Contents lists available at ScienceDirect

Heliyon

Research article

Association of M2 macrophages with EMT in glioma identified through combination of multi-omics and machine learning

Peng Feng [a], Shangyu Liu [a], Guoqiang Yuan [a,b,c,**], Yawen Pan [a,b,c,*]

[a] The Second Medical College of Lanzhou University, Lanzhou, Gansu, 730030, PR China
[b] Department of Neurosurgery, Second Hospital of Lanzhou University, Lanzhou, Gansu, 730030, PR China
[c] Key Laboratory of Neurology of Gansu Province, Lanzhou University, Lanzhou, Gansu, 730030, PR China

ABSTRACT

Background: The incidence of glioma, a prevalent brain malignancy, is increasing, particularly among the elderly population. This study aimed to elucidate the clinical importance of epithelial–mesenchymal transition (EMT) in gliomas and its association with malignancy and prognosis.

Background: The incidence of glioma, particularly among elderly individuals, is on the rise. The malignancy of glioma is determined not only by the oncogenic properties of tumor cells but also by the composition of the tumor microenvironment, which includes immune system macrophages. The prevalence of M2-type macrophages typically fosters tumor progression, yet the underlying mechanism remains elusive. Our study explored the clinical importance of epithelial–mesenchymal transition (EMT) in gliomas and its association with malignancy and prognosis.

Methods: Our study used the gene set variation analysis (GSVA) algorithm to classify different levels of EMT activation based on the transcriptomic and multi-omics data. Machine learning (ML) and single-cell analysis were integrated into our model for comprehensive analysis. A predictive model was constructed and in vitro experiments were performed to validate our findings.

Results: Our study classified 1,641 samples into two clusters based on EMT activation: the EMT-hot group and the EMT-cold group. The EMT-hot group had elevated copy number loss, tumor mutational burden (TMB), and a poorer survival rate. Conversely, the EMT-cold group showed a better survival rate, likely attributed to lower stromal and immune cell scores, as well as decreased expression of human leukocyte antigen-related genes. Driving genes were identified through weighted gene coexpression network analysis (WGCNA) and dimensionality reduction techniques. These genes were then utilized in the construction of a prognostic model using ML and protein-protein interaction (PPI) network analysis. Furthermore, the impact of the core genes identified through single-cell analysis on glioma prognosis was examined.

Conclusion: Our research underscores the efficacy of our model in predicting glioma prognosis and elucidates the connection between the M2 macrophages and EMT. Additionally, core genes such as LY96, C1QB, LGALS1, CSPG5, S100A8, and CHGB were identified as pivotal for mediating the occurrence of EMT induced by M2 macrophages.

* Corresponding author. The Second Medical College of Lanzhou University, Lanzhou, Gansu, 730030, PR China.
** Corresponding author. The Second Medical College of Lanzhou University, Lanzhou, Gansu, 730030, PR China.
E-mail addresses: sjwkfp@163.com (P. Feng), liushy878@163.com (S. Liu), yuangq08@lzu.edu.cn (G. Yuan), pyw@lzu.edu.cn (Y. Pan).

1. Introduction

Gliomas, a type of neuroepithelial tumor originating from the glial cells of the brain, can develop at different stages of glial cell differentiation, resulting in varying glioma subtypes [1]. Neuroepithelial tissue accounts for a large proportion of primary brain tumors [2]. Glioblastoma multiforme (GBM) is the most aggressive and malignant variant and is characterized by the highest incidence rate among gliomas and a dismal prognosis [3]. In recent decades, the incidence of primary brain tumors has increased, with an annual growth rate of 1 %–2 %, particularly among elderly individuals [4]. The World Health Organization (WHO) classification system categorizes gliomas into low-grade gliomas (WHO grades I-II) and high-grade gliomas (WHO grades III-IV). Overall, gliomas pose substantial therapeutic challenges due to their high lethality and susceptibility to recurrence [5,6]. The prognosis for patients with gliomas has not effectively improved in recent years, with an average survival period of only 12–15 months, and a 5-year survival rate of 5.5 %, despite remarkable advances in microscopic technology, imaging technology, and radiotherapy, as well as the advent of treatment modalities such as novel antitumor drugs, molecular targeted drugs, and immunomodulatory therapies [4].

Epithelial–mesenchymal transition (EMT) is the process by which epithelial cells forfeit their polarity to assume a mesenchymal cell phenotype, facilitating the acquisition of migratory properties [7]. In normal human tissue, epithelial cells rely on intercellular connections and polarity to maintain tissue integrity. These connections primarily consist of tight junctions, gap junctions, adherens junctions, and desmosomes, while apical–basal polarity is governed by protein complexes distributed along the cell apical–basal axis. EMT occurs when cells relinquish their epithelial properties, namely intercellular adhesion and polarity.

EMT can be classified into three types based on the underlying biological mechanisms: type I EMT, which is integral to embryonic development and underscores the significance of cellular diversity during this process; type II EMT, which is implicated in wound healing and tissue regeneration, in which fibroblasts are produced through EMT; and type III EMT, which is associated with various diseases such as cancer progression and empowers cancer cells with invasive capabilities, facilitating metastasis from the primary tumor to distant sites. The occurrence of EMT coincides with the transition from predominantly E-cadherin expression in epithelial cells to N-cadherin expression, a hallmark of mesenchymal cells, culminating in the morphological transformation of epithelial cells from polygonal and cobblestone shapes to spindle-like shapes characteristic of mesenchymal cells. Consequently, epithelial cells lose their apical–basal polarity and cell integrity, precipitating tumor cell migration and invasion of tumor cells. While gliomas typically do not metastasize to extracranial sites, they can to infiltrate adjacent healthy brain tissue and disseminate within the brain, intermingling with normal healthy tissue and posing challenges in delineating tumor boundaries from healthy tissue.

In a study by Scherer, 45 % of GBM cases exhibited tumor extension beyond a single brain lobe, with 25 % demonstrating hemisphere-wide infiltration and 25 %–30 % displaying contralateral hemisphere involvement. Nearly 60 % of lateral supratentorial gliomas exhibited anteroposterior dissemination, while approximately 20 % invaded deep supratentorial and infratentorial structures along the fiber bundles. In both high- and low-grade gliomas, glioma invasion patterns varied based on tumor location. Frontal gliomas (nine cases) displayed a predilection for corpus callosum (CC)-mediated frontal lobe invasion, while temporal gliomas tended to invade the midbrain and pons. Gliomas originating below the CC primarily involved basal structures along the corticospinal tract, such as the thalamus and penduculi. Bilateral glioma extension occurred through infiltration of the thalamus, hypothalamus, and anterior commissure. Thalamic and hypothalamic gliomas also showed bilateral extension. Eight cases of pontine tumors often extended to the midbrain and thalamus. Occasionally, caudal gliomas directly invaded the upper cervical spinal cord. Untreated lesions harbored tumor cells within a 3 cm radius of necrotic tissue, whereas recurrent cases exhibited far-reaching tumor cell dissemination beyond the primary lesion. Strikingly, 80 % of the patients demonstrated contralateral hemisphere involvement. Moreover, glioma invasion and dissemination led to residual tumor cell infiltration into surrounding healthy tissue, resulting in tumor recurrence. Another concern is the progression of many low-grade gliomas to lethal high-grade gliomas over time, even when quiescent. Therefore, elucidating the mechanisms governing tumor progression and invasion is imperative for enhancing clinical decision-making efficacy.

