frontiers | Frontiers in Pharmacology

Integrating bulk and single-cell RNA sequencing data reveals epithelial-mesenchymal transition molecular subtype and signature to predict prognosis, immunotherapy efficacy, and drug candidates in low-grade gliomas

Chengcheng Wang[1] and Zheng He[2]*

[1]Department of Pharmacy, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, Qingdao, Shandong, China, [2]Department of Neurosurgery, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, Qingdao, Shandong, China

Objective: Epithelial-mesenchymal transition (EMT) is a tightly regulated and dynamic process occurring in both embryonic development and tumor progression. Our study aimed to comprehensively explore the molecular subtypes, immune landscape, and prognostic signature based on EMT-related genes in low-grade gliomas (LGG) in order to facilitate treatment decision-making and drug discovery.

Methods: We curated EMT-related genes and performed molecular subtyping with consensus clustering algorithm to determine EMT expression patterns in LGG. The infiltration level of diverse immune cell subsets was evaluated by implementing the single-sample gene set enrichment analysis (ssGSEA) and ESTIMATE algorithms. The distinctions in clinical characteristics, mutation landscape, and immune tumor microenvironment (TME) among the subtypes were subjected to further investigation. Gene Set Variation Analysis (GSVA) was performed to explore the biological pathways that were involved in subtypes. The chemo drug sensitivity and immunotherapy of subtypes were estimated through GDSC database and NTP algorithm. To detect EMT subtype-related prognostic gene modules, the analysis of weighted gene co-expression network (WGCNA) was performed. The LASSO algorithm was utilized to construct a prognostic risk model, and its efficacy was verified through an independent CGGA dataset. Finally, the expression of the hub genes from the prognostic model was evaluated through the single-cell dataset and *in-vitro* experiment.

Results: The TCGA-LGG dataset revealed the creation of two molecular subtypes that presented different prognoses, clinical implications, TME, mutation landscapes, chemotherapy, and immunotherapy. A three-gene signature (SLC39A1, CTSA and CLIC1) based on EMT expression pattern were established through WGCNA analysis. Low-risk patients showed a positive outlook, increased immune cell presence, and higher expression of immune checkpoint proteins. In addition, several promising drugs, including birinapant, fluvastatin, clofarabine, dasatinib, tanespimycin, TAK−733, GDC−0152, AZD8330, trametinib and ingenol-

mebutate had great potential to the treatment of high risk patients. Finally, CTSA and CLIC1 were highly expressed in monocyte cell through single-cell RNA sequencing analysis.

Conclusion: Our research revealed non-negligible role of EMT in the TME diversity and complexity of LGG. A prognostic signature may contribute to the personalized treatment and prognostic determination.

1 Introduction

The efficacy of cancer chemotherapy and immunotherapy is often impeded by inter-patient and intra-tumor heterogeneity, as well as multifactorial drug resistance. Low-grade gliomas (LGGs), being a form of malignant brain tumor, are also confronted with the challenge of tumor heterogeneity and resistance to treatment (Weller et al., 2015). Even after undergoing standard surgical resection followed by radiotherapy and chemotherapy, individuals diagnosed with low-grade gliomas continue to have a bleak prognosis, with an average survival rate ranging from 2 to 10 years (Golub et al., 2019). Despite extensive efforts to improve clinical outcomes, more than half of LGG cases progress and evolve into therapy-resistant high-grade aggressive glioma over time (Claus et al., 2015). One of the primary objectives of precision medicine is to precisely identify patients who are likely to benefit from personalized treatment and to prescribe treatments that are tailored to the unique characteristics of their tumors, with the aim of achieving optimal therapeutic outcomes. Hence, in the realm of precision medicine, it is imperative to establish a more precise categorization of tumors to eradicate the heterogeneity and resistance to treatment associated with LGGs.
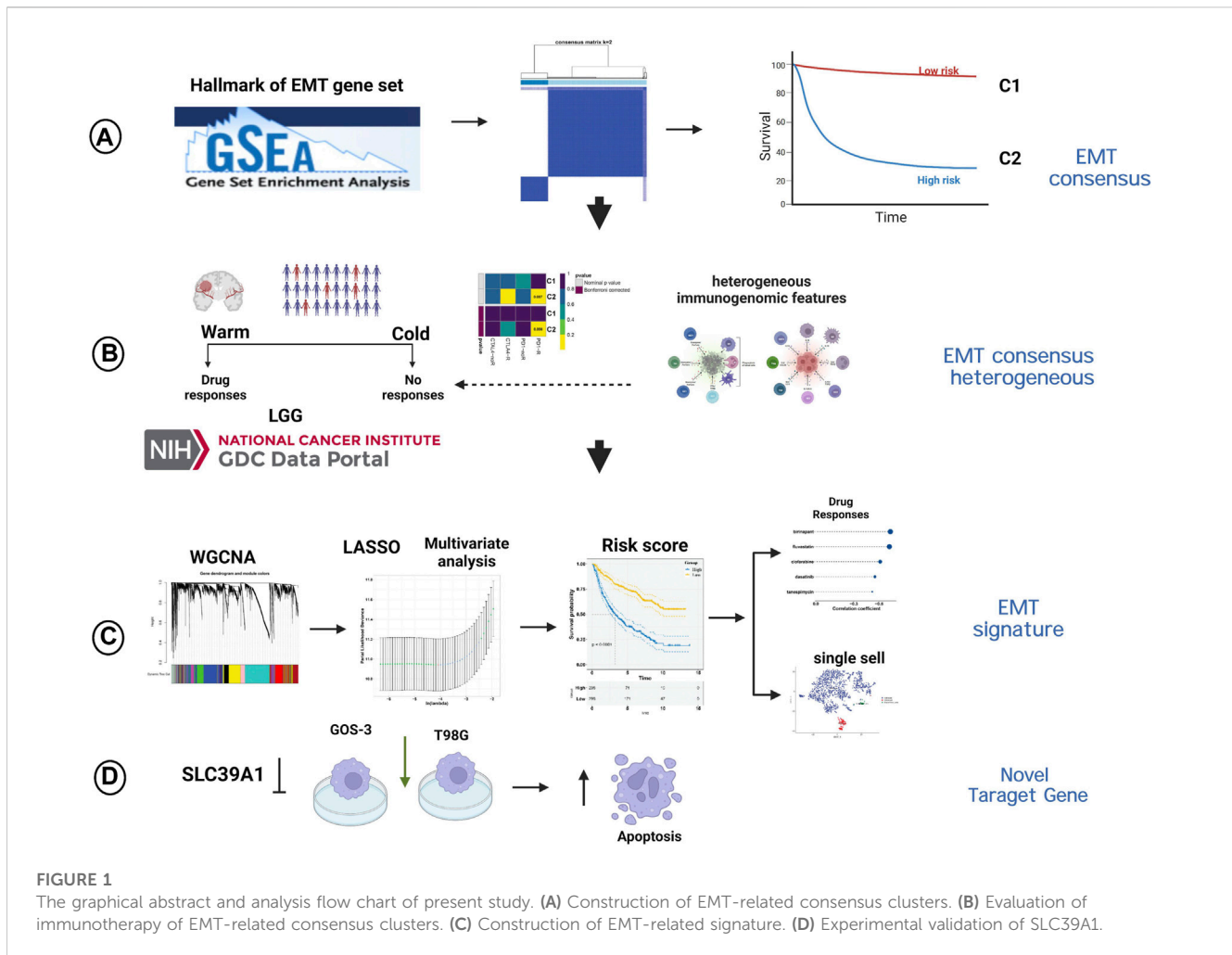
The process of epithelial-mesenchymal transition (EMT) is an essential differentiation program that is necessary for the development of tissues during embryogenesis. During this process, cells lose their epithelial characteristics and acquire mesenchymal migration properties (Yang et al., 2020). Aberrant activation of the EMT process is frequently observed in tumor proliferation, leading to the development of resistance towards conventional therapeutic interventions (Lambert et al., 2017; Shibue and Weinberg, 2017). EMT has the capability to enhance the potential of tumor cells for invasion and metastasis by facilitating their migration, disturbing cell-cell connections, disintegrating the basement membrane, and restructuring the extracellular matrix (ECM) (Miyoshi et al., 2004). Moreover, the process of EMT is linked with a higher percentage of cancer stem cells, reduced immune response against tumors and the development of resistance towards treatment (Akalay et al., 2013; Rhim et al., 2014), the study of therapeutic resistance from the perspective of EMT is one of cancer research focus. Thus, the presence of EMT, which is a hallmark of cancer, has been shown to be intimately linked with tumor invasion, metastasis, and the acquisition of chemotherapy resistance, all of which are critical biological processes in cancer development (van Staalduinen et al., 2018; Loret et al., 2019). Despite the existence of variations among tumor subtypes, the EMT process in LGG could potentially hold significant prognostic and molecular typing value. As a result of the highly hostile and intrusive characteristics of gliomas, it has been progressively acknowledged that the occurrence of EMT in gliomas may hold significant significance in the progression of glioma and the restructuring of the glioma microenvironment (Lu et al., 2012). Given the emergence of innovative immunotherapy techniques that offer groundbreaking cancer treatment alternatives, it is crucial to ascertain the immune profile of distinct EMT expression patterns, and their responsiveness to immunotherapy (Yamaguchi, 2016). In our study, while we have primarily relied on computational analyses of existing datasets, we have also performed experimental validation of the identified key genes. This approach underscores the critical role of EMT-related genes in the context of low-grade gliomas (LGG) and further highlights their significance in future research. By combining computational insights with experimental evidence, we aim to provide a more comprehensive understanding of the role of EMT in the expression, prognosis, immune tumor microenvironment (TME), clinical implications, and personalized treatment strategies for LGG.

2 Materials and methods

2.1 Data acquisition

Transcriptome data, mutation data and clinicopathologic characteristics of LGG samples were retrieved from the Cancer Genome Atlas (TCGA; https://www. cancer. gov/tcga/) through the Genomic Data Commons data portal (GDC; https://portal. gdc.cancer.gov/). In addition, the aforementioned information were obtained from two cohorts (mRNAseq_693, Illumina HiSeq Platform; mRNAseq_325, Illumina HiSeq (2000) and (2,500) platforms) in Chinese Glioma Genome Atlas (CGGA; http://www.cgga.org.cn), and serve as an external dataset. "SVA" (Leek et al., 2012) package was then used to merged the two cohorts after removing the batch effects ("ComBat" algorithm). After converting the raw read count, the values were expressed as transcripts per kilobase million (TPM). The Molecular signatures database (MSigDB; http://www.broadinstitute.org/msigdb) (Liberzon et al., 2015) provided a collection of 200 genes related to EMT, obtained from the gene set named "HALLMARK_EPITHELIAL_ MESENCHYMAL_TRANSITION". Figure 1 showed the graphical abstract and analysis flow chart of present study.

FIGURE 1
The graphical abstract and analysis flow chart of present study. **(A)** Construction of EMT-related consensus clusters. **(B)** Evaluation of immunotherapy of EMT-related consensus clusters. **(C)** Construction of EMT-related signature. **(D)** Experimental validation of SLC39A1.

2.2 Unsupervised clustering analysis

To perform an unsupervised cluster analysis on mRNA expression profiles of genes associated with EMT, the "ConsensusClusterPlus" R package was created (Wilkerson and Hayes, 2010). The optimal number of clusters was determined using the cumulative distribution function (CDF) and consensus matrix through resampling analysis to achieve the best performance. To assess the disparity in survival among clusters, the Kaplan-Meier (KM) analysis was employed. The results displayed the top 20 mutated genes in each cluster after visualizing single-nucleotide polymorphisms (SNP) variations using the R package "maftools" (Mayakonda et al., 2018). The GISTIC2.0 tool from Genome Data Analysis Center (GDAC) Firehose (https://gdac.broadinstitute.org) was used to analyze the copy number alterations (CNA) data among subtypes.

2.3 Landscape of immune TME

To calculate the ratio of immune cells in LGG samples, the Estimation of Stromal and Immune cells in Malignant Tumours using Expression (ESTIMATE) data method was utilized to compute the stromal and immune scores (Yoshihara et al., 2013).

