

A novel prognostic risk-scoring system based on m⁵C methylation regulator-mediated patterns for glioma patients

Yutong Wen,^{1,7} Xiaotong Chen,^{2,3,7} Runtong Li,^{1,7} Haiting Xie,^{1,7} Shuai Zhi,^{2,3} Kaitao Wang,¹ Shang Yi,¹ Wen Liang,⁴ Haiyan Hu,⁵ Shitao Rao,^{2,3} and Xiaoya Gao^{1,6}

¹Department of Neurology, Zhujiang Hospital of Southern Medical University, 253 Gongye Avenue, Guangzhou, Guangdong 510282, P.R. China; ²Department of Bioinformatics, Fujian Key Laboratory of Medical Bioinformatics, School of Medical Technology and Engineering, Fujian Medical University, Fuzhou 350122, China; ³Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou 350122, China; ⁴Department of Radiology, Zhujiang Hospital of Southern Medical University, 253 Gongye Avenue, Guangzhou, Guangdong 510282, P.R. China; ⁵Department of Oncology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, P.R. China; ⁶Department of Pediatric Neurology, Zhujiang Hospital of Southern Medical University, 253 Gongye Avenue, Guangzhou, Guangdong 510282, P.R. China

N5-methylcytosine (m⁵C) methylation modification plays a crucial role in the epigenetic mechanisms underlying tumorigenesis, aggressiveness, and malignancy in diffuse glioma. Our study aimed to develop a novel prognostic risk-scoring system to assess the impact of m⁵C modification in glioma patients. Initially, we identified two distinct m⁵C clusters based on the expression level of m⁵C regulators in The Cancer Genome Atlas glioblastoma (TCGA-GBM) dataset. Differentially expressed genes (DEGs) between the two m⁵C cluster groups were determined. Utilizing these m⁵C regulation-related DEGs, we classified glioma patients into three gene cluster groups: A, B, and C. Subsequently, an m⁵C scoring system was developed through a univariate Cox regression model, quantifying the m⁵C modification patterns utilizing six DEGs associated with disease prognosis. The resulting scoring system allowed us to categorize patients into high- or low-risk groups based on their m⁵C scores. In test (TCGA-GBM) and validation (Chinese Glioma Genome Atlas [CGGA]-1018 and CGGA-301) datasets, glioma patients with a higher m⁵C score consistently exhibited shorter survival durations, fewer isocitrate dehydrogenase (IDH) mutations, less 1p/19q codeletion and higher World Health Organization (WHO) grades. Additionally, distinct immune cell infiltration characteristics were observed among different m⁵C cluster groups and risk groups. Our study developed a novel prognostic scoring system based on m⁵C modification patterns for glioma patients, complementing existing molecular classifications and providing valuable insights into prognosis for glioma patients.

INTRODUCTION

Diffuse glioma is the most common type of primary brain tumor, arising from astrocytes, oligodendrocytes, oligodendroglia-astrocytes or ependyma. In the fifth edition of the World Health Organization (WHO) Classification of Tumors of the Central Nervous System

(WHO CNS 5),¹ diffuse glioma can be graded from WHO grades 1–4, with higher grades indicating aggressive progression. Isocitrate dehydrogenase (IDH) wild-type glioblastoma (GBM) is designated as WHO grade 4, which has extreme malignant consequences even with high-intensity combination treatment of operation and radiation, chemotherapy, hormonotherapy, or immunotherapy.² The median survival time of GBM patients is only 12–15 months, and only 3%–5% of patients live longer than 3 years.³

The traditional classification system for diffuse glioma is largely based on histopathologic features and cannot accurately explain the biological behaviors of tumors.^{4,5} For example, glioma categorized as mixed oligoastrocytoma can be considered either “low grade” or “high grade.”⁶ Misclassification prevents patients from receiving the most suitable therapy strategy, which may cause patients to miss the optimal treatment time. Therefore, a new classification system was established by the WHO in 2021. This classification system is more accurate in clinical diagnosis and prediction for the prognosis of glioma patients by integrating multiple morphological and molecular markers, such as IDH, 1p/19q codeletion, ATRX (a chromatin remodeler protein, is recurrently mutated in H3F3A-mutant pediatric glioblastoma) mutant, TP53 mutant, and others.⁴ Based on the new

Received 13 June 2023; accepted 1 March 2024;
<https://doi.org/10.1016/j.omton.2024.200790>.