Our research revealed distinct EMT enrichment profiles, namely, the EMT-hot and EMT-cold subgroups, which were identified from a cohort of 1,641 samples. The EMT-hot subgroup, characterized by noticeable activation of the interleukin-2 (IL-2) and IL-6 pathways, exhibited significant infiltration of M2 macrophages and elevated expression of genes related to human leukocyte antigen (HLA) and major histocompatibility complex (MHC), correlating with poorer prognosis. The EMT-hot subgroup exhibited increased MGMTp_Unmethylated and 1p19q noncodeleted profiles, as well as a higher prevalence of high-grade glioma. Moreover, higher tumor purity, more gene mutations, and significant amplification at 7q11.2 were found in this subgroup, which was particularly pronounced among elderly patients. Additionally, this subgroup had a greater frequency of "IDH wild-type" gliomas, and 1p/19q commutation was more common in this subgroup. Then, the weighted gene coexpression network analysis (WGCNA) algorithm and "limma" R package were used to identify driver genes in the two subgroups, with a focus on immune- and survival-related genes. Core genes were obtained using random forest (RF) and least absolute shrinkage and selection operator (Lasso) regression techniques, followed by the construction of a prognostic model using the "survival" package in R, resulting in the identification of 16 genes. Subsequent protein–protein interaction (PPI) network analysis identified six key genes—LY96, C1QB, LGALS1, CSPG5, S100A8, and CHGB—indicative of potential M2 macrophage involvement, as revealed by single-cell analysis. Validation of core gene expression was carried out using tissue samples. In brief, our research suggested that the identified core genes may facilitate EMT in gliomas through the activation of the IL-2 and IL-6 pathways, thereby enhancing tumor-associated macrophage infiltration and contributing to the poor prognosis of gliomas (Fig. 1). In this regard, our study can provide insights into molecular-targeted research on glioma based on the underlying biological mechanisms of EMT.
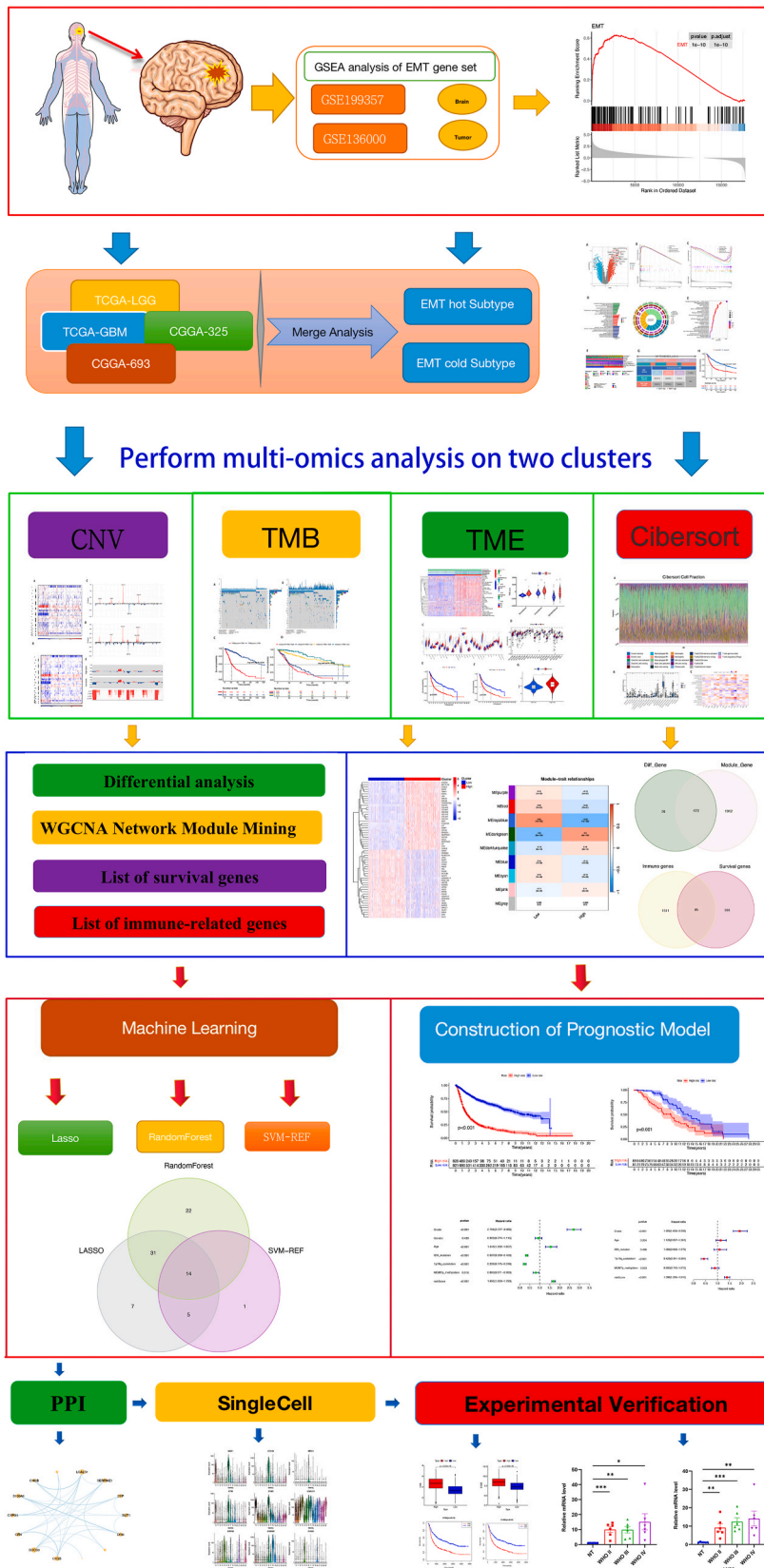
(caption on next page)

Fig. 1. Analyze the flowchart.

2. Materials and methods

2.1. Data source

Transcriptome expression data and clinical data were obtained from two databases: the China Glioma Genome Map (CGGA) database (http://www.cgga.org.cn) and The Cancer Genome Atlas (TCGA) program (https://portal.gdc.cancer.gov/) [8–13] The integration of available samples containing genomic and clinical characteristics produced a total of 1,641 patients with gliomas who were included in subsequent analysis. Clinicopathological characteristics and somatic cell mutations were obtained from the cBio-Portal platform (https://www.cbioportal.org/). Additionally, external datasets including CGGA-301, GSE199375, GSE136000, and GSE184941 [14,15]were utilized to verify our findings.

2.2. Subgrouping and clinical correlation analysis

To explore the potential role of marker gene sets in gliomas, we conducted the single-sample gene set enrichment analysis (ssGSEA) to assess the relationships between 50 tumor marker gene sets and samples. Then, the samples were stratified into two subgroups based on the degree of EMT enrichment: the EMT-hot subgroup (exhibiting high-level enrichment of EMT) and the EMT-cold subgroup (showing low-level enrichment of EMT), which were determined according to the median value of the EMT enrichment score. Differences between subgroups were evaluated using the "limma" package to generate a list of differentially expressed genes (DEGs). Further insights into the roles and interactions of DEGs were obtained using the Gene Ontology (GO) database, exploring cellular component (CC), molecular function (MF), and biological process (BP) perspectives. Additionally, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and specific gene set analysis were performed to examine the relationships between each subgroup and clinical features.

2.3. Weighted gene coexpression network analysis (WGCNA)

WGCNA aims to identify gene modules across different sample subgroups [16]. The optimal soft threshold was determined using the scale-free topology criterion. In our analysis, only gene modules containing at least 50 genes were considered. The dynamic tree-cutting method was used to identify modules. The module eigengene dissimilarity (MEDisThres) parameter was set to 0.25.

2.4. Multi-omics analysis

To better understand the differences between the two subgroups, differences in CNV, TMB, and the tumor microenvironment (TME) were first evaluated. Moreover, the differences in changes in the HLA-related gene sets between the two groups were investigated.

2.5. Gene intersection analysis

After screening 17,453 genes from the merged dataset, 473 survival-related genes were identified. Then, a total of 85 genes were identified through the intersection analysis of 2,385 module-related genes, 549 DEGs, and 1,596 immune-related genes.

2.6. Identification of genes using ML and a prognostic model

The merged dataset served as the training set, while GSE184941 was used as the testing set to verify the accuracy of the model. The performance of the constructed model was evaluated through univariate and multivariate analyses. The genes incorporated into the model and their corresponding risk coefficient values (RCVs) are presented in Table. The risk score for each sample is derived by summing the products of the mRNA expression levels (Tpm values) of the model-building genes and their respective RCVs. Then, the samples were categorized into high-risk and low-risk groups based on the median risk score. Notably, if a sequenced gene remains undetected, the risk score attributed to that specific gene is to be zero. Given that the mRNA expression level of the core genes is pivotal for risk scoring, the reliability of this model hinges on detecting at least 90 % of the gene expression levels. The formula for computing the risk score of an individual sample is as follows: Sample Risk Score = *Gene1** RCV 1 + Gene2* RCV 2 + … + *Gene19** RCV 19.

Table 1
The model building genes and their detailed risk coefficient values (RCV).

Gene	RCV	Gene	RCV	Gene	RCV	Gene	RCV
C3	−0.139698917	SAA1	0.04682605	FCGR2A	0.465167782	TNFRSF12A	0.193838219
CD14	−0.203809676	SOCS3	0.075428925	IFITM2	−0.110570506	TREM1	−0.18034441
CEBPD	−0.10436788	SPON2	0.119368556	LIF	−0.167567541	TRIM67	−0.128779386
CHGB	−0.054990473	SPP1	0.142208349	NPPA	0.054302809	TYMP	0.173686942
CTSS	−0.298866328	TNC	0.113996165	PLA2G2A	0.046645746		

Moreover, machine learning (ML) techniques including RF, LASSO regression, and support vector machine (SVM) were utilized to identify relevant genes, followed by intersection analyses that revealed 14 genes. To further elucidate the relationships between these genes, protein-protein interaction (PPI) network analysis was conducted, resulting in the identification of six core genes: LY96, C1QB, LGALS1, CSPG5, S100A8, and CHGB.

2.7. Single-cell analysis

Single-cell datasets from 10× Genomics were downloaded from the Chinese Glioma Genome Atlas (CGGA) database 6148, and then they were imported into the "Seurat" R package (version 4.2.2) for quality control (QC) and downstream analysis. Low-quality cells were first removed, followed by data normalization using the "Log Normalize" function, which converts the raw gene expression counts into a log scale. The "Find Variable Features" function was applied to select genes with the highest variability as "variable features," returning a specified number of variable features (default value: 2000). Principal component analysis (PCA) was then utilized for dimensionality reduction, followed by clustering analysis using graph-based clustering techniques. The t-distributed stochastic neighbor embedding (t-SNE) was employed for the visualization of the clustering results. Finally, the "Seurat" package was used for visualizing gene expression, where functions such as DotPlot, VlnPlot, FeaturePlot, and Heatmap were applied. Additionally, the "Find All Markers" function was used for identifying cluster markers.