The estimation of tumor purity was done by combining stromal and immune scores. The violin plot displayed variations in immune score, stromal score, and tumor purity across various subtypes.

The single-sample gene set enrichment analysis (ssGSEA) algorithm was utilized to measure the proportion of immune cell infiltration in the immune TME of LGG samples through the computation of enrichment scores. To compute the abundance of 28 tumor infiltrating immune cells (TIICs), a sum of 782 marker genes was collected from Bindea et al. (Bindea et al., 2013) and Broad Institute (http://software.broadinstitute.org/gsea/msigdb/index.jsp). Boxplots were used to present the results of ssGSEA analysis, which compared the level and function of immune infiltration among different subtypes using the gene set variation analysis (GSVA) software "GSVA" R package.

2.4 GSVA and functional annotation

To evaluate dissimilarities in biological processes among various clusters, the R software package "GSVA" was utilized for conducting GSVA analysis. Based on transcriptome data, GSVA is an unsupervised and nonparametric gene enrichment method that estimates alterations in the activity of biological processes and pathways in samples (Hänzelmann et al., 2013). To run GSVA

analysis, the gene set of "C2. cp.kegg.v7.1″was downloaded from MSigDB database. The "clusterProfiler" R software package was utilized to conduct functional annotation of genes related to EMT. Significantly enriched pathways were identified through screening for those with an adjusted p-value < 0.05 and false discovery rate (FDR) < 0.05.

2.5 Therapy response prediction of subtype

The tumor immune dysfunction and exclusion (TIDE) algorithm (Jiang et al., 2018) was utilized to approximate the efficacy of immunotherapy based on tumor immune dysfunction and exclusion. According to the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) (Yang et al., 2012), we predict chemotherapy response of each LGG sample. After performing simple processing of data (such as elimination of low-variation genes and summarize duplicated gene expression data into averages), we estimated the sensitivity of eight chemotherapeutic agents (cisplatin, erlotinib, methotrexate, vincristine, carmustine, temozolomide (TMZ), rapamycin, and doxorubicin) currently under study or widely applied in gliomas by SubMap analysis (Gene Pattern) of GDSC data (Hoshida et al., 2007). The ridge regression was performed to estimate the half-maximal inhibitory concentration (IC50) for LGG by "pRRophetic" R software package (Geeleher et al., 2014a), and the quantitative prediction accuracy was verified by 10-fold cross validation based on the GDSC training set (Geeleher et al., 2014b).

2.6 Weighted correlation network analysis (WGCNA) to identify hub genes from subtypes

For the purpose of identifying subtype associated genes, we adopted WGCNA algorithm through the "WGCNA" R package (Version: 1.71) (Langfelder et al., 2009). At first, genes that exhibited a variance value exceeding 25% were selected for the construction of the coexpression network (Yang and Xu, 2021; Zhou et al., 2021; Feng et al., 2022). Subsequently, the outlier samples were excluded through the "goodSampleGenes" function, and the soft-thresholding value $\beta = 5$ (scale free = 0.85) was applied to ensure a scale-free network. Then, a gene clustering tree was generated based on the computed adjacency among genes, followed by the categorization of genes into distinct modules comprising no less than 100 genes exhibiting similar characteristics within each module. The modules that exhibited most correlation with subtypes were selected to the downstream analysis, such as KEGG analysis. Moreover, the hub genes were screened based on the cutoff gene significance (GS) > 0.6 and module membership (MM) > 0.8. Pathway enrichment analysis on the modules was performed via the "clusterProfiler" R package (Wu et al., 2021).

2.7 Development of EMT-related signature

The "glmnet" package was employed to conduct a LASSO regression analysis on genes associated with prognostic EMT subtyping to evaluate their effect on prognosis. The analysis was subjected to ten-fold cross-validation, and the genes were chosen based on the point with the lowest error rate. Patients were categorized into low or high EMT score subgroups using the median score, which was calculated by merging gene expression and coefficients to compute the EMT risk score. KM survival curves were generated by performing survival analysis with the aid of the "survival" and "survminer" packages. To generate receiver operating characteristic (ROC) curves the "timeROC" package was employed. Prognostic variables were subjected to uni- and multivariate-cox regression analyses. A nomogram was constructed using the "rms" package to estimate survival probability, and calibration curves were used to evaluate the accuracy of the predictions.

2.8 Drug discovery of EMT-related signature

The response of human cancer cell lines to small molecule compounds was estimated using drug sensitivity profiles obtained from either the Cancer Therapeutics Response Portal (CTRP) (Basu et al., 2013) or the PRISM project (Corsello et al., 2020).

## 2.9 Single cell RNA-sequencing (scRNA-seq) revealed the expression level of genes

The GSE202096 dataset provided scRNA-seq data for one LGG samples (GSM6094425) (Chen et al., 2022). After conducting quality control, cells that had more than 20% mitochondrial UMI counts were eliminated using the "Seurat" package (Cao et al., 2022). The selection process involved choosing the top 1,500 genes with high variability, followed by clustering cell populations via the FindClusters function and then mapping them into t-distributed stochastic neighbor embedding (t-SNE). Using the FindAllMarkers function, the markers for each cell cluster were determined. Then, the CellMarker database was utilized to annotate cells based on their cell markers (Hu et al., 2022).

2.10 Cell culture and transfection

The cultivation of HMC3, GOS-3, T98G, and LN-18 cells (ATCC) was carried out in DMEM supplemented with 10% fetal bovine serum (Sigma Aldrich) under an atmosphere of 5% CO_2 at 37°C. After cloning siRNA targeting SLC39A1 into lentiviral vectors, the vectors were transfected into GOS-3 and T98G cells for 3 days and then exposed to 4 μg/mL puromycin for 1 week.

2.11 Real-time quantitative PCR (RT-qPCR)

Total RNA was extracted using TRIzol (Beyotime) and cDNA was synthesized with the PrimeScript RT reagent kit after gDNA Eraser (Takara) treatment. SYBR Green II Mixture (TaKaRa) was used for RT-qPCR, and GAPDH was used as an internal reference to calculate the expression using the $2^{-\Delta\Delta CT}$ method.
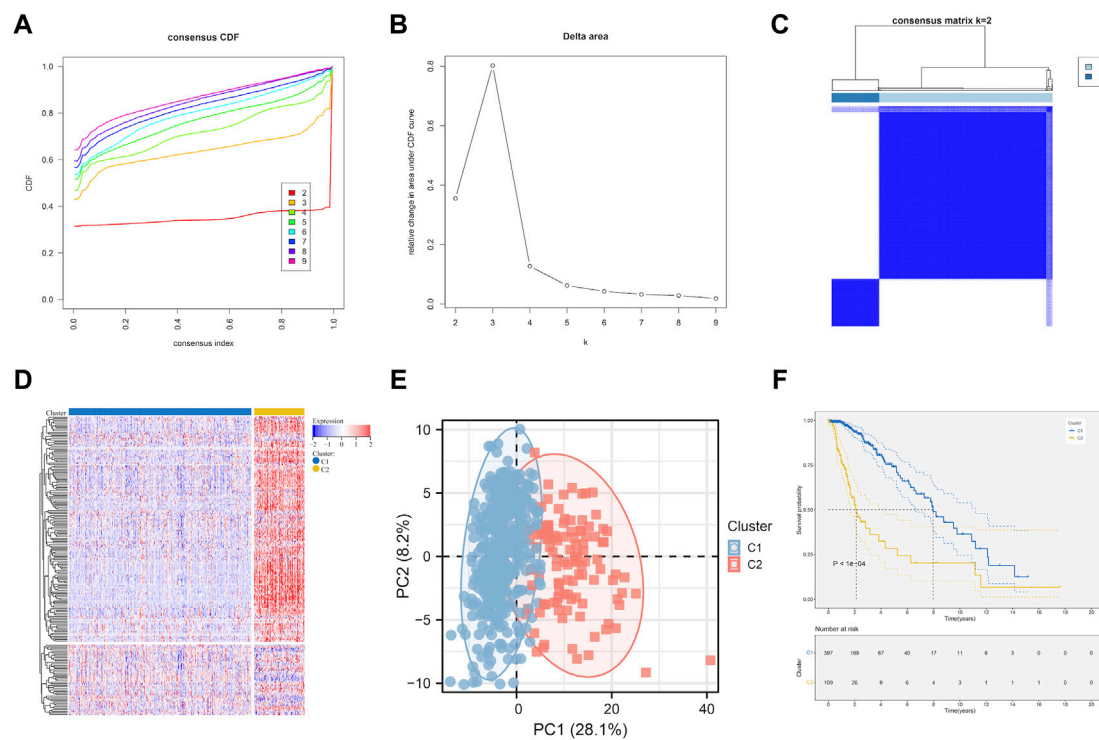
FIGURE 2
EMT-related genes could distinguish LGG in TCGA with different clinical and molecular features. **(A)** Relative change in area under CDF curve for k = 2 to k = 9. **(B)** Delta curve analysis from k = 2 to k = 9. **(C)** Consensus clustering matrix heatmap plots of 506 samples from TCGA datasets for k = 2. **(D)** Expression of EMT-related genes in molecular subtypes. **(E)** PCA analysis of the EMT-related genes expression when k = 2. **(F)** Kaplan-Meier analysis of patients between two subtypes. CDF, cumulative distribution function; PCA, principal components analysis.

2.12 Flow cytometry

Cell apoptosis was detected through flow cytometry and the Cell Apoptosis Detection Kit (KTA0002; Abbkine) was utilized. To summarize, a group of cells was exposed to 5 μL of Annexin V-AbFluor™ 488 binding and 2 μL of PI for 15 min at room temperature while being shielded from light. After adding 400 μL of 1 x Annexin V buffer, the level of apoptosis was assessed using a Beckman Flow cytometer.

2.13 Statistical analysis

The R language (R-project.org) platform and software package from the Bioconductor project (www.bioconductor.org) (R Core Team, Version 4.0.2) were used for statistical analysis and visualization of results in the current study. To compare the two groups and more groups, the Wilcoxon and Kruskal–Wallis tests were utilized, respectively (Hazra and Gogtay, 2016). To detect the differentially expressed genes (DEGs) between the two subtypes, the one-way ANOVA and Tukey's test were conducted with a q-value of less than 0.05 and an absolute value of log2FC greater than 2. The KM curve, analyzed by log-rank test, exhibited the variations in operating systems among different groups. Bilateral p values were used, with a statistically significant difference defined as $p < 0.05$.

3 Results

3.1 Clinicopathological characteristics of two LGG molecular subtypes based on EMT related genes

Based on the prognostic EMT gene expression profile, LGG samples was divided into two molecular subtypes by unsupervised clustering analysis (Figures 2A–C). The EMT gene expression were presented a significant divergence between two subtypes (Figure 2D). Figure 2E showed that the principal components analysis (PCA) distribution patterns were mostly in agreement with the two subtypes designations. Figure 2F suggested the overall prognosis of C2 is worse than that of C1 ($p < 0.0001$). Moreover, we further explore the relationship between subtypes and clinical traits.

As Figures 3A–G showed, the patients survival status (Figure 3A), age (Figure 3B), grade (Figure 3D), IDH status (Figure 3E), 1p19q status (Figure 3F) and methylation status (Figure 3G) were showed a significant difference ($p < 0.05$) between C1 and C2, while no difference was observed in gender (Figure 3C). Subtype C2 corresponding to more deaths, elderly patients, higher grade, 1p19q non-codel, IDH wild type, and unmethylated cases.