⁷These authors contributed equally

Correspondence: Haiyan Hu, MD, PhD, Department of Oncology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, P.R. China.
E-mail: xuri1104@163.com

Correspondence: Shitao Rao, MD, PhD, Department of Bioinformatics, Fujian Key Laboratory of Medical Bioinformatics, School of Medical Technology and Engineering, Fujian Medical University, Fuzhou 350122, China.
E-mail: strao@fjmu.edu.cn

Correspondence: Xiaoya Gao, PhD, Department of Pediatric Neurology, Zhujiang Hospital of Southern Medical University, 253 Gongye Avenue, Guangzhou, Guangdong 510282, P.R. China.
E-mail: gaoxy23@126.com



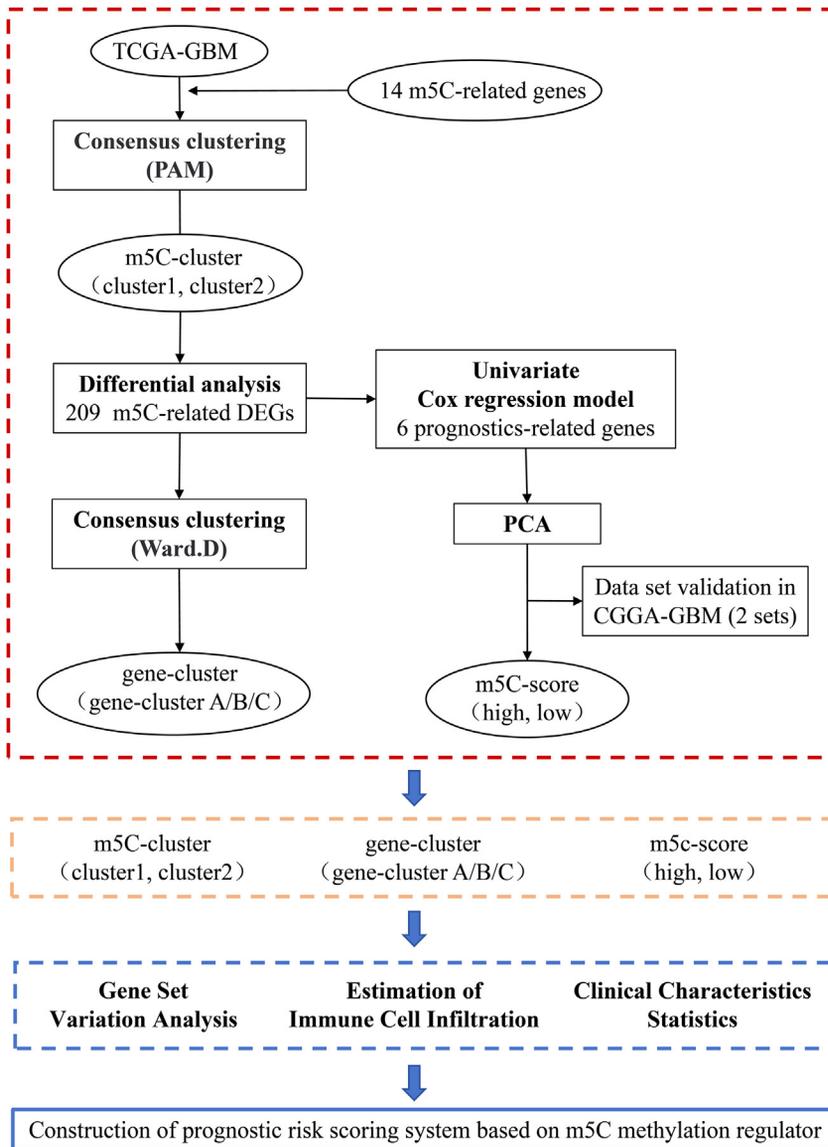


Figure 1. Flowchart of the present study

Three classifications (m⁵C cluster, gene cluster, and m⁵C score) were identified based on the expression of 14 m⁵C regulators in the TCGA-GBM testing dataset. Subsequently, we estimated TME cell infiltration and analyzed patients' clinical characteristics among the three different classifications. PAM: partitioning around medoids; DEGs: differentially expressed genes; PCA: principal component analysis.

RNA could activate oncogenesis-related pathways and create a microenvironment that is conducive to the migration and metastasis of cancer cells in skin cancer, bladder carcinoma, prostate cancer, and breast cancer.^{3,4}

The m⁵C methylation, which occurs at the fifth N position of cytosine nucleotides in coding RNA or noncoding RNA domains, can regulate stem cell stress, cytotoxic stress, mRNA nucleation, and gene expression. The m⁵C methylation and those m⁵C-regulated genes have been reported to be linked to the epigenetic mechanisms for the tumorigenesis, aggressiveness, and malignancy of diffuse glioma,⁷ although the precise mechanism is not yet clear. For example, the loss of ten-eleven translocation 2 (TET2), which is responsible for the reversible conversion of 5-methylcytosine to 5-hydroxymethylcytosine, has been linked to GBM stem cells and poor survival rates of GBM patients.⁹ However, there is still a lack of systematic analyses of the correlations between m⁵C-regulating genes and glioma prognosis.

The present study therefore aimed to systematically integrate all putative m⁵C regulators to construct a reliable scoring system to quantify

the m⁵C modification pattern in individual glioma patients and further investigate the tumor microenvironment (TME) cell infiltration characteristics mediated by all regulators to enhance our understanding of TME immune regulation in patients.