2.8. Cell experiment

The tumor tissue samples were obtained from the Neurosurgery Research Institute of Lanzhou University Second Hospital (LUSH), and the pathological information of the patients was provided by the Department of Pathology of LUSH. Total RNA was extracted using TRIzol reagent (Invitrogen, Thermo Fisher Scientific) following the manufacturer's instructions, and quantitative real-time polymerase chain reaction (qRT–PCR) was used to quantify RNA levels. Reverse transcription (RT) was performed using 5× ABScriptMix III and 20x gDNA remover MIX. Each PCR mixture had a total volume of 20 μl, followed by 45 cycles of PCR amplification. The $2^{-\Delta\Delta Ct}$ method was used to analyze the mRNA expression levels of each gene. The following PCR primers were utilized: LY96 forward primer: ACACCTACTGGGAGAGATTT; reverse primer: GTAGGATGAACAACAACATCACAGCAAA; C1QB forward primer: ACT-GATGTTGCTCCTGCTCC; reverse primer: ATGAGGTTCACGCACAGGTT.

2.9. Immunohistochemistry

Paraffin-embedded tissue sections were deparaffinized and rehydrated, followed by antigen retrieval using microwaves. The tissues were blocked with 0.3 % hydrogen peroxide solution for 10 min. The primary antibodies, C1QB (1:100, DF7283, Affinity) and LY96 (1:100, 11784-1-AP, PROTEINTECH), were added dropwise to the tissues and incubated overnight at 4 °C. The tissue samples were then incubated with biotinylated secondary antibodies at room temperature for 10 min. Following incubation, the samples were stained with diaminobenzidine (DAB) and hematoxylin.

2.10. Western blot analysis

Total protein was extracted from both the cancerous tissue of glioma patients and the normal brain tissue of patients utilizing the complete protein extraction kit protocol (Solarbio, China). After extraction, the proteins were subjected to SDS–PAGE electrophoresis, followed by transfer onto a PVDF membrane (Millipore, USA). The membrane was blocked with a 5 % skim milk solution in Tris-buffered saline/Tween (TBST) and then incubated overnight at 4 °C with primary antibodies. The following day, the PVDF membrane was cleansed with TBST before being incubated with goat anti-rabbit IgG and anti-mouse IgG for 1 h at room temperature. Finally, protein visibility was enhanced using the appropriate developing solution (Thermo Fisher Scientific, USA).

2.11. Statistical analysis

Statistical analysis was performed using the software R version 4.2.2 (http://www.r-project.org). The Wilcoxon test was utilized to assess differences between the two subgroups for continuous data. The Pearson correlation coefficient was used to evaluate the relationship between the two variables. Kaplan-Meier survival analysis was conducted to compare survival rates associated with different subgroups. $P < 0.05$ was considered the threshold of statistical significance.

3. Results

3.1. Subgroups and clinical correlation

The GSE199375 and GSE136000 datasets were downloaded from the Gene Expression Omnibus (GEO) database. Eight normal and 10 glioma tissue samples were obtained after batch effects were removed. A total of 50 marker gene sets were identified through the GSEA (https://www.gseamsigdb.org/gsea/index.jsp), and the enrichment score for each gene set was computed using the ssGSEA algorithm. Then, GSEA was performed on 18 samples, revealing the significant enrichment of EMT in gliomas, compared to normal samples (Fig. 2A). The 1,641 samples were divided into two subgroups based on the median enrichment scores of the EMT gene sets:

the EMT-hot subgroup and the EMT-cold subgroup (Fig. 2B). The overall survival rate of patients in the EMT-hot subgroup was 4 times lower than that of patients in the EMT-cold subgroup (Fig. 2H), with a significant difference between the two groups. A greater proportion of patients in the EMT-hot subgroup had high-grade glioma, and there was no significant difference between the two groups regarding sex (Fig. 2C–E). Moreover, patients in the EMT-hot subgroup tended to be older and have IDH-wild glioma. The 1p19q noncodeleted and MGMTp_Unmethylated gliomas were enriched in the EMT-cold subgroup. Additionally, the EMT-hot subgroup exhibited a greater proportion of high-grade gliomas (Fig. 2F and G). (WHO II (n = 399, low vs high = 291 (47 %): 235 (38 %), P < 0.001), WHO III (n = 472, low vs high = 91 (15 %): 108 (16 %), P < 0.001), WHO IV (n = 409, low vs high = 237 (36 %): 318 (48 %), P < 0.001)).

3.2. Relationship between EMT and multi-omics characteristics

The integration of multi-omics data is beneficial for investigating the overall epigenetic alterations in tumors, among which CNV is
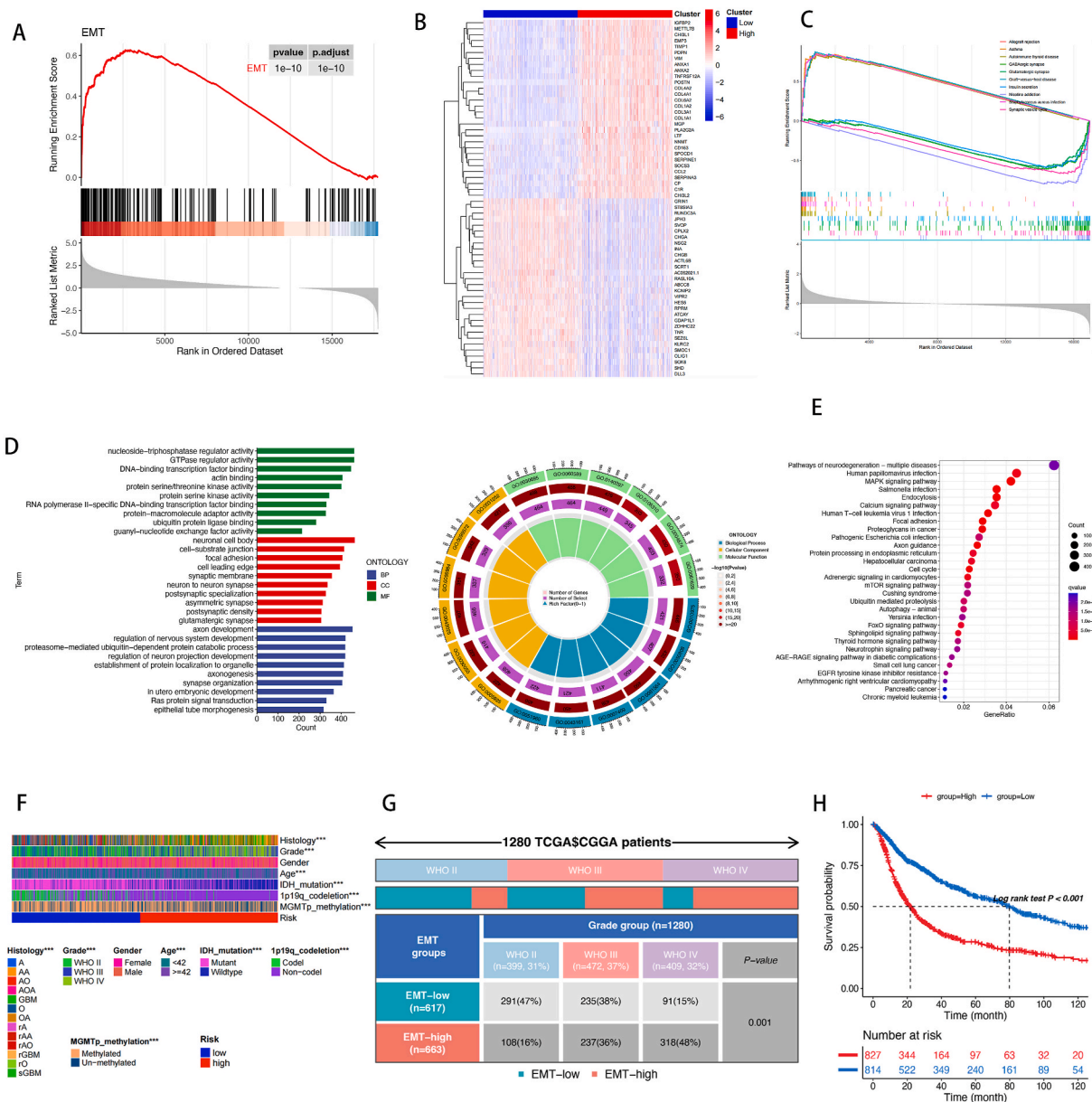

Fig. 2. EMT mediates glioma clinical trait transition. A. High EMT score in tumors. B. Differences between different subtypes gene heatmaps. C-E. Functional analysis between two subtypes. F. Proportion of clinical traits. G. WHO grade distribution. H. Survival differences. (Data are presented as means ± standard deviations; *indicates P < 0.05, **indicates P < 0.01, ***indicates P < 0.001, ****indicates P < 0.0001.)

a prevalent large-scale genetic alteration in cancer. Differences in CNVs between the two subgroups were evaluated using TCGA. T Compared with the EMT-cold subgroup, the EMT-hot subgroup exhibited significantly greater copy numbers (Fig. 3B). Additionally, compared with those in the EMT-cold subgroup, regions with significant copy number acquisition in the EMT-hot subgroup were primarily identified in the 7p11.2 region, while regions with missing copy numbers were identified in the 9q21.3 region (Fig. 3A–E). To understand how changes in the tumor mutational burden (TMB) might impact the proportions of immune and stromal cells in the tumor microenvironment (TME), we compared the TMB and TME between the two subgroups. The analysis revealed a greater TMB in the EMT-hot group than in the EMT-cold group (Fig. 3F–G), and increased TMB and EMT were found to be associated with poorer patient survival (Fig. 3H). Further analysis of the TME demonstrated lower tumor purity and a greater proportion of stromal cells in the EMT-hot subgroup than in the EMT-cold subgroup (Fig. 4A–B). The results of the Tumor Immune Dysfunction and Exclusion (TIDE) analysis results indicated that the gliomas in the EMT-hot group were more prone to evade the immune response and thereby showed poor responses to immunotherapy. Surprisingly, our analysis indicated that the EMT-hot subgroup had a greater proportion of MHC and HLA-related genes (Fig. 4C–D), which was linked to poorer survival in glioma patients (Fig. 4E–F). This finding may partly explain the occurrence of EMT and the transition from low-grade to high-grade glioma, which are associated with poor prognosis.