We assessed the function of each subtype through GSVA analysis (Supplementary Figure S1) to delve deeper into the potential biological process of the two clusters. Results suggested
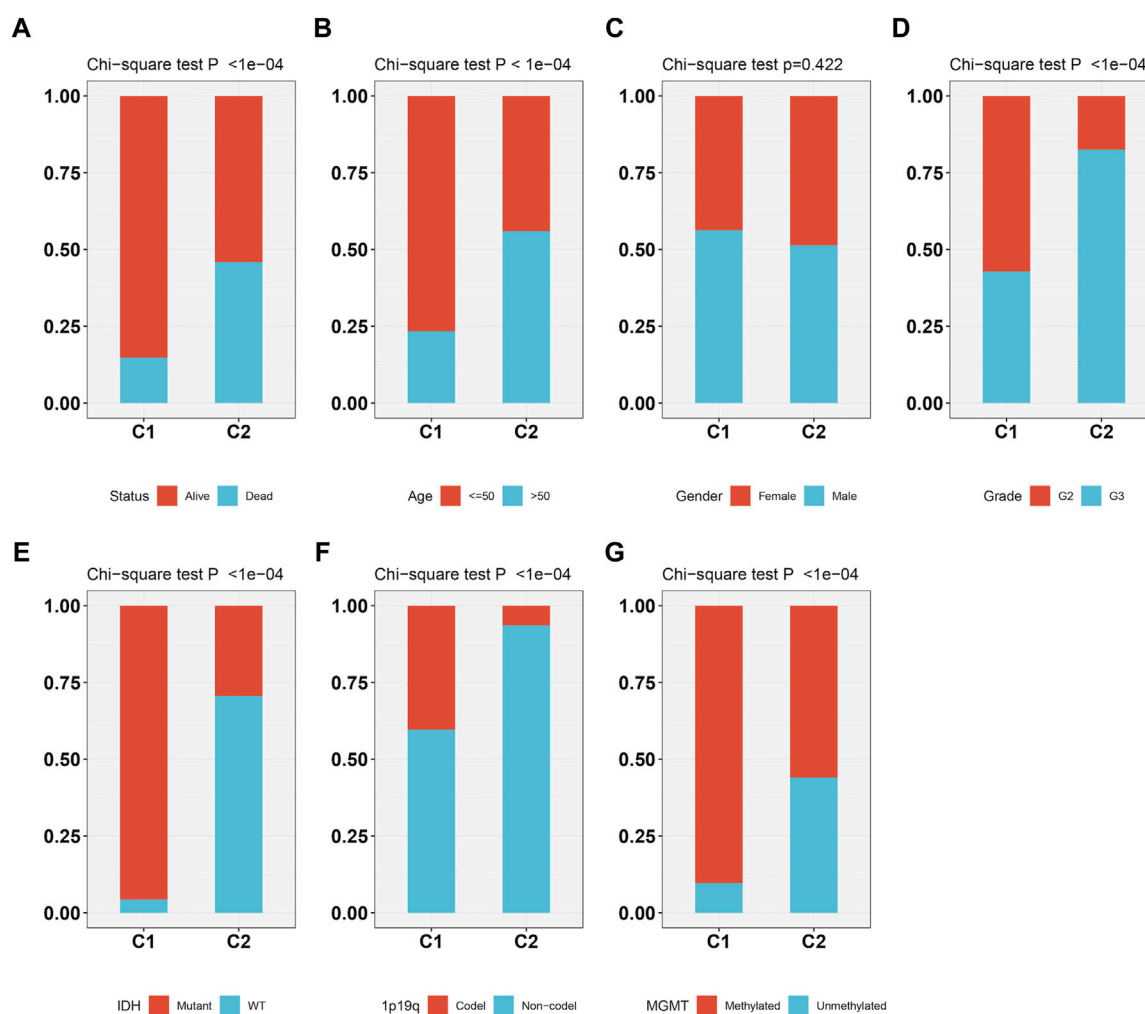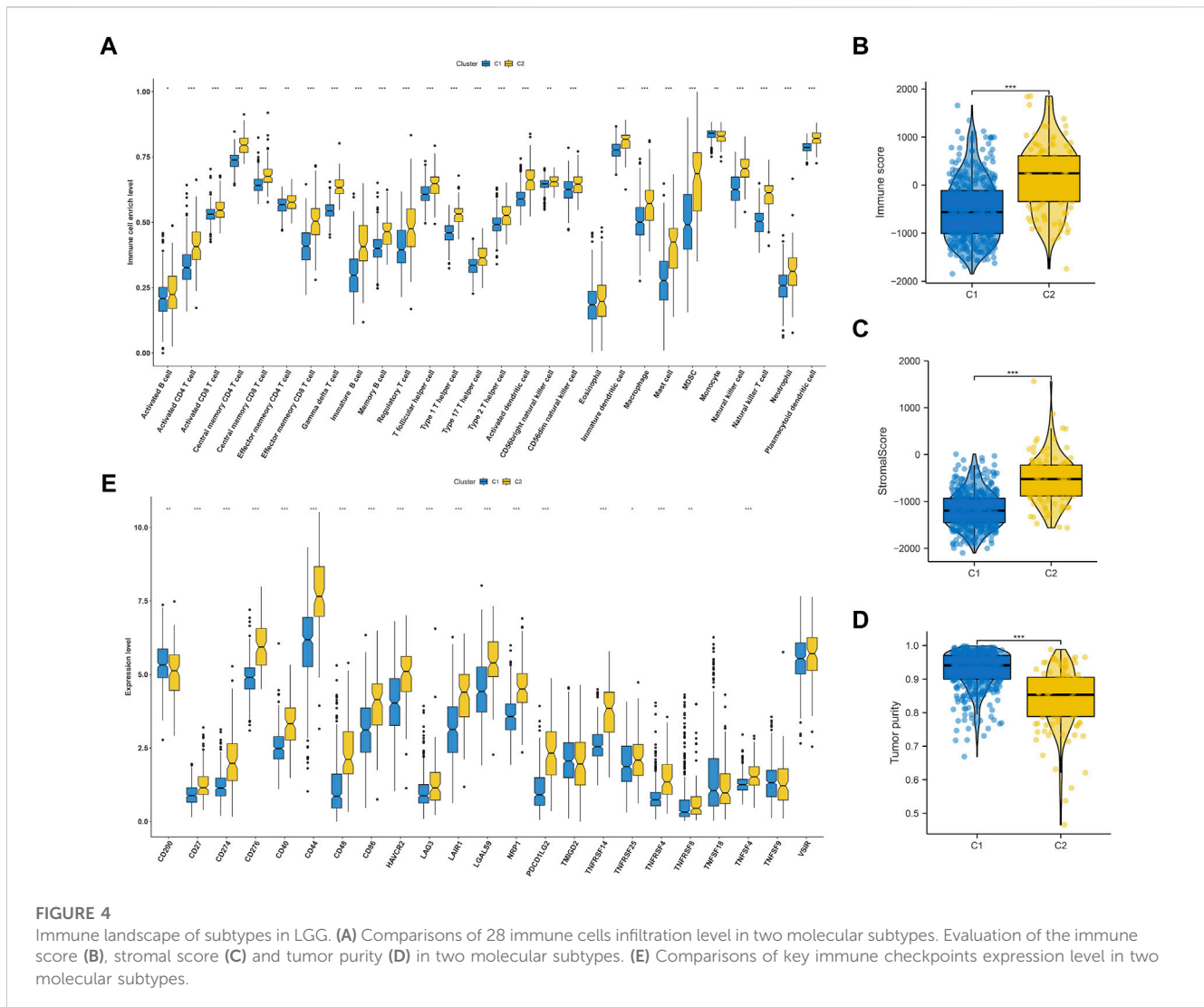
FIGURE 3
The relationship between two molecular subtypes, survival status **(A)**, age **(B)**, gender **(C)**, grade **(D)**, IDH mutation status **(E)**, 1p19q status **(F)** and methylation status **(G)**.

that C1 mainly enriched metabolism-related signaling pathways, such as butanoate metabolism and propanoate metabolism pathways. The primary routes of C2 that were enriched consist of cell adhesion molecules (CAMs), interaction with ECM receptors, the p53 signaling pathway, interaction between cytokines and cytokine receptors, focal adhesion, the JAK-stat signaling pathway, and apoptosis. These pathways were significantly correlated with tumor cell genesis, proliferation, invasion and migration. Furthermore, the GSVA outcomes indicated that C2 was linked to pathways related to DNA damage repair (such as Nucleotide excision repair, DNA replication, and Mismatch_ repair) as well as immune functions (including Natural killer cell mediated cytotoxicity, Toll like receptor signaling pathway, and Antigen processing and presentation). In conclusion, C2 showed more malignant biological behaviors and enrichment signaling pathways than C1.In addition, SNP analysis of the two clusters in TCGA uncovered that C1 had a higher mutation proportion than C2 of IDH1, TP53, ATRX, CIC, FUBP1, NOTCH1 (Rotin et al., 2015) and IDH2 (Supplementary Figure S1A). However, EGFR, PTEN and NF1 (Verhaak et al., 2010; Marques et al., 2019; Koptyra

et al., 2023), the mutations that were common in glioblastoma (GBM), were more frequent in C2 than C1 (Supplementary Figure S2B). Therefore, EMT related genes are important references for LGG molecular subtype. Our cluster-specific CNAs analysis presented chromosome deletions and amplifications, for example, deletion of 9p21.3 were significantly enriched in the C2. The 9p21.3 deletions is associated with the poor survival outcome in LGG (Xia et al., 2021). In addition, gene duplication/amplification at 7p11.2 activates EGFR expression through the formation of new topological associating domain (TAD) and the emergence of new enhancer-promoter interactions between LINC01446 and EGFR, and EGFR amplification and/or mutations are directly associated with poor prognosis of gliomas (Yang et al., 2022).

3.2 Disparity in immune tumor microenvironment between subtypes

We delved deeper into the dissimilarities in the immune tumor microenvironment between the two LGG EMT

FIGURE 4
Immune landscape of subtypes in LGG. **(A)** Comparisons of 28 immune cells infiltration level in two molecular subtypes. Evaluation of the immune score **(B)**, stromal score **(C)** and tumor purity **(D)** in two molecular subtypes. **(E)** Comparisons of key immune checkpoints expression level in two molecular subtypes.
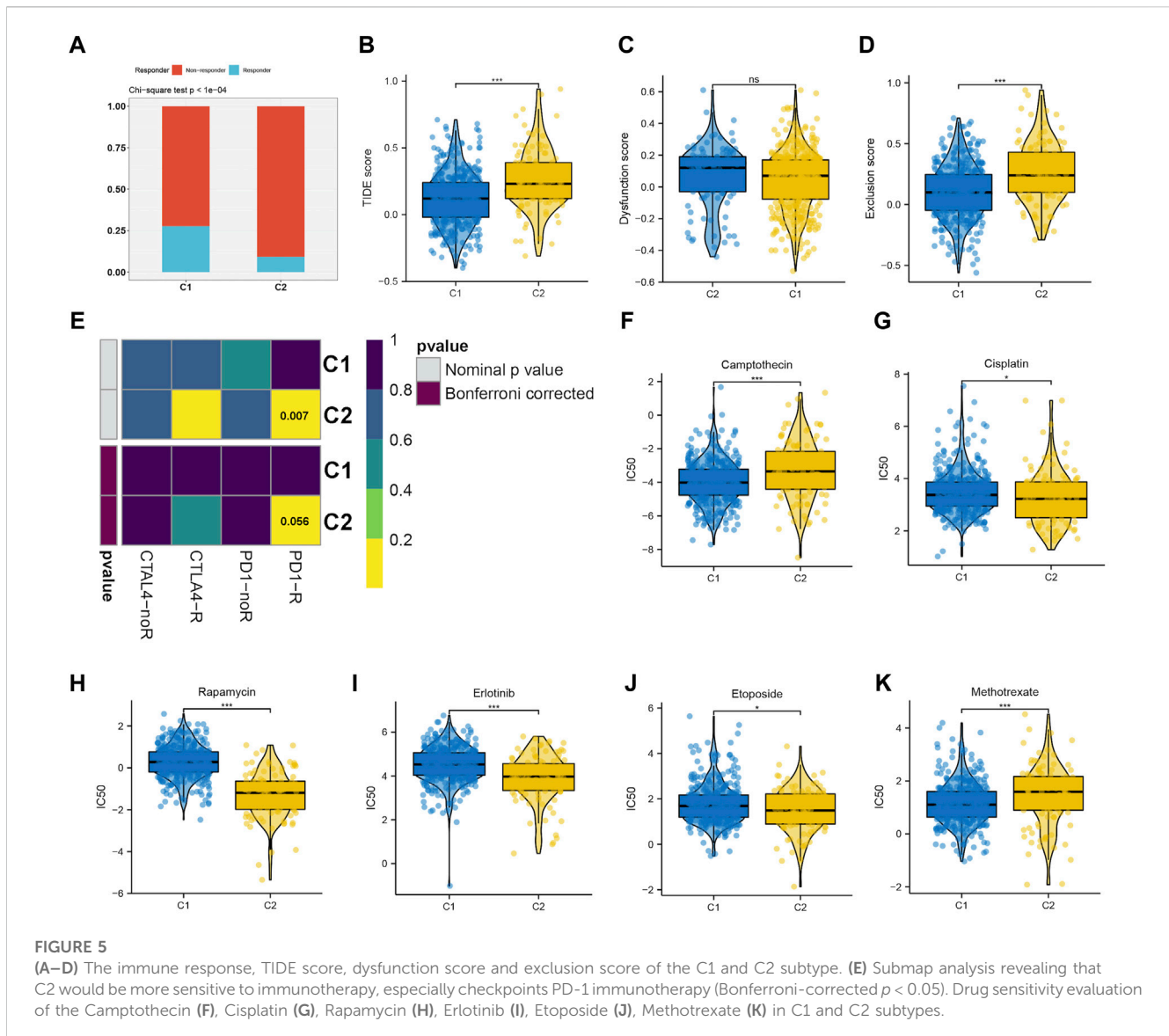
subtypes, given their notable clinicopathologic distinctions. According to Figure 4A, C2 had a higher degree of immune cell infiltration than C1, which included various types of immune cells such as Activated CD8+ T cells, effector memory CD4+ T cells, Regulatory T cells, CD56bright natural Killer cells, Type-1 T-helper 1 cells, Central memory CD4+ T cells, Activated Dendritic cells, Activated CD4+ T cells, CD56dim natural killer cells, myeloid-derived suppressor cells, Immature Dendritic cells, Central memory CD8+ T cells, Effector memory CD8+ T cells, immature B cells, macrophages, mast cells, memory B cells, natural killer cells, natural killer T cells, neutrophils, plasmacytoid dendritic cells, follicular helper T cells, Gamma delta T cells, Type-17 T helper cells, Type-2 T helper cells, activated B cells, eosinophils and monocytes. The findings from analyzing various cluster immune TME in TCGA and CGGA datasets indicated a strong correlation between the EMT molecular subtype and immune TME. It was noticed that C2 had higher stromal and immune scores compared to C1 (both $p < 0.0001$) as shown in Figures 4B,C. On the other hand, the tumor purity in C2 was lower than that in C1 ($p < 0.0001$) as depicted in Figure 4D. Figure 4E showed that C2 had

