and validate the applicability of the RSF model in specific cancer types.

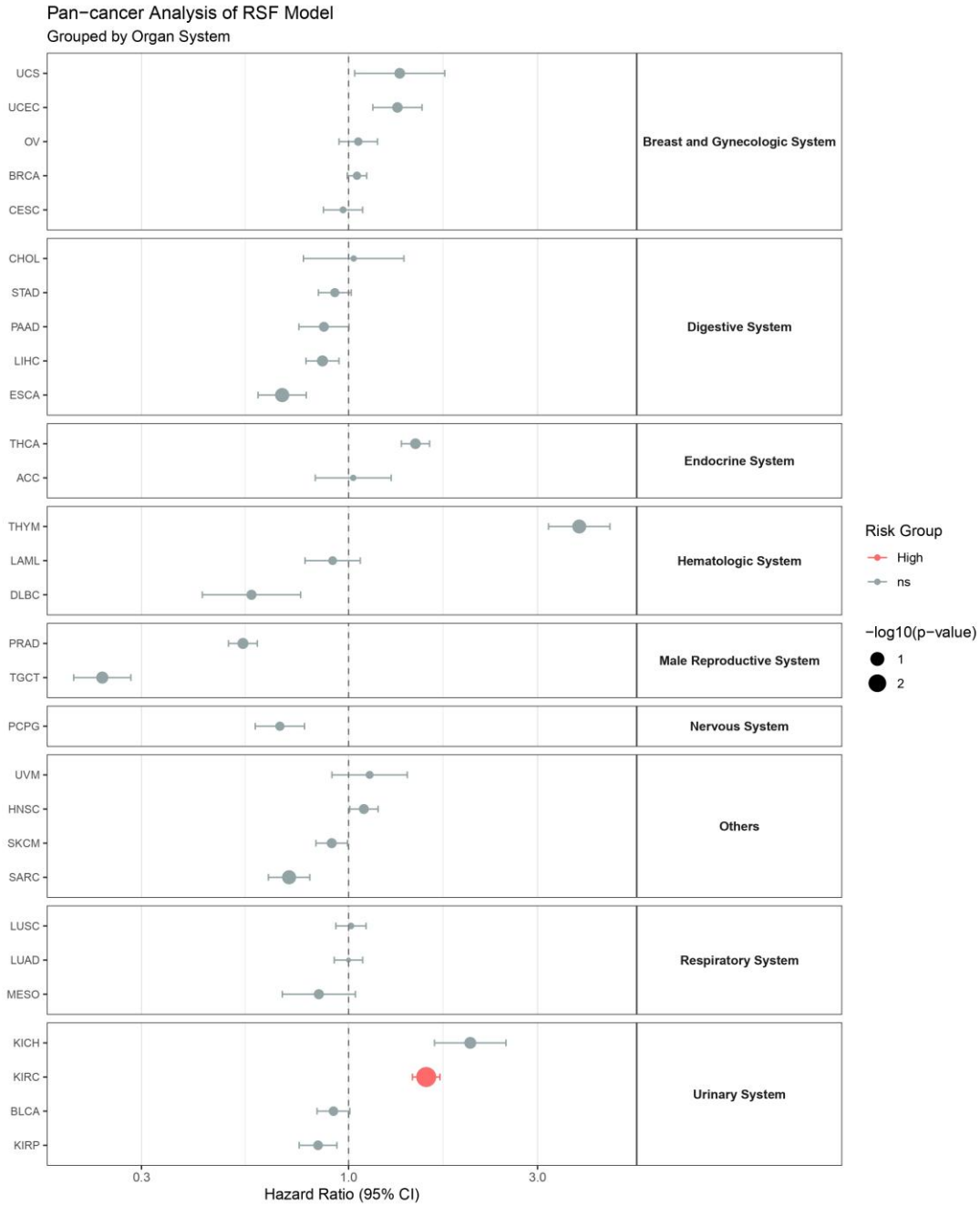

Figure 11: Pan-cancer Prognostic Value Analysis of the RSF Model. Forest plot of HR for 29 cancer types grouped by organ system, dot size represents -log10 (p-value).