3.3. WGCNA to identify modules

The scale-free topology criteria were utilized to calculate soft thresholds. Then the optimal soft threshold power (β) value of 12 was determined from a phylogenetic tree (Fig. 5A–B) for network construction. Coexpression modules were then created (cutting height ≥0.25). Hierarchical clustering revealed similar gene expression patterns within modules belonging to the same branch (Fig. 5C–E).
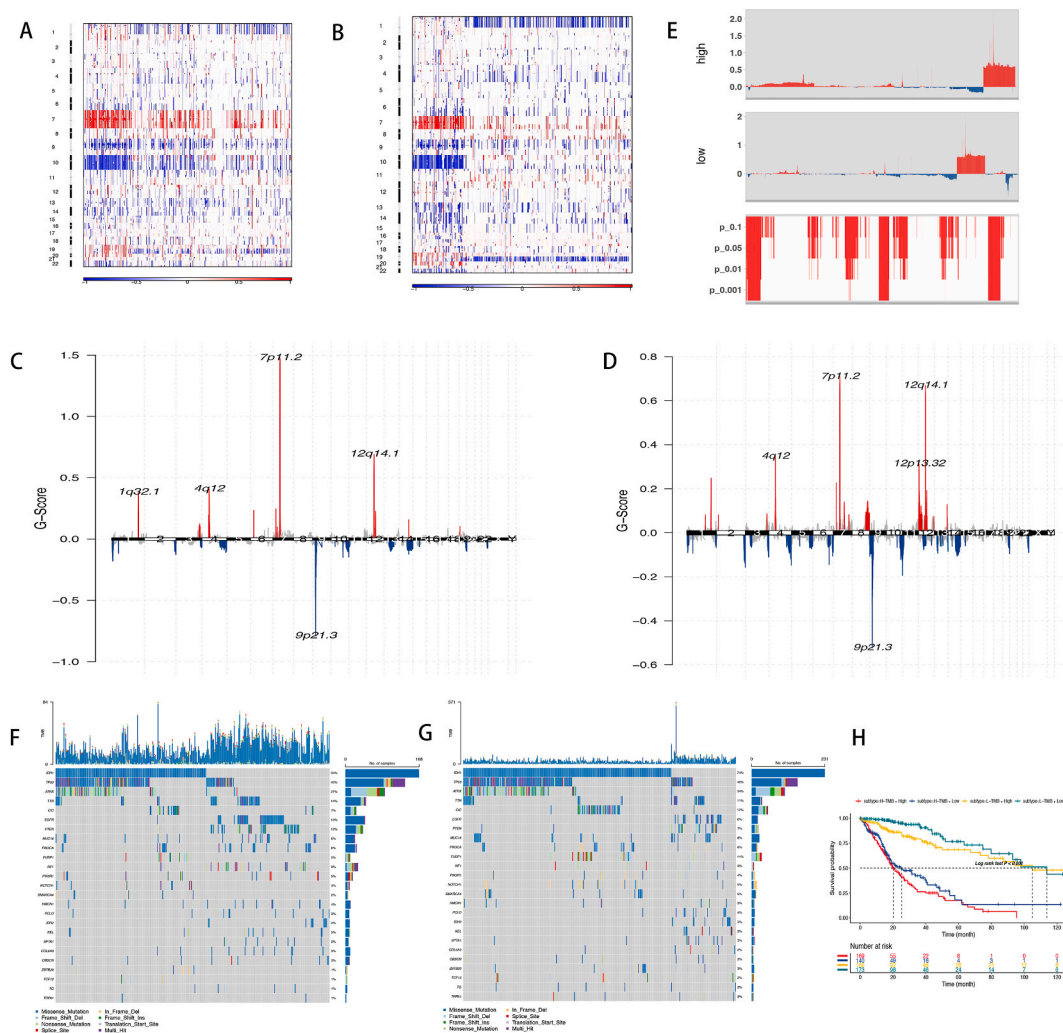

Fig. 3. Evaluation of CNV and TMB between two subtypes. A–E: Analysis of chromosome copy number variation between two isotypes (A, C are hot subtypes, B, D are cold subtypes). F: Manifestations of tumor mutational burden in EMT thermoisotypes. G: Manifestations of tumor mutational burden in EMT cold subtype. H: Survival curve of tumor mutation burden binding subtype.
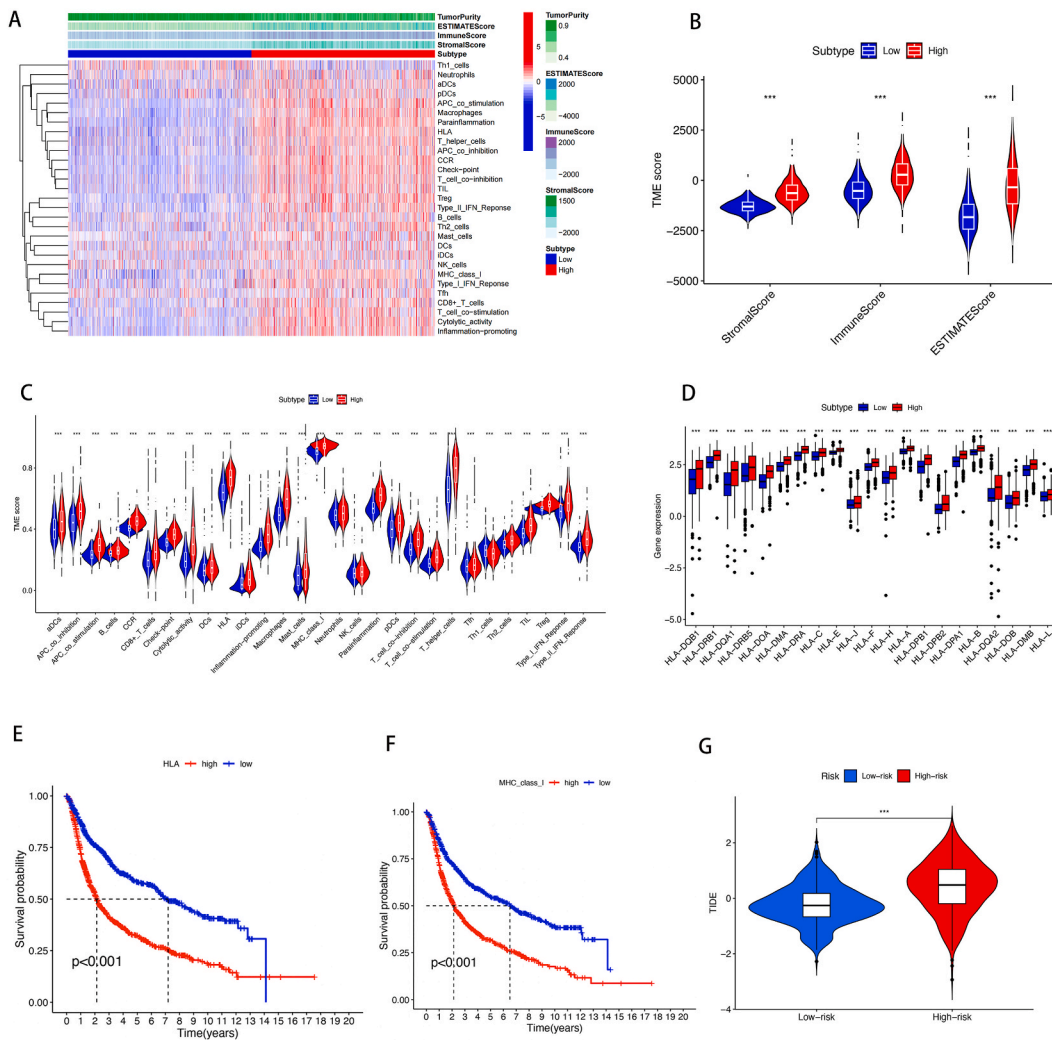
Fig. 4. Tumor microenvironment combined with immunotherapy and immune escape prediction. A: Heat map of the difference in tumor micro-environment scores in the two subtypes. B: Representation of stromal cells, immune cells, and overall scores in both subtypes. C: Overall difference in 28 immune-related function scores between the two subtypes. D: Differential expression of human leukocyte antigen-related genes between the two subtypes. E: Survival analysis of human leukocyte antigen score. F: Survival analysis of high and low histocompatibility complex scores. G:TIDE immune evasion and immunotherapy predict the efficacy of immunotherapy in two subtypes.