significantly higher expression levels of most checkpoints compared to C1.

3.3 Sensitivity prediction of different clusters to chemotherapy and immunotherapy

Immunotherapy has shown a potential application prospect in the treatment of glioma. We evaluated the response of two clusters to immunotherapy through TIDE analysis. The findings revealed that patients belonging to C1 exhibit a higher number of responders in comparison to C2 ($p < 0.05$) (Figure 5A), and individuals with a lower TIDE score are more likely to experience benefits from immunotherapy. The TIDE, Dysfunction, and Exclusion scores in the C1 group were lower, which implies that they could benefit more from immunotherapy (Figures 5B–D). Immune checkpoint blockade (ICB) is one of the most promising therapeutic approaches to change the current pessimistic situation in the treatment of glioma. Therefore, SubMap module analysis was used to compare the expression profile of LGG cluster with other published ICB-therapy datasets. As showed in Figure 5E, Subtype

FIGURE 5
(A–D) The immune response, TIDE score, dysfunction score and exclusion score of the C1 and C2 subtype. (E) Submap analysis revealing that C2 would be more sensitive to immunotherapy, especially checkpoints PD-1 immunotherapy (Bonferroni-corrected $p < 0.05$). Drug sensitivity evaluation of the Camptothecin (F), Cisplatin (G), Rapamycin (H), Erlotinib (I), Etoposide (J), Methotrexate (K) in C1 and C2 subtypes.

C2 is similar to the melanoma patients who reacted positively to anti-PD-1 therapy.

We then evaluated the different responses of two clusters to drug chemotherapy. The regimen of combined treatment for gliomas includes maximum safe surgical resection and postoperative chemotherapy with radiotherapy. Glioma can be treated with chemotherapy, which is considered a fundamental therapy. The GDSC database was utilized to predict the efficacy of 8 frequently used drugs in LGG chemotherapy, including cisplatin, erlotinib, methotrexate, camptothecin, etoposide, rapamycin, and doxorubicin. According to the findings, camptothecin (Figure 5F) and methotrexate (Figure 5K) were found to be more effective on C1, whereas cisplatin (Figure 5G), rapamycin (Figure 5H), erlotinib (Figure 5I), and etoposide (Figure 5J) were more effective on C2.

Genomic instability is a hallmark of cancer, particularly LGG, which facilitates its growth and the ability to withstand treatments (Dionellis et al., 2021). In our result, we found that Aneuploidy score, fraction altered, number of segments, tumor mutation burden (TMB) were elevated in C2, while Cancer Testicular Antigens (CTA)

score were lower in C2, which consistent with previously reports (Nie et al., 2020) (Figure 6).

3.4 Identification of EMT-related genes through WGCNA analysis

We utilized WGCNA to build correlation networks in order to identify genes associated with EMT. The optimal soft threshold was set to 5, based on a chosen R2 of 0.90 (Figures 7A,B). Eleven modules were identified based on their connectivity in Figure 7C, each containing at least 100 genes. The "module-trait" relationship heatmap revealed that brown module have the highest correlation associated with molecular subtypes (R = 0.70) (Figure 7D). In addition, six hub genes were identified, namely, SLC39A1, CTSA, TMSB4X, CAST, S100A11, and Chloride Intracellular Channel 1 (CLIC1), based on their GS being greater than 0.6 and their MM being greater than 0.8, as shown in Figure 7E. Further function enrichment of the brown module indicated that
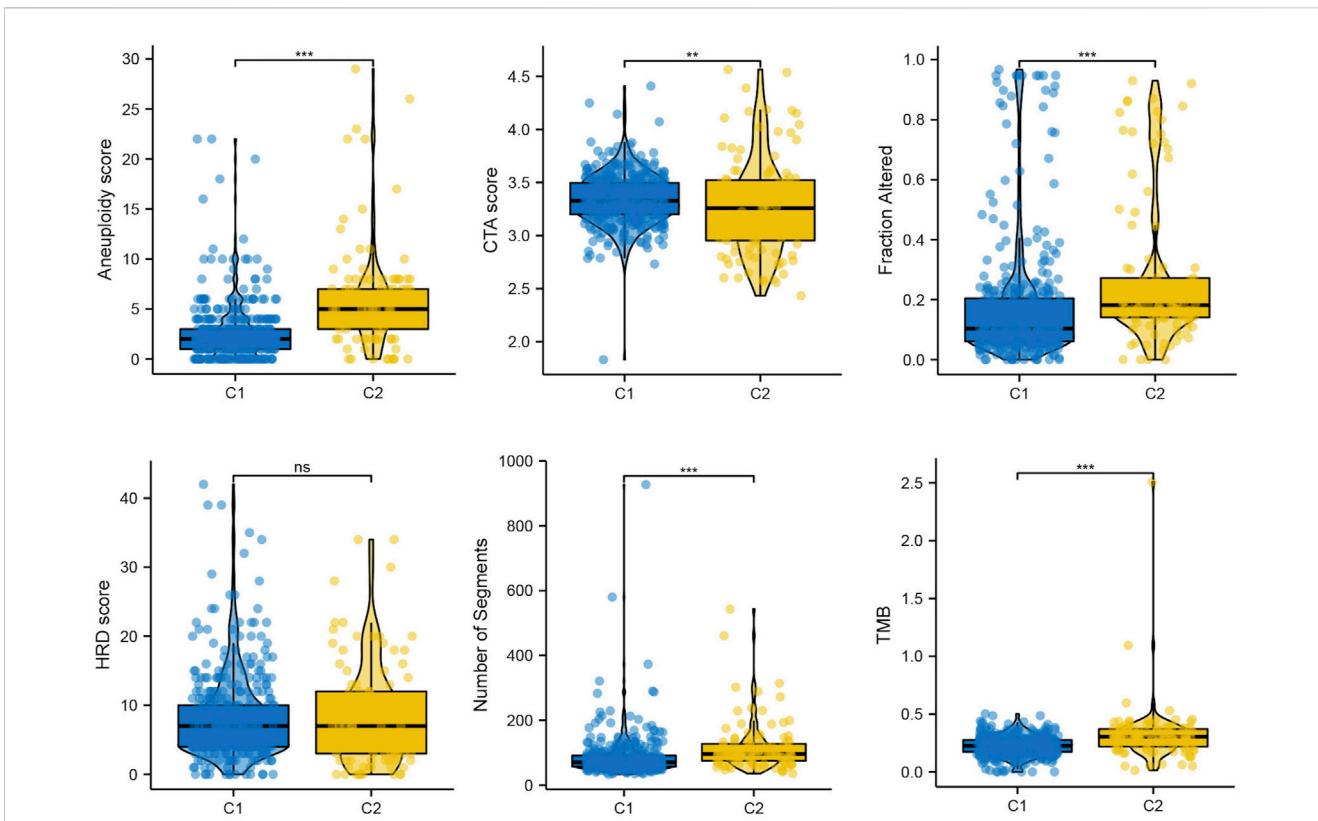
FIGURE 6
The mutation profile in subtypes, including Aneuploidy score, CTA score, Fraction altered, HRD score, Number of segments, and TMB.
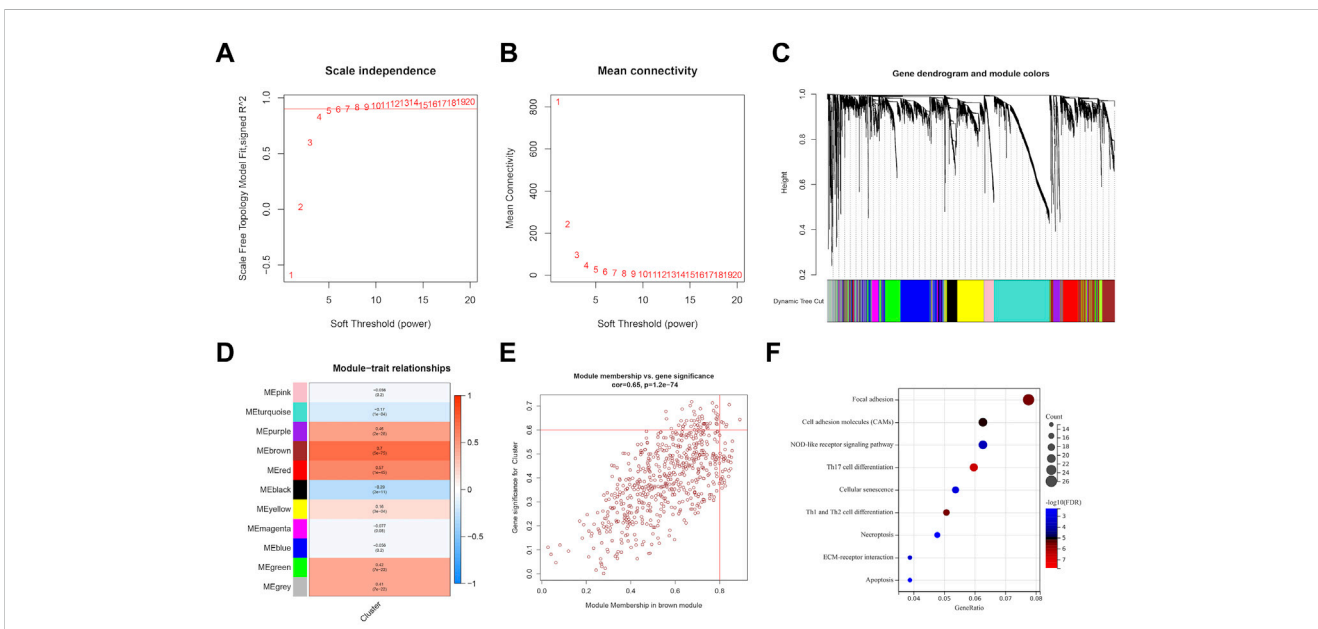
FIGURE 7
Identification of EMT-related genes through WGCNA analysis. **(A, B)** soft threshold determination through scale-free network analysis. **(C)** Clustering dendrograms were used to classify genes with similar expression patterns in LGG into co-expression modules. **(D)** module-traits heatmap showed the correlation between Cluster and gene modules. **(E)** hub genes were characterized by brown module with module membership >0.8 and gene significance >0.6. **(F)** Pathway enrichment analysis.