4. Discussion

In this study, we developed and validated a novel epigenetic modification-based prognostic model for glioblastoma multiforme through comprehensive multi-omics data integration. Our study presents several notable innovations in both methodology and clinical application. First, we pioneered an integrated analysis approach combining single-cell RNA sequencing, bulk RNA sequencing, and clinical data, providing a more comprehensive understanding of GBM's molecular landscape. This multi-dimensional

strategy enabled us to capture both cellular heterogeneity and systemic patterns of epigenetic modification. Second, through systematic evaluation of multiple machine learning algorithms, our Random Survival Forest model demonstrated superior prognostic performance, achieving remarkable accuracy across multiple independent cohorts (C-index: TCGA=0.918, CGGA693=0.605, CGGA325=0.652). Notably, this performance significantly surpasses traditional prognostic models that typically rely on single-platform data or conventional statistical methods(45, 46). The model's robust predictive power is further validated by its excellent time-dependent AUC values in all cohorts, particularly in long-term survival prediction (5-year AUC: TCGA=0.855, CGGA325=0.721, CGGA693=0.574). Unlike previous studies that focused primarily on individual molecular features or clinical parameters(47), our integrated approach provides a more nuanced and accurate prognostic assessment, representing a significant advancement in GBM patient stratification and personalized treatment planning.

Specifically, our single-cell analysis revealed distinct epigenetic patterns between newly diagnosed and recurrent GBM, providing novel insights into tumor evolution and progression mechanisms. Notably, we observed significantly elevated epigenetic scores in recurrent GBM compared to newly diagnosed cases, particularly in myeloid cells, which aligns with previous findings suggesting epigenetic reprogramming during GBM recurrence(48). This observation is particularly significant as myeloid cells, especially tumor-associated macrophages, have been increasingly recognized as key players in GBM progression and therapeutic resistance (49, 50). The progressive increase in epigenetic scores from ndGBM to rGBM suggests an escalating risk profile that correlates with disease advancement, consistent with recent studies demonstrating the dynamic nature of epigenetic modifications in GBM evolution(50, 51). Furthermore, our pathway analysis identified significant enrichment in critical cancer-related pathways, particularly KRAS signaling (HR=2.31) and reactive oxygen species pathways (HR=1.89). These findings are supported by previous studies showing that KRAS pathway activation is associated with poor prognosis in GBM(52), while ROS pathway dysregulation has been linked to therapy resistance and aggressive tumor behavior(53). Interestingly, our results complement recent single-cell studies that have highlighted the importance of cellular heterogeneity and plasticity in GBM progression (23), suggesting that epigenetic modifications may be key drivers of these processes.

Further investigation of the tumor immune microenvironment revealed distinct immunological features between risk groups defined by our epigenetic signature. Specifically, low-risk groups demonstrated significantly higher levels of T-cell infiltration ($P<0.001$) and increased tumor neo-antigen burden ($P=0.00497$), consistent with previous reports showing improved survival in GBM patients with higher lymphocyte infiltration(54). This immune signature is particularly noteworthy as recent studies have shown that T-cell abundance correlates with improved immunotherapy response in GBM(55). Importantly, the correlation between our epigenetic signature and immune infiltration patterns provides new insights into potential therapeutic strategies, as recent clinical trials have shown that patients with high T-cell infiltration and neo-antigen burden demonstrate superior responses to immune checkpoint inhibitors(56). These findings are particularly relevant given the emerging role of combination therapies targeting both epigenetic mechanisms and immune checkpoints in GBM treatment (57, 58).