These modules, characterized by similar gene expression patterns, were combined to form nine coexpression modules, represented by the colors purple, red, royal blue, dark green, dark turquoise, blue, cyan, pink, and gray (Fig. 5F). Next, gene clusters were visualized to evaluate the correlation between modules. Notably, the blue and cyan modules showed a strong correlation with cluster A (Fig. 5G–H), with the cyan module highly correlated with the EMT hot subgroup (cor = 0.58, p = 1.5e-159), making it an important module for further research (Fig. 5I).

3.4. Dimensionality reduction and identification of feature genes

We extracted 2,385 module-related genes and 549 DEGs for dimensionality reduction for the identification of feature genes. Subsequently, a total of 473 genes related to the survival of patients with gliomas (P ≤ 0.001) were identified using the "survival" and "survminer" R packages, and a list of 1596 immune-related genes was compiled. Upon intersection, a total of 85 DEGs associated with high-level enrichment of EMT, the survival of patients with glioma, and the TME of glioma were identified. Then, survival models for the two glioma subgroups were constructed using the GSE184941 dataset (Fig. 6A–B), yielding 19 feature genes. Univariate and multivariate analyses demonstrated the good predictive ability of these genes (Fig. 6C–D). Additionally, we used the CGGA-301 dataset to confirm the prediction efficiency of the established score. The results indicated the good ability of the model to predict survival, as evidenced by univariate and multivariate analyses (Fig. SI–K).
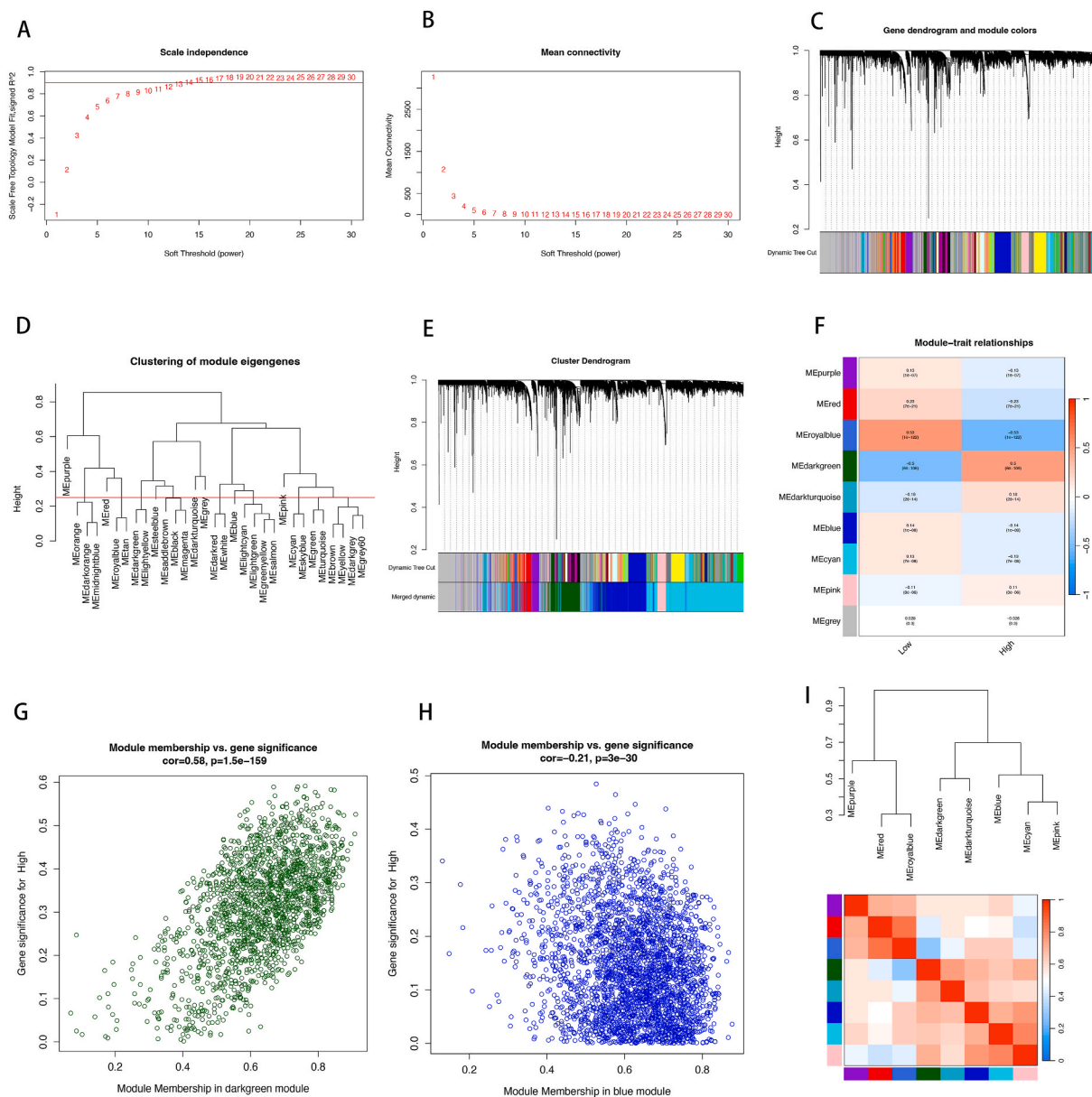
Fig. 5. WGCNA analysis for the two clusters. A: Confirming the best scale-free index for various soft-threshold powers (β). B: The mean connectivity for various soft-threshold powers. C:The gene tree map and nodule color. D: Hierarchical clustering analysis. E: The gene dendrogram is based on clustering. F: Heatmap of the correlation between the module genes and the two clusters. G-H:Correlation of gensets with EMT thermotypes. I:The heatmap of all genes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.5. ML

ML pertains to data learning that can help machines learn patterns from complex existing data to predict future outcomes and trends. In this study, 85 DEGs relevant to survival and immunity were analyzed using LASSO regression, RF, and SVM, of which 57 genes were identified by LASSO regression (Fig. 6E); subsequently, genes with an importance score greater than 5 from the RF model were selected for further analysis (Fig. 6F). After the intersection with the 20 genes selected by SVM, where 57 genes were identified through Lasso for further analysis, we obtained 14 core genes (Fig. 6G). The correlation between these 14 genes was then assessed (Fig. 6H), suggesting potential interactions between the identified core genes contributing to the onset of EMT in glioma.
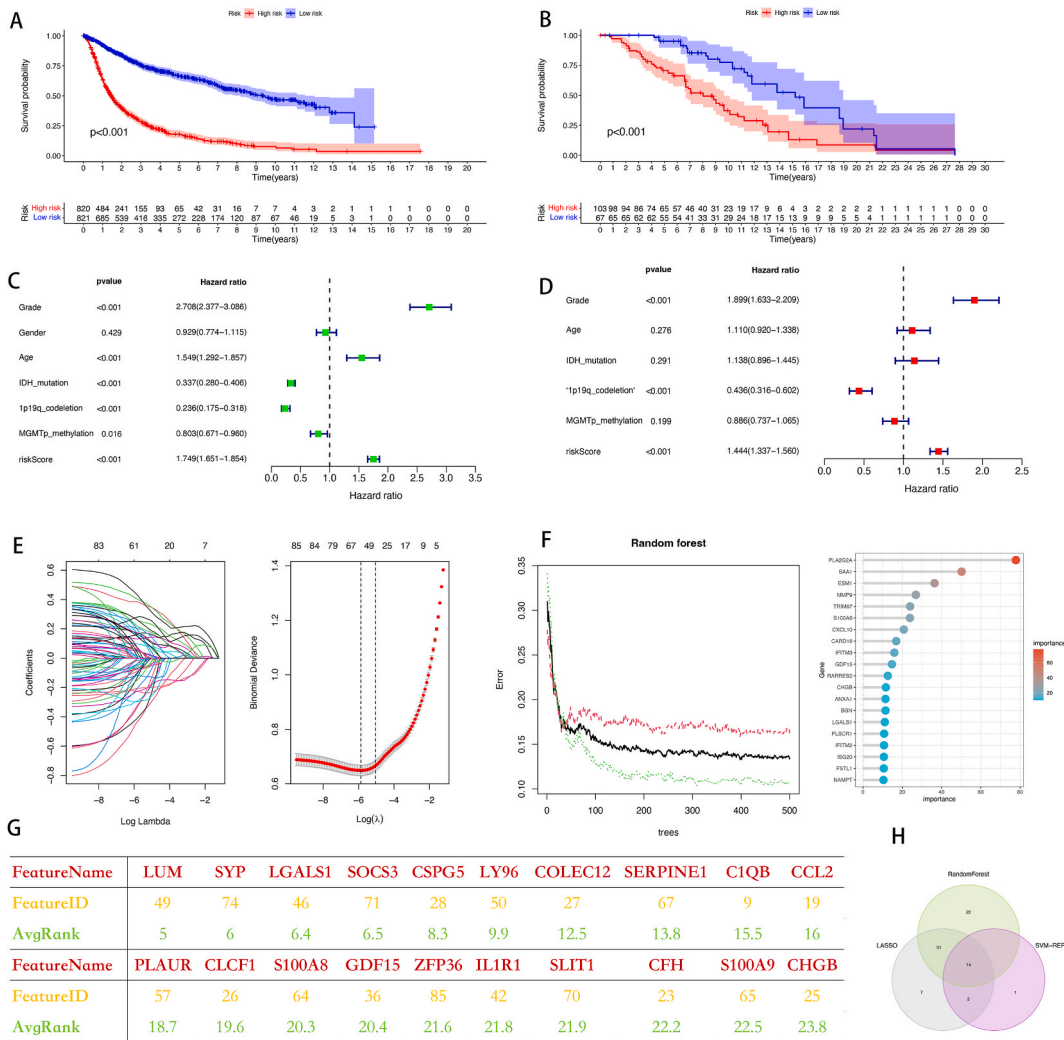
Fig. 6. Prognostic model construction, machine learning to find core genes. A. Prognostic model train group. B. Prognostic model test group. C. Univariate analysis. D. Multivariate analysis. E. lasso returns; F. Random forest. G. Support vector machine. H. Intersecting genes.