FIGURE 8
Development of EMT-related signature. **(A, B)** LASSO analysis to determine the EMT-related signature. **(C)** K-M curve survival analysis between high- and low-risk group in TCGA-LGG dataset. **(D)** ROC curve analysis of the EMT-related signature in TCGA-LGG dataset. **(E)** K-M curve survival analysis between high- and low-risk group in CGGA dataset. **(F)** ROC curve analysis of the of EMT-related signature in CGGA dataset. Determination of the independence of Univariate cox regression **(G)** and multivariate cox regression **(H)** analysis of the EMT-related signature. **(I)** Construction of nomogram on the basis of independent prognostic factors. **(J)** Calibration plot for the nomogram in 1-year, 3-year and 5-year OS.
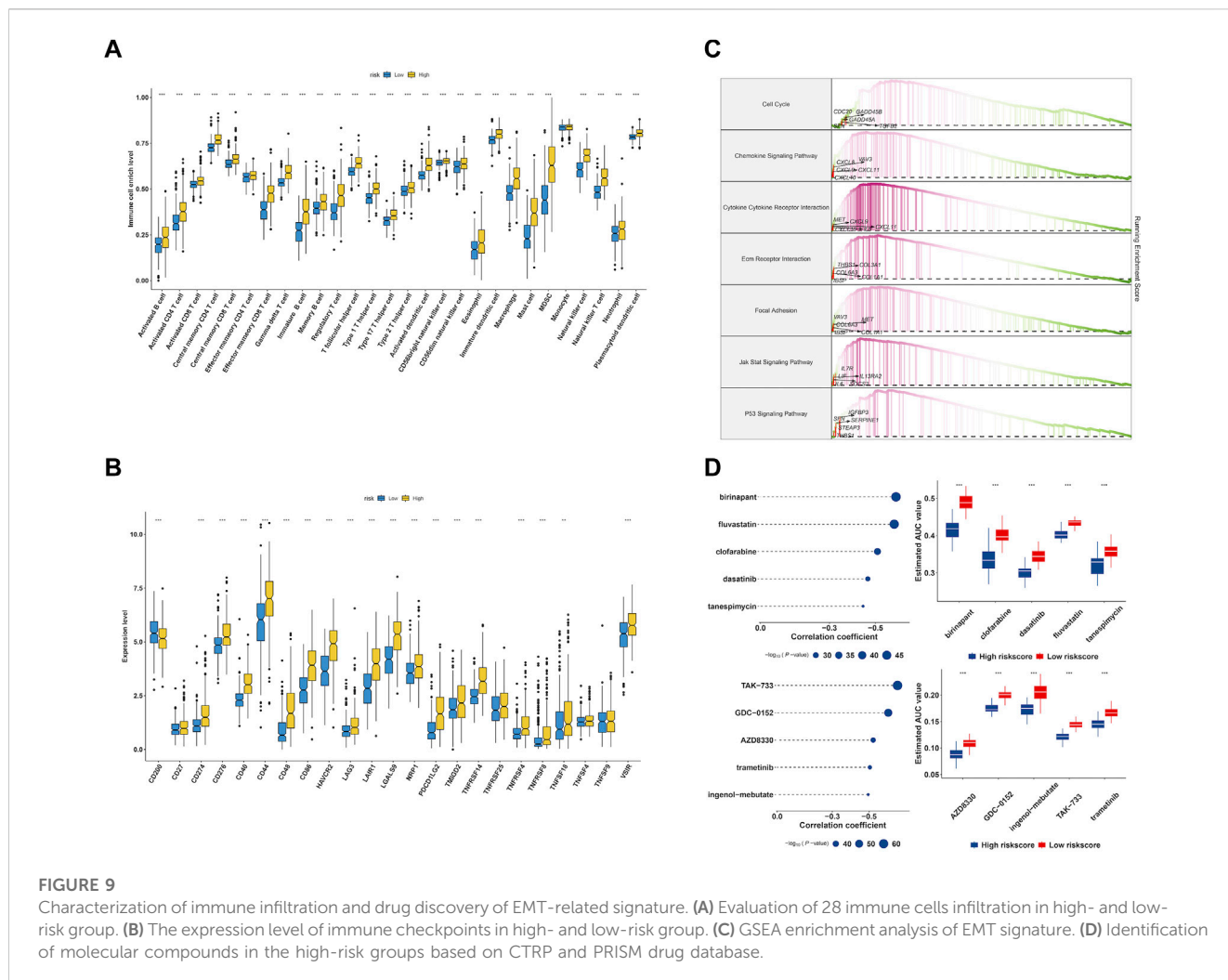
genes may involve in focal adhesion, Th17 cell differentiation, necroptosis, etc (Figure 7F).

To identify prognostic genes related to EMT from six hub genes, we performed lasso analysis and found three genes, namely, CLIC1, CTSA, and SLC39A1, based on the optimal lambda value (Figures 8A,B). The CGGA dataset was used as an external validation cohort to guarantee its reliability. The score for the risk model was computed utilizing the risk equation Risk score = (SLC39A1exp * −1.559) + (CTSAexp* −1.139) + (CLIC1exp * 1.474). After analyzing the TCGA and CGGA datasets, patients were divided into two groups: high-risk and low-risk, with the median score serving as the differentiating factor for each group. Figure 8C and Figure 8E demonstrated that patients who had a high-risk score in both TCGA and CGGA datasets had a poor survival outcome, whereas those with a low risk score had a favorable survival. Furthermore, the precision of the risk model was assessed using ROC curve analysis, yielding a score of 0.86, 0.76, and 0.69 for 1-, 3-, and 5-year periods in the TCGA and CGGA datasets, as shown in Figure 8D and 0.71, 0.74, and 0.76 in Figure 8F, respectively. The independent prognostic factors were

identified through the results of both univariate and multivariate Cox regression analyses, demonstrating that the risk score can function in this manner, as depicted in Figures 8G,H. Then, the independent factors including risk score, grade were including to construct nomogram to predict patients overall survival (OS) (Figure 8I). According to Figure 8J, the calibration curves indicated that the estimated OS values were consistent with the real values, particularly for the 3-year OS.

3.5 Immune infiltration landscape and drug discovery of EMT-related model

Though the immune cell infiltration analysis, we evaluated the divergence of immune cells between risk groups. The result of Figure 9A revealed that most of immune cells infiltration level in high risk group were elevated compared to low risk group, such as Activated CD8[+] T cells (CD8[+] Ta), effector memory CD4[+] T cells, Regulatory T cells, CD56bright natural Killer cells, Type-1 T-helper 1 cells, Central memory CD4[+] T cells, Activated

FIGURE 9
Characterization of immune infiltration and drug discovery of EMT-related signature. **(A)** Evaluation of 28 immune cells infiltration in high- and low-risk group. **(B)** The expression level of immune checkpoints in high- and low-risk group. **(C)** GSEA enrichment analysis of EMT signature. **(D)** Identification of molecular compounds in the high-risk groups based on CTRP and PRISM drug database.
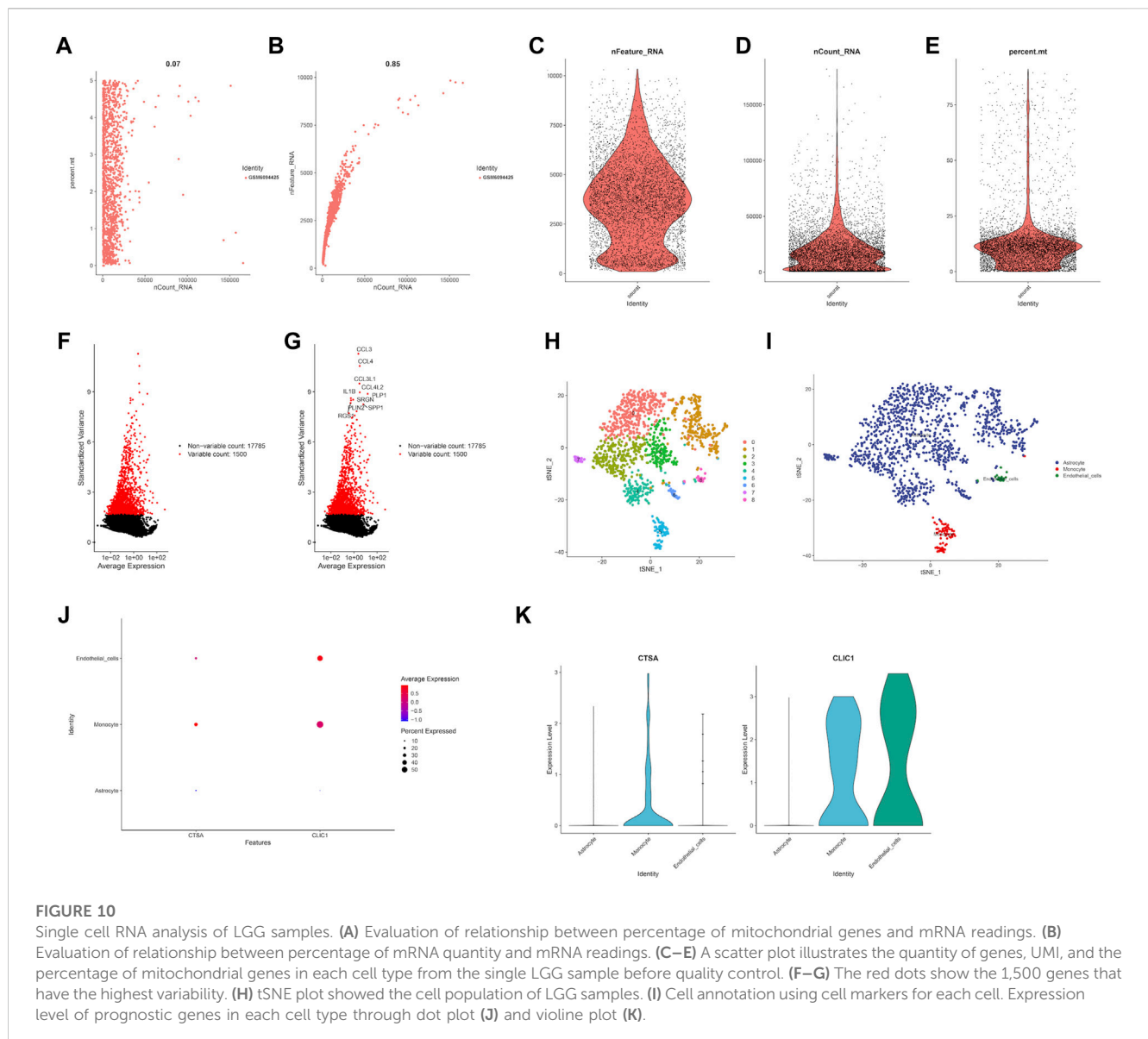
Dendritic cells, Activated CD4+ T cells, CD56dim natural killer cells, myeloid-derived suppressor cells, Immature Dendritic cells, Central memory CD8+ T cells, Effector memory CD8+ T cells, immature B cells, macrophages, mast cells, memory B cells, natural killer cells, natural killer T cells, neutrophils, plasmacytoid dendritic cells, follicular helper T cells, Gamma delta T cells, Type-17 T helper cells, Type-2 T helper cells, activated B cells, eosinophils and monocytes. We further checked the expression of immune checkpoint between groups and found that the genes showed a high expression level in high risk group (Figure 9B). Accumulative evidence have demonstrated that anti-tumor effects of a high T cell infiltration are counteracted by the immunosuppressive pathways that are triggered by the over-expression of immune checkpoint proteins (Vuong et al., 2019; Dionellis et al., 2021). The GSEA analysis was performed between groups, and found cell cycle, chemokine signaling pathway, cytokine-cytokine receptor interaction, ECM receptor interaction, p53 signaling pathway were significantly enriched in high risk group (Figure 9C).