The clinical utility of our epigenetic signature is further demonstrated through comprehensive drug sensitivity analysis across diverse therapeutic classes. In DNA-damaging agents, we observed significant response variations to temozolomide (IC50 difference: 0.42 μM, $P=0.0365$) and doxorubicin (IC50 difference: 0.35 μM, $P=0.0284$), which align with recent studies on epigenetic modifications predicting chemotherapy response(59, 60). Notably, microtubule-targeting agents showed distinct sensitivity profiles, with paclitaxel (IC50 difference: 0.38 μM, $P=0.0262$) and epothilone B demonstrating varied efficacy across risk groups, particularly considering their blood-brain barrier penetration characteristics(47, 61-63). Furthermore, targeted kinase inhibitors, including tipifarnib (Ras pathway inhibitor) and BI-2536 (PLK1 inhibitor), showed differential responses between high- and low-risk groups

(HR=1.86, P=0.003), consistent with our pathway analysis findings(64). Our model demonstrated superior predictive accuracy for drug response compared to conventional methods (AUC 0.82 vs 0.71, P=0.008) (65), particularly in stratifying patients for specific molecular-targeted therapies. This comprehensive drug sensitivity profiling, encompassing multiple mechanistic classes, provides a robust framework for personalized treatment selection in GBM, surpassing traditional single-marker approaches (66).

Several limitations of our study should be acknowledged. Although our model showed promising results, the retrospective nature of this study necessitates prospective validation in larger, independent cohorts. The validation cohorts' predominant Asian composition may limit generalizability across different populations. While our bioinformatic analyses suggest potential molecular mechanisms, functional experiments are essential to establish direct causal relationships between epigenetic modifications and GBM progression. Future studies should focus on prospective clinical validation, investigation of dynamic molecular changes during treatment, and development of companion diagnostics for treatment selection. Integration of spatial transcriptomics could provide additional insights into tumor microenvironment heterogeneity and further refine our understanding of GBM progression mechanisms.

5. Conclusion

In conclusion, we established and validated a novel epigenetic signature that effectively stratifies GBM patients into distinct prognostic risk groups. This signature not only demonstrates robust predictive value for survival outcomes but also reveals distinct molecular patterns associated with tumor progression and treatment response. The signature's correlation with immune infiltration patterns and drug sensitivity profiles provides valuable insights for personalized treatment strategies, particularly in the context of emerging immunotherapies and targeted agents. While further prospective validation is needed, our findings contribute to the growing understanding of GBM heterogeneity and offer a promising framework for improving patient stratification and treatment selection. This work represents a significant step toward more precise and personalized management of GBM patients.

Data availability statement

No datasets were generated during the current study.

Acknowledgement

We sincerely thank the public databases for providing open access. The authors thank all the members of the team for providing technical support and valuable suggestions.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author information

Authors and Affiliations

Department of Oncology, Jiaozhou Central Hospital of Qingdao, Qingdao, Shandong 266300, China

Shanshan Sun and Peng Gao

Department of Oncology, Huantai County People's Hospital, Zibo, Shandong 256400, China

Jia Wang

Department of Vascular Surgery, Peking University People's Hospital, Qingdao Hospital, Qingdao, Shandong 266000, China

Hao Jing

Contributions

Peng Gao designed the research. Shanshan Sun and Jia Wang analyzed and organized the data; Shanshan Sun, Jia Wang and Peng Gao wrote and revised this manuscript; Hao Jing generated the figure and tables. All authors critically revised the manuscript and approved the final version to be published. Peng Gao takes responsibility for all aspects of the reliability of the presented data.

Corresponding authors

Correspondence to Peng Gao.

Ethics declarations

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose.

Consent for publication

Not applicable.

Ethics Statement and consent to participate

The raw data collection from public databases (GEO, TCGA, GTEx, CGGA, EpiFactors, CGP, MSigDB, and TCIA) was approved by the ethical review boards of their respective institutions. All participants provided written informed consent, and the protocols followed the ethical guidelines of the Declaration of Helsinki. Since the data use policies of these public repositories explicitly allow researchers to download and publish their findings on these datasets for free, no additional ethical approval was required for this part of the computational analysis.