3.6. Possible association of EMT with tumor-associated macrophages

Initially, given the uniqueness of the 14 core genes, the scores of the 50 marker gene sets in the two subgroups and the scores of e genes in 50 gene sets were visualized (Fig. 7A–B). Then, the infiltration of 22 types of immune cells in the two EMT subgroups was evaluated using the "cibersort" package revealing increased infiltration of M2 macrophages in the EMT-hot group, as anticipated. Subsequently, we assessed the correlation between the 14 core genes and the 22 types of immune cells, along with the interplay between genes. The results indicated a strong correlation between LY96 and C1QB (Fig. 7G), as well as their association with M2 macrophages. PPI network analysis was conducted to identify the targets of tumor-associated macrophages. Finally, six genes, namely, LY96, C1QB, LGALS1, CSPG5, S100A8, and CHGB, were included in further analysis. Single-cell analysis suggested a potential association of these six genes with macrophages (Fig. 7H–J), consistent with the ability of marker genes, namely MSR1, CD163, MRC1, and CSF1R, to identify M2 macrophages. Notably, LY96 and C1QB have garnered particular attention due to their similarity to the marker genes for M2 macrophage. This resemblance in gene expression suggested that LY96 and C1QB may be involved in M2 macrophage functions or could be used as supplementary markers for identifying M2 macrophages.

3.7. Validation through experiments

Little work has investigated the roles of LY96 and C1QB in gliomas. In our study, the evaluation of these two genes revealed increased expression of these two genes in the EMT hot subgroup compared with the EMT-cold subgroup. Importantly, their expression levels were positively correlated with poor prognosis in patients with gliomas (FigS A-N). Moreover, increased expression of the two
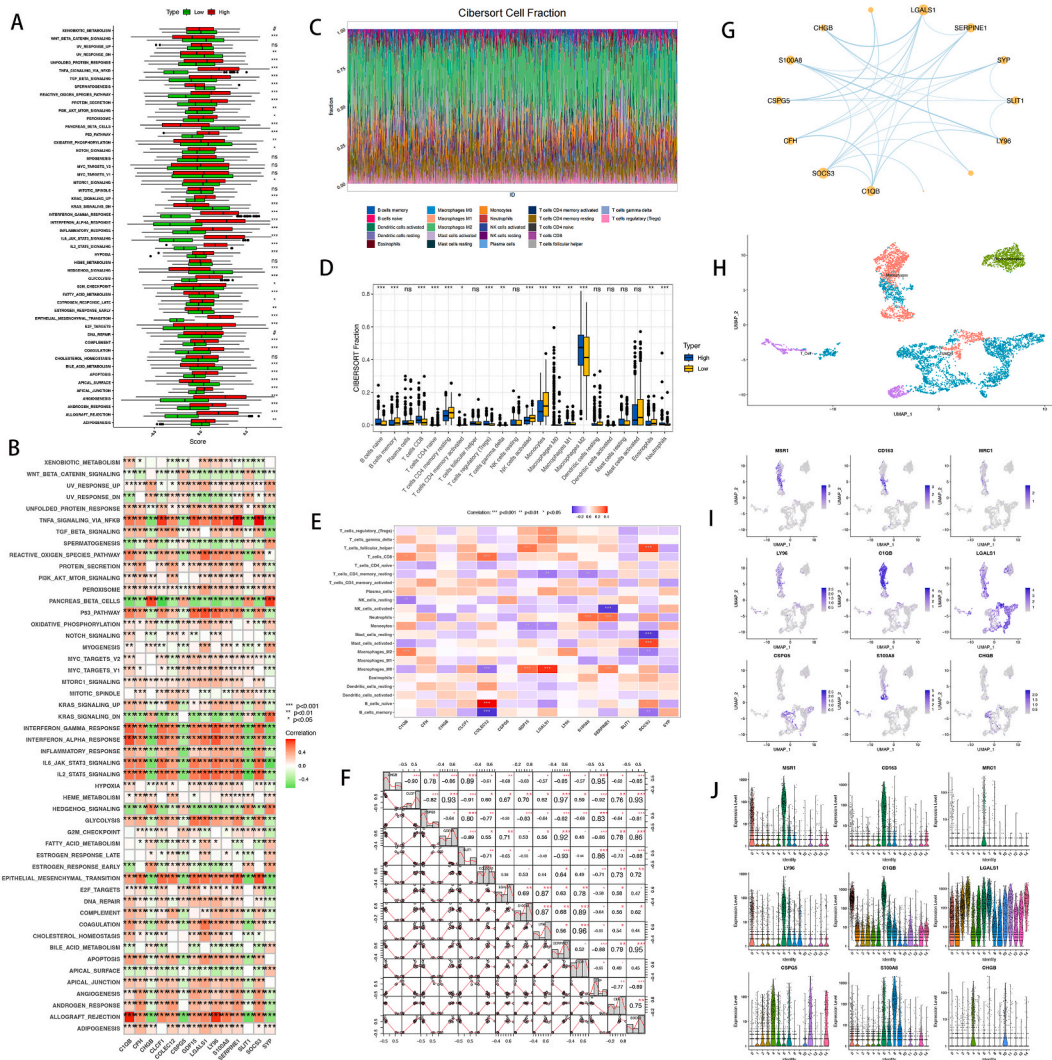
Fig. 7. The immunological landscape of core genes and typing. A. Performance of two subtypes in Hhallmark gene sets. B. Representation of core genes in Hhallmark gene sets. C. Distribution landscape of 22 types of immune cell infiltration in glioma. D. Difference box plot of subtypes on 22 immune cell infiltration. E. Differences in the performance of core genes on 22 types of immune cells. F. Correlation analysis of core genes. G. Core gene protein network interaction analysis. H. Single-cell Umap grouping. I-J. Distribution of specific genes between groups.

genes was found in patients with high-grade gliomas and in older patients (aged >42), and they exhibited heightened expression in noncodeleted (1p19q_codeletion) gliomas and unmethylated (MGMTp_methylation) gliomas (FigS A-N). This finding is consistent with the clinical characteristics observed in both glioma subgroups. Subsequently, experimental validation was conducted using glioma tissue samples. The qRT–PCR and Western blot analysis results demonstrated elevated expression of LY96 and C1QB in gliomas compared with normal tissue samples (Fig. 8A–C), thus corroborating our initial findings. Immunohistochemistry results further demonstrated a progressive increase in C1QB (Fig. 8D–F) and LY96 (Fig. 8G–I) expression from normal tissue to WHO II、 WHO III and WHO IV tissues (Fig. 8D–I).