To explore potential molecular drugs to the treatment of LGG in high risk group, we integrated CTRP and PRISM drug database. As showed in Figure 9D, the top panel of CTRP result revealed five drugs including birinapant, fluvastatin, clofarabine, dasatinib and tanespimycin may have the potential value to the treatment of high risk LGG patients, and the bottom panel of PRISM results indicated five drugs including TAK−733, GDC−0152, AZD8330, trametinib and ingenol−mebutate were identified.

3.6 Evaluation of gene expression level through single cell analysis

In order to estimate the gene expression level of hub genes, we retrieved LGG single cell sample from GSE202096 dataset (Figures 10A–E). Following data filtering and standardization, 1,500 genes exhibiting the most variance were selected for cell classification (Figures 10F,G). Dimensionality reduction of the expression levels of 1,500 genes was done by PCA analysis, resulting in PC1 to PC20. T-SNE was then applied to the PC1-20 to classify all cells into 9 clusters (Figure 10H). Afterwards, we annotated the cells in each cluster. The main cell types identified were astrocyte, monocyte, and endothelial cells (Figure 10I). The expression level of hub genes (CTSA, CLIC1) in cell clusters were showed in Figures 10J,K, of which CTSA was highly expressed in monocyte, and CLIC1 highly expressed in monocyte and endothelial cells.

FIGURE 10
Single cell RNA analysis of LGG samples. **(A)** Evaluation of relationship between percentage of mitochondrial genes and mRNA readings. **(B)** Evaluation of relationship between percentage of mRNA quantity and mRNA readings. **(C–E)** A scatter plot illustrates the quantity of genes, UMI, and the percentage of mitochondrial genes in each cell type from the single LGG sample before quality control. **(F–G)** The red dots show the 1,500 genes that have the highest variability. **(H)** tSNE plot showed the cell population of LGG samples. **(I)** Cell annotation using cell markers for each cell. Expression level of prognostic genes in each cell type through dot plot **(J)** and violine plot **(K)**.

3.7 Experiment validation of the hub genes

As depicted in the previously result, we obtained six hub genes (CAST, CLIC1, CTSA, S100A1, SLC39A1 and TMSB4X) associated with EMT from WGCNA result. Next, we investigated the level of expression of the six genes in four LGG cell lines, namely, HMC3, GOS-3, T98G, and LN-18. Figure 11A demonstrates that the majority of genes were expressed at high levels in both COS-3 and T98G cell lines, as indicated by the RT-PCR results. Using the gene expression profiling interactive analysis (GEPIA) database, we delved deeper into the expression levels of normal and tumor tissue. Our findings revealed that CLIC1, S100A1, and SLC39A1 were significantly highly expressed in tumor tissue compared to normal tissue. However, CAST, CTSA, and TMSB4X showed no significant difference (Supplementary Figure S3). Prior research indicated that CLIC1 and S100A1 are involved in the advancement of LGG, thus our focus shifted to examining the function of SLC39A1 in LGG. Firstly, we examined the expression level through the RT-PCR

experiment. Two LGG cell lines, GOS-3 and T98G, were treated with si-SLC39A1 #one to three, resulting in a notable reduction in SLC39A1 expression as determined by RT-qPCR analysis (Figure 11B). Meanwhile, we screened the stably transfected cell lines with puromycin after transfection of si-SLC39A1 #one to three, and the construction of stably transfected cell lines was observed by light microscopy 1 week after screening. The results showed that the stable-transformed cell line of si-SLC39A1 was successfully constructed by puromycin screening (Figure 11C). Flow cytometry was conducted to assess the apoptosis of GOS-3 and T98G cells, by comparing the number of cells distributed in early apoptosis and late apoptosis, the results showed that the apoptotic cell number in Q2 and Q3 regions of the si-SLC39A1 group was much higher than that of the si-NC group in both GOS-3 and T98G cell lines, revealing that the apoptosis rate was greater in cells with SLC39A1 knockout, indicating that the reduction of SLC39A1 significantly induced apoptosis of LGG cells (Figures 11D,E).
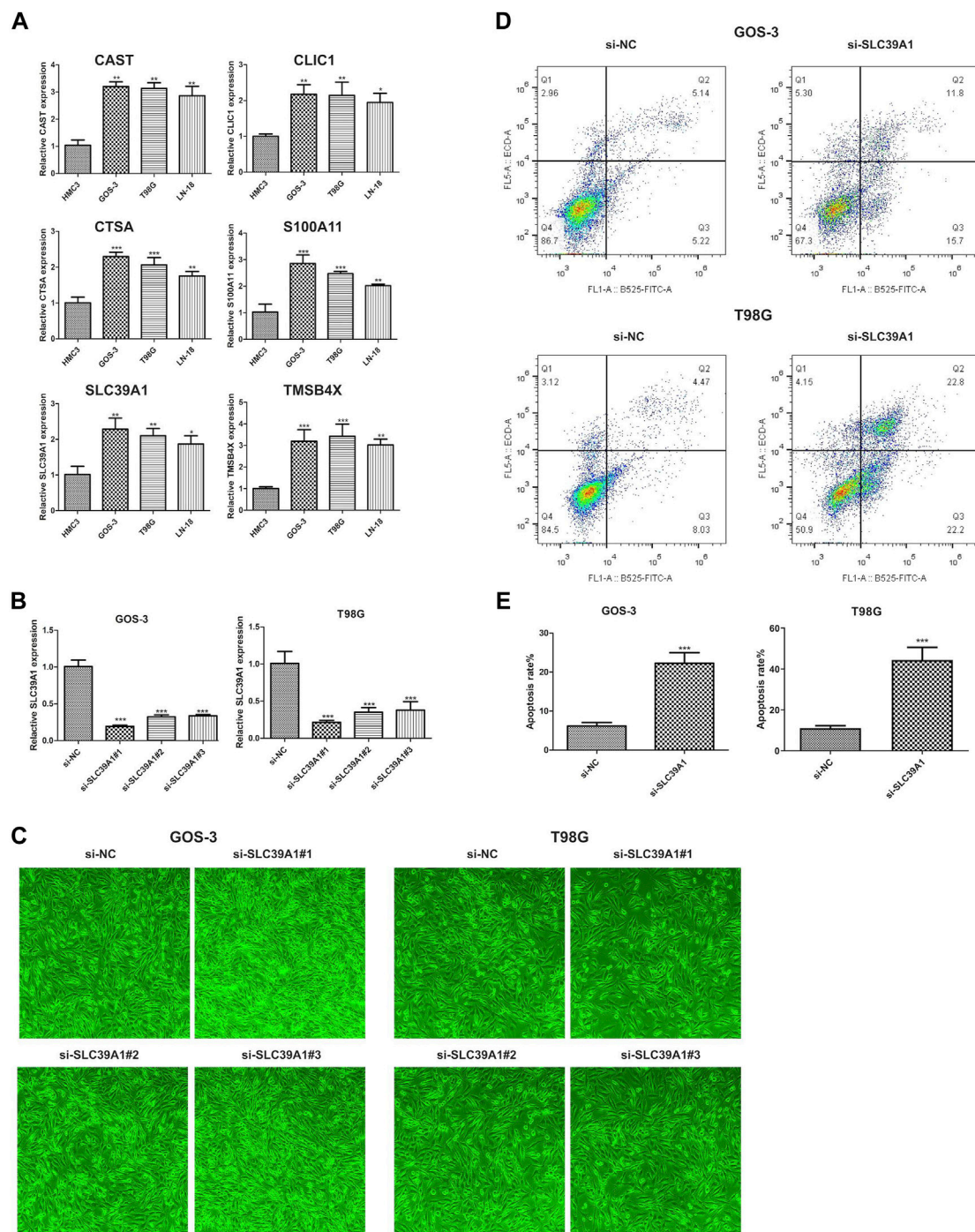
FIGURE 11
Validation of hub gene through *in-vitro* experiment. **(A)** Six hub genes expression level in HMC3, GOS-3, T98G, LN-18 cell lines. **(B)** SLC39A1 expression level in GOS-3 and T98G cell lines with SLC39A1 silence. **(C)** Transfection of si-SLC39A1 was filtered with puromycin for 1 week, and the expression of si-NC and si-SLC39A1 was observed by light microscopy. **(D, E)** Apoptotic level of SLC39A1-knockout GOS-3 and T98G cells.

4 Discussion

In the present study, a comprehensive analysis of LGG in TCGA and CGGA datasets was conducted based on EMT-related genes. By data mining analysis, we identified two clusters with markedly different prognosis and clinical phenotype in LGG. Further analysis showed that the two clusters were significantly different in TME immune cell infiltration and functional pathways, and presented inconsistent sensitivity to chemotherapy and immunotherapy. Finally, we developed a EMT-related signature and screen small molecule compounds with potential therapeutic effect targeting EMT. Additionally, we have undertaken further experimental validation of the identified EMT-related genes and their role in LGG. To some extent, our research contributed to

improving the current LGG treatment dilemma in precision therapy such as molecular classification, drug development, chemotherapy and immunotherapy.

The ESTIMATE and ssGSEA algorithms were used to measure the number of immune cells and stromal cells in every LGG sample from the TCGA and CGGA datasets. According to our data, immune and stromal scores as well as tumor purity vary among various LGG-EMT subtypes. Furthermore, our findings indicate that C2, which represents the LGG-EMT subtype with unfavorable prognosis, is associated with a greater level of infiltration of various immune cells such as CD56bright and CD56dim natural killer cells, eosinophils, monocytes, plasmacytoid dendritic cells, and others. Monocytes, which are believed to have tumor-promoting and immunosuppressive effects, are the primary infiltrating immune cells (Filipazzi et al., 2012; Marvel and Gabrilovich, 2015). It is anticipated that directing attention towards monocytes or other stromal components will alter the gliomas "cold" TME to a more "hot" TME phenotype. The efficiency of conventional first-line immunotherapy for glioma may be enhanced by this alteration (Tomaszewski et al., 2019). It may be possible to develop more effective therapeutic strategies if we understand the interaction between LGG and immune cells. Our findings using the TIDE algorithm supported the conclusion that C2 exhibited greater responsiveness to immunotherapy compared to C1. Furthermore, the SubMap method was employed to juxtapose the expression patterns of the LGG group with other previously published immune checkpoint datasets, which additionally validates the heightened susceptibility of C2 to immunotherapy, particularly PD-1.