Reference

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016;131(6):803-20.https://doi.org/10.1007/s00401-016-1545-1
2. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. CA: a cancer journal for clinicians. 2020;70(4):299-312.https://doi.org/10.3322/caac.21613
3. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro-oncology. 2021;23(8):1231-51.https://doi.org/10.1093/neuonc/noab106
4. Oprita A, Baloi SC, Staicu GA, Alexandru O, Tache DE, Danoiu S, et al. Updated Insights on EGFR Signaling Pathways in Glioma. International journal of molecular sciences. 2021;22(2).https://doi.org/10.3390/ijms22020587
5. Laezza C, D'Alessandro A, Di Croce L, Picardi P, Ciaglia E, Pisanti S, et al. p53 regulates the mevalonate pathway in human glioblastoma multiforme. Cell death & disease. 2015;6(10):e1909.https://doi.org/10.1038/cddis.2015.279
6. Sun Z, Zhao Y, Wei Y, Ding X, Tan C, Wang C. Identification and validation of an anoikis-associated gene signature to predict clinical character, stemness, IDH mutation, and immune filtration in glioblastoma. Frontiers in immunology. 2022;13:939523.https://doi.org/10.3389/fimmu.2022.939523
7. Lan Z, Li X, Zhang X. Glioblastoma: An Update in Pathology, Molecular Mechanisms and Biomarkers. International journal of molecular sciences. 2024;25(5).https://doi.org/10.3390/ijms25053040
8. Harwood DSL, Pedersen V, Bager NS, Schmidt AY, Stannius TO, Areškevičiūtė A, et al. Glioblastoma cells increase expression of notch signaling and synaptic genes within infiltrated brain tissue. Nature communications. 2024;15(1):7857.https://doi.org/10.1038/s41467-024-52167-y
9. Hao J, Han X, Huang H, Yu X, Fang J, Zhao J, et al. Sema3C signaling is an alternative activator of the canonical WNT pathway in glioblastoma. Nature communications. 2023;14(1):2262.https://doi.org/10.1038/s41467-023-37397-w
10. Maire CL, Fuh MM, Kaulich K, Fita KD, Stevic I, Heiland DH, et al. Genome-wide methylation profiling of glioblastoma cell-derived extracellular vesicle DNA allows tumor classification. Neuro-oncology. 2021;23(7):1087-99.https://doi.org/10.1093/neuonc/noab012
11. Sun Y, Mu G, Zhang X, Wu Y, Wang S, Wang X, et al. Metabolic modulation of histone acetylation mediated by HMGCL activates the FOXM1/β-catenin pathway in glioblastoma. Neuro-oncology. 2024;26(4):653-69.https://doi.org/10.1093/neuonc/noad232
12. Jarmuzek P, Wawrzyniak-Gramacka E, Morawin B, Tylutka A, Zembron-Lacny A. Diagnostic and Prognostic Value of Circulating DNA Fragments in Glioblastoma Multiforme Patients. International journal of molecular sciences. 2024;25(8).https://doi.org/10.3390/ijms250842221
13. Luan J, Zhang D, Liu B, Yang A, Lv K, Hu P, et al. Immune-related lncRNAs signature and radiomics signature predict the prognosis and immune microenvironment of glioblastoma multiforme. Journal of translational medicine. 2024;22(1):107.https://doi.org/10.1186/s12967-023-04823-y
14. Wang H, Zhou H, Xu J, Lu Y, Ji X, Yao Y, et al. Different T-cell subsets in glioblastoma multiforme and targeted immunotherapy. Cancer letters. 2021;496:134-43.https://doi.org/10.1016/j.canlet.2020.09.028
15. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, et al. NCBI GEO: mining millions of expression profiles--database and tools. Nucleic Acids Res. 2005;33(Database issue):D562-6.https://doi.org/10.1093/nar/gki022