4. Discussion

Gliomas, the most common primary malignant brain tumors, are characterized by poor prognosis and a tendency to cause disability, imposing great burdens on patients and society. Among gliomas, glioblastoma has the highest degree of malignancy. Despite the implementation of standard treatment approaches, such as surgical resection and concurrent radiotherapy and chemotherapy with temozolomide, the median survival of patients remains less than 15 months. Gliomas possess complex characteristics that distinguish them from tumors in other organs. Malignant glioma cells are characterized by robust invasiveness, leading to early infiltration of surrounding normal tissues. As a result, even following surgical excision of tumor tissue, recurrence at the margins of the surgical cavity is frequently inevitable [17–19].
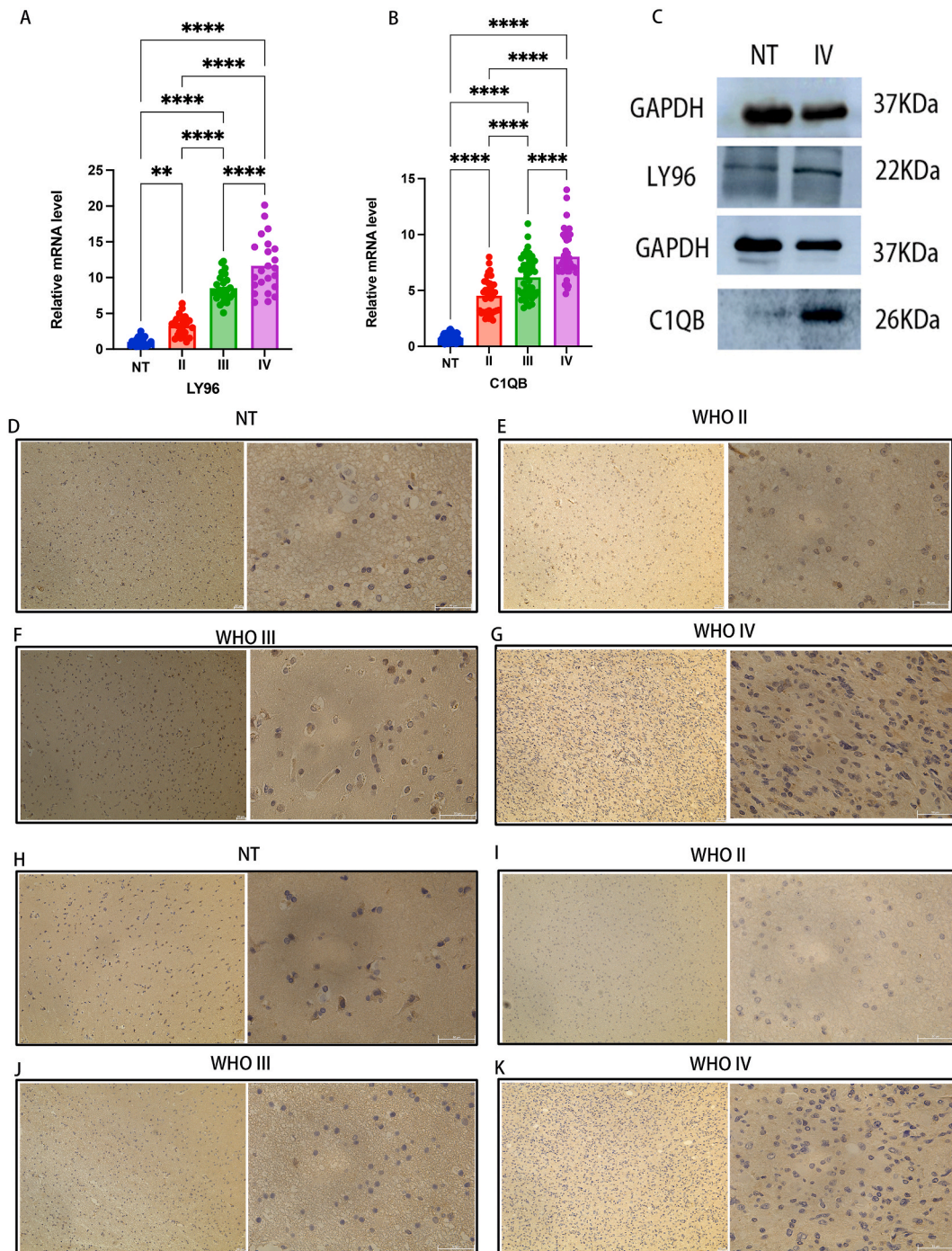
Fig. 8. Vitro experiments were performed to verify the expression of LY96 and C1QB. A:qPCR results showed that LY96 was highly expressed in gliomas. B:qPCR results showed that C1QB was highly expressed in gliomas. C: Western blot of LY96 and C1QB in normal tissue and tumor tissue. D-k: Immunohistochemistry staining of gliomas tissue sections with anti-human LY96 and C1QB antibody.

EMT is a cellular process wherein epithelial cells lose their epithelial characteristics and acquire mesenchymal traits [20]. Throughout this transformation, epithelial cells lose polarity and separate from each other to gain deformation and mobility, thereby enabling migration to distant tissues, which is marked by decreased expression of E-cadherin, a pivotal molecule for cell – cell adhesion in epithelial tissues, and increased expression of molecules facilitating mesenchymal characteristics, such as vimentin, N-cadherin, and Snail [21–24]. EMT is essential for normal tissue development and wound healing. It has also emerged as a driving force in the pathogenesis of cancer and fibrosis [20,25,26]. Notably, EMT has been identified as a pivotal contributor to the

development of malignant phenotypes, driving tumor cell invasion and metastasis [27–29] in various epithelial cancers, such as breast [30], liver [31], bladder [32], and lung [27] cancers. Previous studies [33] have underscored the importance of characterizing distinct disease subgroups, which is crucial for refining tumor classification parameters and devising personalized and efficacious treatment strategies. Such subclassification also aids in pinpointing malignant gliomas heavily reliant on EMT. If the concept of EMT-dependent GBM infiltration holds true, the application of effective EMT inhibitors, as explored in other systems, could revolutionize the treatment landscape for patients with malignant gliomas. Consequently, numerous scholars have dedicated significant efforts to delineating various therapeutic targets based on glioma classification, especially those related to EMT. Among these endeavors, studies have elucidated the role of GAMs as potent drivers of glioma cell invasion through their regulation of the EMT process [34], suggesting that GAMs are a prognostic tool. Concurrently, cellular experiments have identified ZNF207 as a potential promoter of glioma migration by modulating the EMT process [35], offering valuable insights into glioma pathogenesis and treatment strategies. Based on prior research, we used the EMT gene set as a distinguishing characteristic to investigate glioma classification and identify potential marker genes.

To elucidate the molecular mechanisms underlying the biological process of EMT in gliomas, we conducted a comprehensive analysis. Initially, we compared the expression of EMT-related gene sets between patients with tumors and healthy controls, revealing a significant enrichment of EMT gene sets in patients (Fig. 1A). Then, we used the ssGSEA algorithm to construct a scoring system to classify 1,641 glioma samples into two clusters: the EMT-hot subgroup and the EMT-cold subgroup. Subsequent investigations of these two subgroups highlighted significant differences between them. Compared with the EMT-cold subgroup, the EMT-hot subgroup exhibited more aggressive infiltration, a greater proportion of high-grade gliomas, and a tendency to be associated with elderly patients. Moreover, the worse prognosis observed in the EMT hot subgroup, which had a worse prognosis, may be attributed to the presence of IDH mutations, 1p19q combined deletion, and MGMTp_ methylation.

Moreover, a stronger association of EMT with the IL2_STAT5_SIGNALING, TNFA_SIGNALING_VIA_NFKB, and IL6_JAK_-STAT3_SIGNALING pathways was observed in the EMT-hot group (cor>0.7), which also displayed a greater proportion of tumor-associated macrophages than did the EMT-cold subgroup. In this regard, the activation of these pathways and cell infiltration may act as catalysts for EMT development. Furthermore, our evaluation of CNVs between the two subgroups indicated that significant copy number amplification was primarily concentrated in the 7p11.2 region within the EMT-hot subgroup. Previous studies have highlighted strong amplification at chromosome 7p11.2 in glioblastoma, with frequent amplification of two genes, SEC61c and EGFR. The EGFR gene, also known as c-erbB-1, encodes the epidermal growth factor receptor, representing the first member of the epidermal growth factor receptor family. Moreover, the TMB in two EMT subgroups was investigated in this study, as it is a hallmark of malignant tumors and a candidate for predicting immunotherapy outcomes. The EMT-hot subgroup exhibited a greater TMB than its cold counterpart, indicating the potential of the EMT subgroup as a method to evaluate immunotherapy for glioma. In this context, our TIDE analysis revealed the potential of EMT to promote immune escape in gliomas, thereby facilitating the failure of immunotherapy. Another important factor associated with glioma development is tumor purity. Our analysis revealed lower tumor purity in the EMT-hot subgroup than in the EMT-cold subgroup, which was associated with a poorer prognosis, consistent with previous findings. Furthermore, the EMT-hot subgroup exhibited significant enrichment of genes related to HLA and MHC, indicating potential immune-related effects. Therefore, we evaluated the expression of HLA-related genes (HLA-DOB, HLA-G, HLA-L, HLA-F, HLA-DPB1, HLA-DPA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DQA1, HLA-DRB5, HLA-DQB2, HLA-DDMA, HLA-A, HLA-H, HLA-DRB6, HLA-DRA, HLA-C, HLA-DRB1, HLA-J, HLA-E, HLA-DPB2, and HLA-DMB) in both groups. All of these genes exhibited increased expression in the EMT hot subgroup, which could be linked to the poor prognosis of glioma patients.

To gain insight into the mechanism underlying the aforementioned characteristics, WGCNA was conducted, and the "limma" R package was utilized to identify the modular genes and DEGs in the two subgroups, resulting in the identification of 85 survival and immune-related genes. These genes were used to construct survival models for the two EMT subgroups, from which feature genes for the genetic risk model were derived.

As a result, 14 genes were identified through intersection analysis using ML techniques including LASSO regression, SVM, and RF. Finally, six core genes were delineated through PPI analysis of the 14 genes. Notably, our single-cell data analysis suggested that these six core genes facilitated the infiltration of M2 macrophages. Furthermore, the expression of two out of the six genes was detected in patients with glioma via glioma tissue samples, and the results were consistent with our data analysis.