The EMT-related survival prognosis model constructed by WGCNA and LASSO algorithm contains a series of hub genes, such as CLIC1, CTSA and SLC39A1. Multiple studies underscored the tumor-driver roles of the hub genes identified in present study. These hub genes were potential to be novel therapeutic targets and prognostic predictors in LGG. CLIC1, extensively studied within the CLIC family concerning tumors, has potential as a diagnostic indicator and therapeutic target. It has been linked to various cancers, influencing cell processes, including cell viability and mitochondrial function (Dehghan-Nayeri et al., 2017; Singha et al., 2018). In breast cancer, elevated CLIC1 expression correlates with tumor characteristics like size, TNM classification, grade, lymph node metastasis, and Ki67, while lower expression associates with extended OS and progression-free survival, suggesting its diagnostic potential (Xia et al., 2022). In esophageal squamous cell carcinoma (ESCC), CLIC1 is linked to clinical TNM classifications (Geng et al., 2022). CLIC1 knockdown in ESCC cell lines inhibits mTOR signaling, affecting cell proliferation and protein expression (Geng et al., 2022). High CLIC1 expression in lung adenocarcinoma predicts shorter overall survival and functions as an independent prognostic factor (Yasuda et al., 2022). In gastric cancer, CLIC1 absence impedes invasion and migration by affecting AMOT-p130 expression, possibly contributing to metastasis (Qiu et al., 2020). In hepatocellular carcinoma (HCC), upregulated CLIC1 is associated with aggressiveness, metastasis, and poor prognosis (Peng et al., 2020). GBM exhibits high CLIC1 expression (Setti et al., 2013). Reducing CLIC1 expression impairs cell proliferation and self-renewal in GBM, while CLIC1-mediated channel activity correlates with tumor aggressiveness (Setti et al., 2013). CLIC1 modulates reactive oxygen species and pH in human GBM stem cells, impacting motility and proliferation, making it a potential therapeutic target (Peretti et al.,

2018). The study by Biasiotta et al. identified CLIC1 among nine genes with significant alterations in ion channels in solid tumors and vascular malformations, particularly in GBM and bladder cancers (Biasiotta et al., 2016). CLIC1 expression is correlated with the drug-resistant protein MRP1. Knockdown of CLIC1 in human choriocarcinoma cell lines reduces MRP1 expression (Wu and Wang, 2016). Additionally, CLIC1 has been found to transfer from GBM cells to microvascular epithelial cells through extracellular vesicles, potentially impacting metastasis (Thuringer et al., 2018). In summary, CLIC1 plays a crucial role in various cancers, including breast cancer, ESCC, lung adenocarcinoma, gastric cancer, and HCC, making it a potential target for cancer treatment due to its influence on cell proliferation, migration, invasion, and metastasis. Research into CLIC1's role in cancer and glioma progression and patient survival is promising, but further studies are needed to fully understand the mechanisms of action and develop targeted therapies.

Research on HCC has shown that SLC39A1 overexpression is linked to immune infiltration and promotes tumor progression (Zhang et al., 2021). Yuan et al. (2022) conducted a study to explore SLC39A1's potential as a tumor suppressor in renal cell carcinoma (RCC) using integrated omics analyses. They found that SLC39A1 significantly impacts various metabolic pathways and triggers communication among multiple signaling pathways. This research provides valuable insights into RCC development and the molecular changes induced by SLC39A1 (Yuan et al., 2022). Furthermore, glioma tissues exhibit increased SLC39A1 expression, strongly associated with clinical features like grade, IDH mutation status, and 1p19q co-deletion status. Elevated SLC39A1 levels are linked to reduced survival chances and contribute to glioma malignancy by promoting cell growth, inhibiting cell death, and influencing immune cell infiltration in the tumor microenvironment (Wang et al., 2020). SLC39A1 shows promise as a novel prognostic biomarker and therapeutic target for gliomas.

CTSA, a lysosomal protease, is upregulated in various cancers, including HCC and prostate cancer (Park et al., 2020; Wang et al., 2021; Luo et al., 2022). Its role as a potential diagnostic and prognostic biomarker has been explored. Petrera et al. (2016) demonstrated that inhibiting CTSA in a mouse model improved cardiac functionality in heart failure (Petrera et al., 2016). High CTSA levels can differentiate HCC from healthy liver tissue and are linked to poorer survival rates (Wang et al., 2021). Luo et al. (2022) found high CTSA expression correlated with poor HCC patient outcomes, reinforcing its prognostic value (Luo et al., 2022). In prostate cancer, Park et al. (2020) showed that suppressing CTSA gene expression reduced proliferation, migration, and tumorigenesis (Park et al., 2020). CTSA is a potential therapeutic target and prognostic biomarker in various cancers. Hu et al., Zhang et al., Toss et al., and Kim et al. explored CTSA's role in lung adenocarcinoma (Hu et al., 2020), glioma (Zhang et al., 2022), breast ductal carcinoma in situ (Toss et al., 2019), and canine inflammatory mammary adenocarcinoma (Kim et al., 2020), respectively. Hu et al. found that CTSA promotes lung adenocarcinoma progression and may be a promising target for treatment (Hu et al., 2020). Zhang et al. linked CTSA to poor prognosis in glioma (Zhang et al., 2022). Toss et al. associated CTSA with unfavorable outcomes in breast ductal carcinoma in situ (Toss et al., 2019). Kim et al. demonstrated the impact of leptin on inflammatory mammary adenocarcinoma in dogs by modulating

CTSA expression (Kim et al., 2020). In summary, CTSA is significant in different cancers as a potential treatment target and predictor. Further research is needed to elucidate its precise molecular mechanisms and develop tailored treatments.

5 Conclusion

In conclusion, the present study comprehensively demonstrated the difference in molecular subtype of LGG based on EMT-related genes, thus revealing the tumor heterogeneity of LGG. We identified two distinct subtypes of LGG-EMT with varying clinical prognoses, clinicopathological characteristics, mutation statuses, immune cell infiltration, tumor microenvironment, and signaling pathway activities. The heterogeneity among LGG-EMT subtypes leads to divergent responses to immunotherapy and chemotherapy, guiding the precision treatment of LGG. Moreover, we developed and validated a reliable EMT-signature for LGG prognosis, and speculated about several small molecule compounds that could enhance the clinical practical value of LGG. Further experimental studies can help us understand the underlying mechanisms of EMT in LGG, thereby providing additional support for the clinical utility of the prognostic signature.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

CW: Writing–original draft. ZH: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing–original draft.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1276466/full#supplementary-material

References

Akalay, I., Janji, B., Hasmim, M., Noman, M. Z., Andre, F., Cremoux, P., et al. (2013). Epithelial-to-Mesenchymal transition and autophagy induction in breast carcinoma promote escape from T-cell-mediated lysis. *Cancer Res.* 73, 2418–2427. doi:10.1158/0008-5472.CAN-12-2432

Basu, A., Bodycombe, N., Cheah, J., Price, E., Liu, K., Schaefer, G., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell.* 154, 1151–1161. doi:10.1016/j.cell.2013.08.003

Biasiotta, A., D'Arcangelo, D., Passarelli, F., Nicodemi, E., and Facchiano, A. (2016). Ion channels expression and function are strongly modified in solid tumors and vascular malformations. *J. Transl. Med.* 14, 285. doi:10.1186/s12967-016-1038-y

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Cao, Y., Fu, L., Wu, J., Peng, Q., Nie, Q., Zhang, J., et al. (2022). Integrated analysis of multimodal single-cell data with structural similarity. *Nucleic Acids Res.* 50, e121. doi:10.1093/nar/gkac781

Chen, X. Y., Chen, Y., Fang, W. H., Wu, Z. Y., Wang, D. L., Xu, Y. W., et al. (2022). Integrative and comparative single-cell analysis reveals transcriptomic difference between human tumefactive demyelinating lesion and glioma. *Commun. Biol.* 5, 941. doi:10.1038/s42003-022-03900-0

Claus, E., Walsh, K., Wiencke, J., Molinaro, A., Wiemels, J., Schildkraut, J., et al. (2015). Survival and low-grade glioma: the emergence of genetic information. *Neurosurg. focus* 38, E6. doi:10.3171/2014.10.FOCUS12367

Corsello, S., Nagari, R., Spangler, R., Rossen, J., Kocak, M., Bryan, J., et al. (2020). Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* 1, 235–248. doi:10.1038/s43018-019-0018-6

Dehghan-Nayeri, N., Eshghi, P., Pour, K., Rezaei tavirani, M., Omrani, m.d., and Gharehbaghian, A. (2017). Differential expression pattern of protein markers for predicting chemosensitivity of dexamethasone-based chemotherapy of B cell acute lymphoblastic leukemia. *Cancer Chemother. Pharmacol.* 80, 177–185. doi:10.1007/s00280-017-3347-0

Dionellis, V., Norkin, M., Karamichali, A., Rossetti, G., Huelsken, J., Ordóñez-Morán, P., et al. (2021). Genomic instability profiles at the single cell level in mouse colorectal cancers of defined genotypes. *Cancers* 13, 1267. doi:10.3390/cancers13061267

Feng, S., Xu, Y., Dai, Z., Yin, H., Zhang, K., and Shen, Y. (2022). Integrative analysis from multicenter studies identifies a WGCNA-derived cancer-associated fibroblast signature for ovarian cancer. *Front. Immunol.* 13, 951582. doi:10.3389/fimmu.2022.951582

Filipazzi, P., Huber, V., and Rivoltini, L. (2012). Phenotype, function and clinical implications of myeloid-derived suppressor cells in cancer patients. *Cancer Immunol. Immunother.* 61, 255–263. doi:10.1007/s00262-011-1161-9

Geeleher, P., Cox, N., and Huang, R. (2014a). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS one* 9, e107468. doi:10.1371/journal.pone.0107468

Geeleher, P., Cox, N., and Huang, R. (2014b). Abstract 5561: clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol. 15, R47. doi:10.1186/gb-2014-15-3-r47

Geng, H., Feng, C., Sun, Z., Fan, X., Xie, Y., Gu, J., et al. (2022). Chloride intracellular channel 1 promotes esophageal squamous cell carcinoma proliferation via mTOR signalling. Transl. Oncol. 27, 101560. doi:10.1016/j.tranon.2022.101560

Golub, D., Iyengar, N., Dogra, S., Wong, T., Bready, D., Tang, K., et al. (2019). Mutant isocitrate dehydrogenase inhibitors as targeted cancer therapeutics. Front. Oncol. 9, 417. doi:10.3389/fonc.2019.00417

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinforma. 14, 7. doi:10.1186/1471-2105-14-7

Hazra, A., and Gogtay, N. (2016). Biostatistics series module 3: comparing groups: numerical variables. Indian J. Dermatology 61, 251–260. doi:10.4103/0019-5154.182416

Hoshida, Y., Brunet, J. P., Tamayo, P., Golub, T., and Mesirov, J. (2007). Subclass mapping: identifying common subtypes in independent disease data sets. PloS one 2, e1195. doi:10.1371/journal.pone.0001195

Hu, B., Zhu, X., and Lu, J. (2020). Cathepsin A knockdown decreases the proliferation and invasion of A549 lung adenocarcinoma cells. Mol. Med. Rep. 21, 2553–2559. doi:10.3892/mmr.2020.11068

Hu, C., Li, T., Xu, L., Zhang, X., Li, F., Bai, J., et al. (2022). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic acids Res. 51, D870–D876. doi:10.1093/nar/gkac947

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Kim, J., Mahiddine, F., and Kim, G. (2020). Leptin modulates the metastasis of canine inflammatory mammary adenocarcinoma cells through downregulation of lysosomal protective protein cathepsin A (CTSA). Int. J. Mol. Sci. 21, 8963. doi:10.3390/ijms21238963

Koptyra, M., Rahti, K., Zhu, Y., Farrow, B., Miller, D., Kraya, A., et al. (2023). Abstract 3566: expansion of the pediatric brain tumor Atlas: children's brain tumor network, kids first data resource and childhood cancer data initiative open science effort. Cancer Res. 83, 3566. doi:10.1158/1538-7445.am2023-3566

Lambert, A., Pattabiraman, D., and Weinberg, R. (2017). Emerging biological principles of metastasis. Cell. 168, 670–691. doi:10.1016/j.cell.2016.11.037

Langfelder, P., Horvath, S., Langfelder, P., and Horvath, S. (2009). WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma. 9, 559. doi:10.1186/1471-2105-9-559

Leek, J., Johnson, W., Parker, H., Jaffe, A., and Storey, J. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinforma. Oxf. Engl. 28, 882–883. doi:10.1093/bioinformatics/bts034

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell. Syst. 1, 417–425. doi:10.1016/j.cels.2015.12.004

Loret, N., Denys, H., Tummers, P., and Berx, G. (2019). The role of epithelial-to-mesenchymal plasticity in ovarian cancer progression and therapy resistance. Cancers 11, 838. doi:10.3390/cancers11060838

Lu, K., Chang, J., Parachoniak, C., Pandika, M., Aghi, M., Meyronet, D., et al. (2012). VEGF inhibits tumor cell invasion and mesenchymal transition through a MET/VEGFR2 complex. Cancer Cell. 22, 21–35. doi:10.1016/j.ccr.2012.05.037

Luo, L., Wang, X., Wang, H., Yang, C., Zhang, Y., Li, X., et al. (2022). High cathepsin A protein expression predicts poor prognosis and tumor recurrence of hepatocellular carcinoma patients after curative hepatectomy. Am. J. cancer Res. 12, 3843–3856.