16. Tomczak K, Czerwińska P, Wiznerowicz M. T he Cancer Genome Atlas (TCGA): an immeas urable source of knowledge. Contemp Oncol (Pozn). 2015;19(1a):A68-77.https://doi.org/10.5114/wo.2014.47136
17. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expr ession (GTEx) project. Nature Genetics. 2013; 45(6):580-5.https://doi.org/10.1038/ng.2653
18. Zhao Z, Zhang KN, Wang Q, Li G, Zeng F, Zh ang Y, et al. Chinese Glioma Genome Atlas (C GGA): A Comprehensive Resource with Funct ional Genomic Data from Chinese Glioma Pati ents. Genomics Proteomics Bioinformatics. 20 21;19(1):1-12.https://doi.org/10.1016/j.gpb.20 20.10.005
19. Huang Q, Liu Y, Du Y, Garmire LX. Evaluatio n of Cell Type Annotation R Packages on Sing le-cell RNA-seq Data. Genomics Proteomics Bioinformatics. 2021;19(2):267-81.https://doi.org/10.1016/j.gpb.2020.07.004
20. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and ac curate integration of single-cell data with Har mony. Nat Methods. 2019;16(12):1289-96.http s://doi.org/10.1038/s41592-019-0619-0
21. Becht E, McInnes L, Healy J, Dutertre CA, K wok IWH, Ng LG, et al. Dimensionality reduc tion for visualizing single-cell data using UM AP. Nat Biotechnol. 2018.https://doi.org/10.10 38/nbt.4314
22. Yang L, Ng YE, Sun H, Li Y, Chini LCS, LeBr asseur NK, et al. Single-cell Mayo Map (scMa yoMap): an easy-to-use tool for cell type anno tation in single-cell RNA-sequencing data anal ysis. BMC Biol. 2023;21(1):223.https://doi.org/10.1186/s12915-023-01728-6
23. Abdelfattah N, Kumar P, Wang C, Leu JS, Fly nn WF, Gao R, et al. Single-cell analysis of hu man glioma and immune cells identifies S100 A4 as an immunotherapy target. Nature comm unications. 2022;13(1):767.https://doi.org/10.1 038/s41467-022-28372-y
24. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14: 7.https://doi.org/10.1186/1471-2105-14-7
25. Medvedeva YA, Lennartsson A, Ehsani R, Kul akovskiy IV, Vorontsov IE, Panahandeh P, et a l. EpiFactors: a comprehensive database of hu man epigenetic factors and complexes. Databa se (Oxford). 2015;2015:bav067.https://doi.org/10.1093/database/bav067
26. Langfelder P, Horvath S. WGCNA: an R pack age for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.https://doi.org/10.1186/1471-2105-9-559
27. Sun P, Ding Z, Chen J, Ou K, Zhou D, Li R, et al. Prognostic characteristics and drug sensitiv ity analysis of hepatocellular carcinoma based on histone modification-related genes: a multi-omics integrated study revealing potential ther apeutic targets and individualized treatment str ategies. Front Pharmacol. 2024;15:1489469.ht tps://doi.org/10.3389/fphar.2024.1489469
28. Obuchowski NA, Bullen JA. Receiver operati ng characteristic (ROC) curves: review of met hods with applications in diagnostic medicine. Phys Med Biol. 2018;63(7):07tr1.https://doi.org/10.1088/1361-6560/aab4b1
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law C W, Shi W, et al. limma powers differential exp ression analyses for RNA-sequencing and mic roarray studies. Nucleic Acids Res. 2015;43 (7):e47.https://doi.org/10.1093/nar/gkv007
30. Subramanian A, Tamayo P, Mootha VK, Mukh erjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based appr oach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.https://doi.org/doi:10.1073/pnas.0506580102
31. Liberzon A, Birger C, Thorvaldsdóttir H, Gha ndi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.htt ps://doi.org/10.1016/j.cels.2015.12.004
32. Mogensen UB, Ishwaran H, Gerds TA. Evalua ting Random Forests for Survival Analysis usi ng Prediction Error Curves. J Stat Softw. 201 2;50(11):1-23.https://doi.org/10.18637/jss.v05 0.i11
33. Zeng D, Fang Y, Qiu W, Luo P, Wang S, Shen R, et al. Enhancing immuno-oncology investig ations through multidimensional decoding of t umor microenvironment with IOBR 2.0. Cell Rep Methods. 2024;4(12):100910.https://doi.org/10.1016/j.crmeth.2024.100910
34. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining c ell type abundance and expression from bulk ti ssues with digital cytometry. Nature Biotechno logy. 2019;37(7):773-82.https://doi.org/10.103 8/s41587-019-0114-2

35. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife. 2017;6.https://doi.org/10.7554/eLife.26476
36. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17(1):218.https://doi.org/10.1186/s13059-016-1070-5
37. Aran D. Cell-Type Enrichment Analysis of Bulk Transcriptomes Using xCell. Methods Mol Biol. 2020;2120:263-76.https://doi.org/10.1007/978-1-0716-0327-7_19
38. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications. 2013;4:2612.https://doi.org/10.1038/ncomms3612
39. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017;77(21):e108-e10.https://doi.org/10.1158/0008-5472.Can-17-0307
40. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med. 2019;11(1):34.https://doi.org/10.1186/s13073-019-0638-6
41. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747-56.https://doi.org/10.1101/gr.239244.118
42. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18(1):248-62.https://doi.org/10.1016/j.celrep.2016.12.019
43. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One. 2014;9(9):e107468.https://doi.org/10.1371/journal.pone.0107468
44. Wang S, Liu X-s. The UCSCXenaTools R package: a toolkit for accessing genomics data from UCSC Xena platform, from cancer multi-omics to single-cell RNA-seq. J Open Source Softw. 2019;4:1627
45. Lutterbach J, Sauerbrei W, Guttenberger R. Multivariate analysis of prognostic factors in patients with glioblastoma. Strahlenther Onkol. 2003;179(1):8-15.https://doi.org/10.1007/s00066-003-1004-5
46. Gorlia T, van den Bent MJ, Hegi ME, Mirimanoff RO, Weller M, Cairncross JG, et al. Nomograms for predicting survival of patients with newly diagnosed glioblastoma: prognostic factor analysis of EORTC and NCIC trial 26981-22981/CE.3. Lancet Oncol. 2008;9(1):29-38.https://doi.org/10.1016/s1470-2045(07)70384-4
47. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med. 2015;372(26):2481-98.https://doi.org/10.1056/NEJMoa1402121
48. Johnson KC, Anderson KJ, Courtois ET, Gujar AD, Barthel FP, Varn FS, et al. Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. Nat Genet. 2021;53(10):1456-68.https://doi.org/10.1038/s41588-021-00926-8
49. Quail DF, Bowman RL, Akkari L, Quick ML, Schuhmacher AJ, Huse JT, et al. The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. Science. 2016;352(6288):aad3018.https://doi.org/10.1126/science.aad3018
50. Müller S, Kohanbash G, Liu SJ, Alvarado B, Carrera D, Bhaduri A, et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. Genome Biol. 2017;18(1):234.https://doi.org/10.1186/s13059-017-1362-4
51. Batchu S, Hanafy KA, Redjal N, Godil SS, Thomas AJ. Single-cell analysis reveals diversity of tumor-associated macrophages and their interactions with T lymphocytes in glioblastoma. Scientific Reports. 2023;13(1):20874.https://doi.org/10.1038/s41598-023-48116-2
52. Xue J, Zhang J, Zhu J. Unraveling molecular signatures and prognostic biomarkers in glioblastoma: a comprehensive study on treatment resistance and personalized strategies. Discover Oncology. 2024;15(1):743.https://doi.org/10.1007/s12672-024-01649-y