In summary, the present study identified EMT-related genes exhibiting distinct clinical and molecular characteristics in gliomas. The analysis findings underscore the potential of these six EMT-related genes as prognostic markers or therapeutic targets for glioma patients. Additionally, we have developed a risk profile based on EMT-related gene expression, enabling better prediction of overall survival for patients with gliomas. The precise mechanisms through which these EMT-related genes affect the prognosis of glioma patients remain elusive, highlighting the need for further research in this domain.

However, our current study has some limitations. The biggest limitation of the present study is that we did not fully confirm the correlation between molecular mechanism and EMT through in vitro experiments. The mechanism by which core genes promote EMT in gliomas has not been fully confirmed and deeply explored by in vitro experiments. Therefore, further experiments are needed to evaluate the underlying mechanisms we discovered in this study.

Consent for publication

All authors have provided their final approval for the version to be published, have agreed to the submission of the article to the journal, and have consented to assume responsibility for all aspects of the work.

Funding

Ethics approval and consent to participate

All patients included in the study signed an informed consent form, and the study was approved by the Medical Ethics Committee of the Second Hospital of Lanzhou University (No.2022A-515). The study was carried out in accordance with the recommendations contained in the Declaration of Helsinki. We adhere to Cell Press's Editorial Ethics Policy as well as Elsevier's Publication Ethics Policy. We ensure that all published research meets the journal's strict technical and ethical standards.

Data availability

This study utilized publicly available datasets, which can be accessed from the following sources: The China Glioma Genome Atlas (CGGA) database (http://www.cgga.org.cn), The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/), and Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The raw files for Western Blot and Immunohistochemistry used in the manuscript can be found in the Supplementary Materials.

CRediT authorship contribution statement

Peng Feng: Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Shangyu Liu:** Validation. **Guoqiang Yuan:** Writing – review & editing, Supervision, Funding acquisition. **Yawen Pan:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e34119.

References

[1] Q.T. Ostrom, H. Gittleman, P. Liao, et al., CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014, Neuro Oncol. 19 (suppl_5) (2017) v1–v88.
[2] J.C. Dewitt, A. Mock, D.N. Louis, The 2016 WHO classification of central nervous system tumors: what neurologists need to know, Curr. Opin. Neurol. 30 (6) (2017) 643–649.
[3] J.A. Schwartzbaum, J.L. Fisher, K.D. Aldape, et al., Epidemiology and molecular pathology of glioma, Nat. Clin. Pract. Neurol. 2 (9) (2006) 494–503.
[4] S.L. Gerlach, R.A. Dunlop, J.S. Metcalf, et al., Cyclotides chemosensitize glioblastoma cells to temozolomide, J. Nat. Prod. 85 (1) (2022) 34–46.
[5] J.G. Cairncross, M. Wang, R.B. Jenkins, et al., Benefit from procarbazine, lomustine, and vincristine in oligodendroglial tumors is associated with mutation of IDH, J. Clin. Oncol. 32 (8) (2014) 783–790.
[6] C. Neftel, J. Laffy, M.G. Filbin, et al., An integrative model of cellular states, plasticity, and genetics for glioblastoma, Cell 178 (4) (2019) 835–849.e21.
[7] A. Puisieux, T. Brabletz, J. Caramel, Oncogenic roles of EMT-inducing transcription factors, Nat. Cell Biol. 16 (6) (2014) 488–494.
[8] Z. Zhao, K.N. Zhang, Q. Wang, et al., Chinese glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese glioma patients, Dev. Reprod. Biol. 19 (1) (2021) 1–12.
[9] Z. Zhao, F. Meng, W. Wang, et al., Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas, Sci. Data 4 (1) (2017) 170024.
[10] Z.S. Bao, H.M. Chen, M.Y. Yang, et al., RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas, Genome Res. 24 (11) (2014) 1765–1773.
[11] K. Zhang, X. Liu, G. Li, et al., Clinical management and survival outcomes of patients with different molecular subtypes of diffuse gliomas in China (2011–2017): a multicenter retrospective study from CGGA, Cancer Biology & Medicine 19 (10) (2022) 1460–1476.
[12] Y. Wang, T. Qian, G. You, et al., Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping, Neuro Oncol. 17 (2) (2015) 282–288.

[13] X. Liu, Y. Li, Z. Qian, et al., A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas, Neuroimage: Clinical 20 (2018) 1070–1077.

[14] S. Grebinoski, Q. Zhang, A.R. Cillo, et al., Autoreactive CD8+ T cells are restrained by an exhaustion-like program that is maintained by LAG3, Nat. Immunol. 23 (6) (2022) 868–877.

[15] A.B. Khan, S. Lee, A.S. Harmanci, et al., CXCR4 expression is associated with proneural-to-mesenchymal transition in glioblastoma, Int. J. Cancer 152 (4) (2023) 713–724.

[16] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (1) (2008) 559.

[17] W. Wick, S. Dettmer, A. Berberich, et al., N2M2 (NOA-20) phase I/II trial of molecularly matched targeted therapies plus radiotherapy in patients with newly diagnosed non-MGMT hypermethylated glioblastoma, Neuro Oncol. 21 (1) (2019) 95–105.

[18] A. Berberich, T. Kessler, C.M. Thomé, et al., Targeting resistance against the MDM2 inhibitor RG7388 in glioblastoma cells by the MEK inhibitor trametinib, Clin. Cancer Res. 25 (1) (2019) 253–265.

[19] S. Lam, Y. Lin, P. Zinn, et al., Patient and treatment factors associated with survival among pediatric glioblastoma patients: a Surveillance, Epidemiology, and End Results study, J. Clin. Neurosci. 47 (2018) 285–293.

[20] J. Yang, P. Antin, G. Berx, et al., Guidelines and definitions for research on epithelial–mesenchymal transition, Nat. Rev. Mol. Cell Biol. 21 (6) (2020) 341–352.

[21] B.D. Craene, G. Berx, Regulatory networks defining EMT during cancer initiation and progression, Nat. Rev. Cancer 13 (2) (2013) 97–110.

[22] T.R. Samatov, A.G. Tonevitsky, U. Schumacher, Epithelial-mesenchymal transition: focus on metastatic cascade, alternative splicing, non-coding RNAs and modulating compounds, Mol. Cancer 12 (1) (2013) 107.

[23] D.S Jeevan, J.B. Cooper, A. Braun, et al., Molecular pathways mediating metastases to the brain via Epithelial-to-Mesenchymal transition: genes, proteins, and functional analysis, Anticancer Res 36 (2) (2016) 523–532.

[24] A. Toll, E. Masferrer, M.E. Hernández-Ruiz, et al., Epithelial to mesenchymal transition markers are associated with an increased metastatic risk in primary cutaneous squamous cell carcinomas but are attenuated in lymph node metastases, J. Dermatol. Sci. 72 (2) (2013) 93–102.

[25] H. Peinado, D. Olmeda, C.A.N.O.A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat. Rev. Cancer 7 (6) (2007) 415–428.

[26] Y. Ding, K. Gelfenbeyn, L. Freire-De-Lima, et al., Induction of epithelial-mesenchymal transition with O-glycosylated oncofetal fibronectin, FEBS (Fed. Eur. Biochem. Soc.) Lett. 586 (13) (2012) 1813–1820.

[27] M.L. Franco-Chuaire, M.C. Sánchez-Corredor, L. Chuaire-Noack, Epithelial-mesenchymal Transition (EMT): Principles and Clinical Impact in Cancer Therapy, Investigación Clínica, 2013, p. 54.

[28] C. Foroni, M. Broggini, D. Generali, et al., Epithelial–mesenchymal transition and breast cancer: role, molecular mechanisms and clinical impact, Cancer Treat Rev. 38 (6) (2012) 689–697.

[29] M. Iwatsuki, K. Mimori, T. Yokobori, et al., Epithelial-mesenchymal transition in cancer development and its clinical significance, Cancer Sci. 101 (2) (2010) 293–299.

[30] G. Jiang, H. Fang, X. Shang, et al., CHFR-mediated epithelial-to-mesenchymal transition promotes metastasis in human breast cancer cells, Mol. Med. Rep. 23 (6) (2021) 451.

[31] Y. Lei, W. Yan, Z. Lin, et al., Comprehensive analysis of partial epithelial mesenchymal transition-related genes in hepatocellular carcinoma, J. Cell Mol. Med. 25 (1) (2021) 448–462.

[32] S. Monteiro-Reis, J. Lobo, R. Henrique, et al., Epigenetic mechanisms influencing epithelial to mesenchymal transition in bladder cancer, Int. J. Mol. Sci. 20 (2) (2019) 297.

[33] U.D. Kahlert, G. Nikkhah, J. Maciaczyk, Epithelial-to-mesenchymal(-like) transition as a relevant molecular event in malignant gliomas, Cancer Lett. 331 (2) (2013) 131–138.

[34] X. He, Y. Guo, C. Yu, et al., Epithelial-mesenchymal transition is the main way in which glioma-associated microglia/macrophages promote glioma progression, Front. Immunol. 14 (2023) 1097880.

[35] C. Zhao, Y. Guo, Y. Chen, et al., Zinc finger Protein207 orchestrates glioma migration through regulation of epithelial-mesenchymal transition, 2024.