Marques, C., Unterkircher, T., Kroon, P., Izzo, A., Gargiulo, G., Kling, E., et al. (2019). NF1 regulates mesenchymal glioblastoma plasticity and aggressiveness through the AP-1 transcription factor FOSL1. Elife 10, e64846. doi:10.7554/eLife.64846

Marvel, D., and Gabrilovich, D. I. (2015). Myeloid-derived suppressor cells in the tumor microenvironment: expect the unexpected. J. Clin. Invest. 125, 3356–3364. doi:10.1172/JCI80005

Mayakonda, A., Lin, D., Assenov, Y., Plass, C., and Koeffler, H. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28, 1747–1756. doi:10.1101/gr.239244.118

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer cell 17 (1), 98–110. doi:10.1016/j.ccr2009.12.020

Miyoshi, A., Kitajima, Y., Sumi, K., Sato, K., Hagiwara, A., Koga, Y., et al. (2004). Snail and SIP1 increase cancer invasion by upregulating MMP family in hepatocellular carcinoma cells. Br. J. cancer 90, 1265–1273. doi:10.1038/sj.bjc.6601685

Nie, W., Xu, M. D., Gan, L., Yi, Z., Qian, J., Gu, K., et al. (2020). Advanced non-small cell lung cancer patients with low tumor mutation burden might derive benefit from immunotherapy. J. Immunother. Hagerst. Md, 1997) 43, 189–195. doi:10.1097/CJI.0000000000000318

Park, S., Kwon, W., Park, J. K., Baek, S. M., Lee, S. W., Cho, G. J., et al. (2020). Suppression of cathepsin a inhibits growth, migration, and invasion by inhibiting the p38 MAPK signaling pathway in prostate cancer. Archives Biochem. Biophysics 688, 108407. doi:10.1016/j.abb.2020.108407

Peng, J. M., Lin, S. H., Yu, M. C., and Hsieh, S. Y. (2020). CLIC1 recruits PIP5K1A/C to induce cell-matrix adhesions for tumor metastasis. J. Clin. Investigation 131, e133525. doi:10.1172/JCI133525

Peretti, M., Raciti, F., Carlini, V., Verduci, I., Sertic, S., Barozzi, S., et al. (2018). Mutual influence of ROS, pH, and CLIC1 membrane protein in the regulation of G1-S phase progression in human glioblastoma stem cells. Mol. Cancer Ther. 17, 2451–2461. doi:10.1158/1535-7163.MCT-17-1223

Petrera, A., Gassenhuber, J., Ruf, S., Gunasekaran, D., Eßer, J., Shahinian, J., et al. (2016). Cathepsin A inhibition attenuates myocardial infarction-induced heart failure on the functional and proteomic levels. J. Transl. Med. 14, 153. doi:10.1186/s12967-016-0907-8

Qiu, Y., Mao, Y. T., Zhu, J. H., Zhao, K., Wang, J. F., Huang, J. M., et al. (2020). CLIC1 knockout inhibits invasion and migration of gastric cancer by upregulating AMOT-p130 expression. Clin. Transl. Oncol. 23, 514–525. doi:10.1007/s12094-020-02445-0

Rhim, A., Aiello-Couzo, N., Mirek, E., and Stanger, B. (2014). Abstract IA5: epigenetic genome control by heterochromatin machinery and non-coding RNAs. Cancer Res. 72, IA5. doi:10.1158/1538-7445.nonrna12-ia5

Rotin, D., Brat, D. J., Verhaak, R. G. W., Aldape, K. D., Yung, W. K. A., Salama, S. R., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Eng. J. Med. 372, 2481–2498. doi:10.1056/NEJMoa1402121

Setti, M., Savalli, N., Osti, D., Richichi, C., Angelini, M., Brescia, P., et al. (2013). Functional role of CLIC1 ion channel in glioblastoma-derived stem/progenitor cells. J. Natl. Cancer Inst. 105, 1644–1655. doi:10.1093/jnci/djt278

Shibue, T., and Weinberg, R. (2017). EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat. Rev. Clin. Oncol. 14, 611–629. doi:10.1038/nrclinonc.2017.44

Singha, B., Harper, S., Goldman, A., Bitler, B., Aird, K., Borowsky, M., et al. (2018). CLIC1 and CLIC4 complement CA125 as a diagnostic biomarker panel for all subtypes of epithelial ovarian cancer. Sci. Rep. 8, 14725. doi:10.1038/s41598-018-32885-2

Thuringer, D., Chanteloup, G., Winckler, P., and Garrido, C. (2018). The vesicular transfer of CLIC1 from glioblastoma to microvascular endothelial cells requires TRPM7. Oncotarget 9, 33302–33311. doi:10.18632/oncotarget.26048

Tomaszewski, W., Sanchez-Perez, L., Gajewski, T., and Sampson, J. (2019). Brain tumor micro-environment and host state - implications for immunotherapy. Clin. Cancer Res. 25, 4202–4210. doi:10.1158/1078-0432.CCR-18-1627

Toss, M., Miligy, I., Haj-Ahmad, R., Gorringe, K., Al-Kawaz, A., Mittal, K., et al. (2019). The prognostic significance of lysosomal protective protein (Cathepsin A) in breast ductal carcinoma in situ. Histopathology 74, 1025–1035. doi:10.1111/his.13835

van Staalduinen, J., Baker, D., Dijke, P., and Dam, H. (2018). Epithelial–mesenchymal-transition-inducing transcription factors: new targets for tackling chemoresistance in cancer? Oncogene 37, 6195–6211. doi:10.1038/s41388-018-0378-x

Vuong, L., Kotecha, R., Voss, M., and Hakimi, A. (2019). Tumor microenvironment dynamics in clear-cell renal cell carcinoma. Cancer Discov. 9, 1349–1357. doi:10.1158/2159-8290.CD-19-0499

Wang, H., Xu, F., Yang, F., Lv, L., and Jiang, Y. (2021). Prognostic significance and oncogene function of cathepsin A in hepatocellular carcinoma. Sci. Rep. 11, 14611. doi:10.1038/s41598-021-93998-9

Wang, P., Zhang, J., He, S., Xiao, B., and Peng, X. (2020). SLC39A1 contribute to malignant progression and have clinical prognostic impact in gliomas. Cancer Cell. Int. 20, 573. doi:10.1186/s12935-020-01675-0

Weller, M., Weber, R., Willscher, E., Riehmer, V., Hentschel, B., Kreuz, M., et al. (2015). Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. Acta neuropathol. 129, 679–693. doi:10.1007/s00401-015-1409-0

Wilkerson, M., and Hayes, D. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinforma. Oxf. Engl. 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, J., and Wang, D. (2016). CLIC1 induces drug resistance in human choriocarcinoma through positive regulation of MRP1. Oncol. Res. 25, 863–871. doi:10.3727/096504016X14772315906527

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141. doi:10.1016/j.xinn.2021.100141

Xia, J., Wang, Q., Ju, F., Luo, X., Wang, F., Zhou, Y., et al. (2022). Chloride Intracellular Channel 1 is a potential biomarker for breast cancer. Breast Cancer Targets Ther. 14, 247–258. doi:10.2147/BCTT.S367519

Xia, Y., Liu, Y., Yang, C., Simeone, D., Sun, T. T., DeGraff, D., et al. (2021). Dominant role of CDKN2B/p15INK4B of 9p21.3 tumor suppressor hub in inhibition of cell-cycle and glycolysis. *Nat. Commun.* 12, 2047. doi:10.1038/s41467-021-22327-5

Yamaguchi, Y. (2016). *Immunotherapy of cancer: an innovative treatment comes of age*. Japan: Springer.

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., et al. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell. Biol.* 21, 341–352. doi:10.1038/s41580-020-0237-9

Yang, Q., Jiang, N., Zou, H., Fan, X., Liu, T., Huang, X., et al. (2022). Alterations in 3D chromatin organization contribute to tumorigenesis of EGFR-amplified glioblastoma. *Comput. Struct. Biotechnol. J.* 20, 1967–1978. doi:10.1016/j.csbj.2022.04.007

Yang, W., Soares, J., Greninger, P., Edelman, E., Lightfoot, H., Forbes, S., et al. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids Res.* 41, D955–D961. doi:10. 1093/nar/gks1111

Yang, Y., and Xu, X. (2021). Identification of key genes in coronary artery disease: an integrative approach based on weighted gene co-expression network analysis and their correlation with immune infiltration. *Aging (Albany NY)* 13, 8306–8319. doi:10.18632/ aging.202638

Yasuda, Y., Nagano, T., Jimbo, N., Kiriu, T., Suraya, R., Hazama, D., et al. (2022). Chloride Intracellular Channel 1 expression is associated with poor prognosis of lung adenocarcinoma. *Anticancer Res.* 42, 271–277. doi:10.21873/anticanres.15482

Yoshihara, K., Shahmoradgoli, M., Martinez-Ledesma, J. E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yuan, Y., Liu, Z., Li, B., Gong, Z., Piao, C., Du, Y., et al. (2022). Integrated analysis of transcriptomics, proteomics and metabolomics data reveals the role of SLC39A1 in renal cell carcinoma. *Front. Cell. Dev. Biol.* 10, 977960. doi:10.3389/fcell.2022.977960

Zhang, M., Huang, J., Wang, Y., Nie, Q., Zhang, X., Yang, Y., et al. (2022). Cathepsin A upregulation in glioma: a potential therapeutic target associated with immune infiltration. *J. Med. Biochem.* 41, 459–465. doi:10.5937/jomb0-35677

Zhang, Q., Pan, J., An, F., Nie, H., and Zhan, Q. (2021). Decreased SLC39A1 (solute carrier family 39 member 1) expression predicts unfavorable prognosis in patients with early-stage hepatocellular carcinoma. *Bioengineered* 12, 8147–8156. doi:10.1080/ 21655979.2021.1987131

Zhou, H., Huang, L., Liang, L., Chen, L., Zou, C., Li, Z., et al. (2021). Identification of an miRNA regulatory network and candidate markers for ischemic stroke related to diabetes. *Int. J. Gen. Med.* 14, 3213–3223. doi:10.2147/IJGM.S319503

Glossary

CAMs	cell adhesion molecules
CDF	cumulative distribution function
CGGA	the Chinese Glioma Genome Atlas
CLIC1	Chloride Intracellular Channel 1
CNA	copy number alterations
CTA	Cancer Testicular Antigens
CTRP	Cancer Therapeutics Response Portal
DCIS	ductal carcinoma *in situ*
DEGs	differentially expressed genes
DSS	disease-specific survival
ECM	extracellular matrix
EMT	epithelial-mesenchymal transition
ESCC	esophageal squamous cell carcinoma
ESTIMATE	Estimation of Stromal and Immune cells in MAlignant Tumours using Expression
FDR	false discovery rate
GBM	glioblastoma
GDAC	Genome Data Analysis Center
GDSC	Genomics of Drug Sensitivity in Cancer
GEPIA	gene expression profiling interactive analysis
GS	gene significance
GSVA	gene set variation analysis
HCC	hepatocellular carcinoma
IC50	half-maximal inhibitory concentration
ICB	immune checkpoint blockade
KM	Kaplan-Meier
LGG	low-grade glioma
MM	module membership
MSigDB	Molecular signatures database
OS	overall survival
PCA	principal components analysis
RCC	renal cell carcinoma
ROC	receiver operating characteristic
RT-qPCR	Real-time quantitative PCR
scRNA-seq	single cell RNA-sequencing
SNP	single-nucleotide polymorphisms
ssGSEA	single-sample gene set enrichment analysis
TAD	topological associating domain
TCGA	the Cancer Genome Atlas
TIDE	tumor immune dysfunction and exclusion

TIICs	tumor infiltrating immune cells
TME	tumor microenvironment
TMZ	temozolomide
TNM	tumor node metastasis
TPM	transcripts per kilobase million
t-SNE	T-distributed stochastic neighbor embedding
WGCNA	weighted correlation network analysis