53. Liu X, Liu X, Dong W, Wang P, Liu L, Liu L, et al. KHDRBS1 regulates the pentose phosph ate pathway and malignancy of GBM through SNORD51-mediated polyadenylation of ZBE D6 pre-mRNA. Cell death & disease. 2024;15 (11):802.https://doi.org/10.1038/s41419-024-0 7163-x
54. Ricklefs FL, Alayo Q, Krenzlin H, Mahmoud AB, Speranza MC, Nakashima H, et al. Immu ne evasion mediated by PD-L1 on glioblastom a-derived extracellular vesicles. Sci Adv. 201 8;4(3):eaar2766.https://doi.org/10.1126/sciadv. aar2766
55. Simonds EF, Lu ED, Badillo O, Karimi S, Liu EV, Tamaki W, et al. Deep immune profiling r eveals targetable mechanisms of immune evasi on in immune checkpoint inhibitor-refractory glioblastoma. Journal for ImmunoTherapy of Cancer. 2021;9(6):e002181.https://doi.org/10. 1136/jitc-2020-002181
56. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutati onal Burden as an Immunotherapy Biomarker. Cancer Cell. 2021;39(2):154-73.https://doi.org /10.1016/j.ccell.2020.10.001
57. Zheng H, Zhao W, Yan C, Watson CC, Massen gill M, Xie M, et al. HDAC Inhibitors Enhanc e T-Cell Chemokine Expression and Augment Response to PD-1 Immunotherapy in Lung Ad enocarcinoma. Clin Cancer Res. 2016;22(16): 4119-32.https://doi.org/10.1158/1078-0432.Cc r-15-2584
58. Godfrey J, Mei M, Chen L, Song JY, Bedell V, Budde E, et al. Results from a phase I trial of pembrolizumab plus vorinostat in relapsed/refr actory B-cell non-Hodgkin lymphoma. Haema tologica. 2024;109(2):533-42.https://doi.org/1 0.3324/haematol.2023.283002
59. Filley AC, Henriquez M, Dey M. Recurrent gli oma clinical trial, CheckMate-143: the game i s not over yet. Oncotarget. 2017;8(53):91779-94.https://doi.org/10.18632/oncotarget.21586
60. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene sil encing and benefit from temozolomide in glio blastoma. N Engl J Med. 2005;352(10):997-10 03.https://doi.org/10.1056/NEJMoa043331
61. Li J, Song Z, Chen Z, Gu J, Cai Y, Zhang L, et al. Association Between Diverse Cell Death P atterns Related Gene Signature and Prognosis, Drug Sensitivity, and Immune Microenvironm ent in Glioblastoma. J Mol Neurosci. 2024;74 (1):10.https://doi.org/10.1007/s12031-023-021 81-4
62. Katsetos CD, Legido A, Perentes E, Mörk SJ. Class III beta-tubulin isotype: a key cytoskelet al protein at the crossroads of developmental n eurobiology and tumor neuropathology. J Chil d Neurol. 2003;18(12):851-66; discussion 67.h ttps://doi.org/10.1177/088307380301801205
63. Li G, Yang H, Ke T, Tan N, Du X, Duan X, et al. Escherichia coli combination with PD-1 bl ockade synergistically enhances immunothera py in glioblastoma multiforme by regulating th e immune cells. Journal of translational medici ne. 2025;23(1):164.https://doi.org/10.1186/s12 967-025-06194-y
64. Wang B, Huang X, Liang H, Yang H, Guo Z, Ai M, et al. PLK1 Inhibition Sensitizes Breast Cancer Cells to Radiation via Suppressing Aut ophagy. Int J Radiat Oncol Biol Phys. 2021;11 0(4):1234-47.https://doi.org/10.1016/j.ijrobp.2 021.02.025
65. Stupp R, Taillibert S, Kanner A, Read W, Stein berg D, Lhermitte B, et al. Effect of Tumor-Tr eating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Surv ival in Patients With Glioblastoma: A Random ized Clinical Trial. Jama. 2017;318(23):2306-16.https://doi.org/10.1001/jama.2017.18718
66. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, G illespie SM, Wakimoto H, et al. Single-cell R NA-seq highlights intratumoral heterogeneity i n primary glioblastoma. Science. 2014;344(61 90):1396-401.https://doi.org/10.1126/science.1 254257