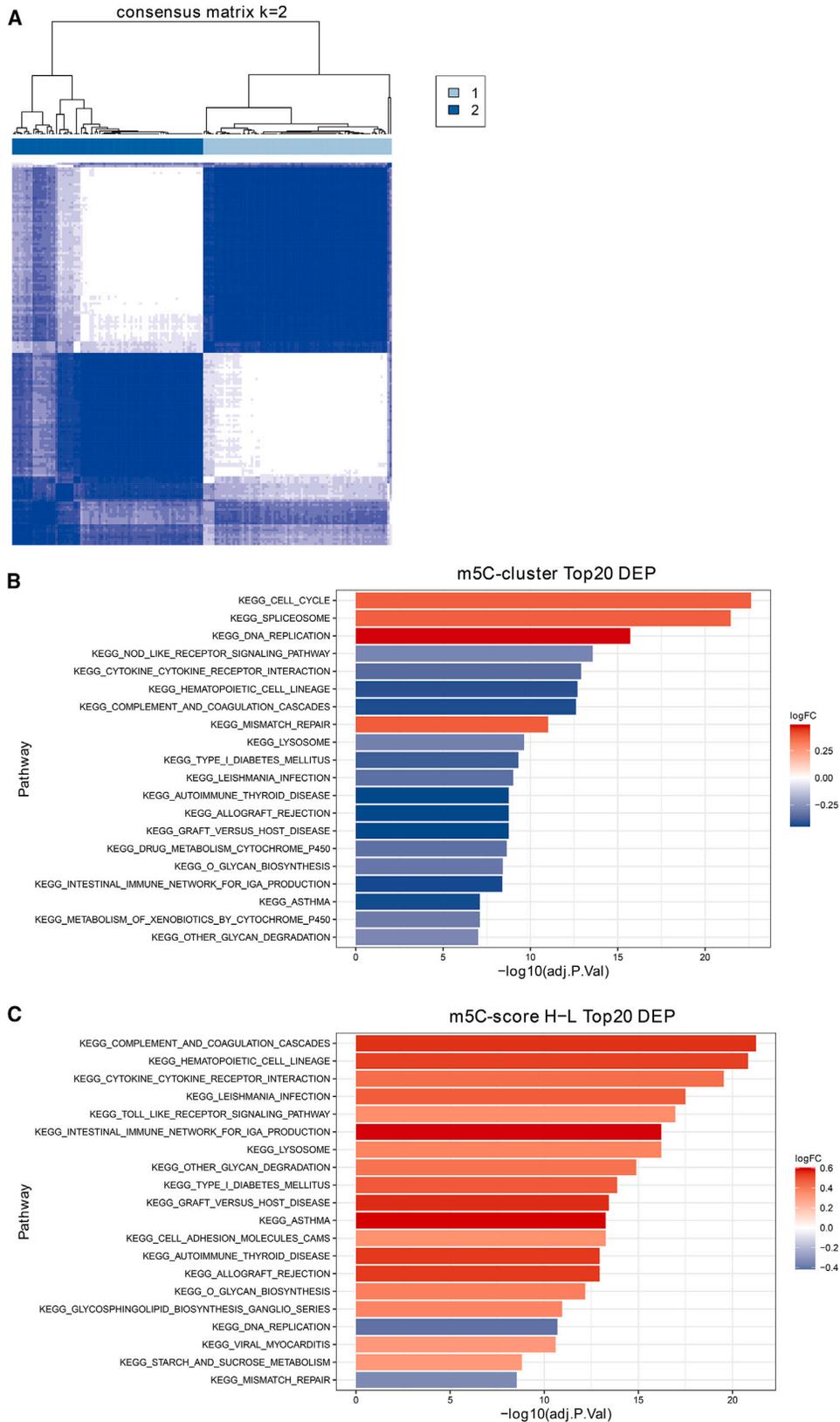
RESULTS

Construction and functional annotations of m⁵C cluster modification patterns

From the test dataset (The Cancer Genome Atlas [TCGA]-GBM), we successfully extracted the gene expression of 10 m⁵C regulators and performed unsupervised clustering analysis with the "ConsensusClusterPlus" package, which eventually determined 2 stable modification patterns (m⁵C cluster 1, 83 subjects; m⁵C cluster 2, 84 subjects) (Figures 1 and 2A; Tables S2–S4). To further investigate the differential biological behaviors between the two clusters, we performed gene

system, adult diffuse gliomas are divided into three main categories: astrocytoma with IDH mutant, oligodendroglioma with IDH mutant and 1p/19q codeleted, and GBM with IDH wild type. In this manner, low-grade diffuse gliomas (WHO grades 2 and 3) are characterized by the presence of IDH mutations. In contrast, IDH wild-type astrocytoma GBM has invasive biological behaviors.⁷

Recently, several studies highlighted the role of RNA modifications, such as N5-methylcytosine (m⁵C) or N6-methyladenine (m⁶A) methylation, in regulating the initiation and progression of glioma.⁶ RNA methylation is responsible for more than 60% of RNA epigenetic modifications in eukaryotes, regulating the expression of genes and affecting the biological behavior of cells.³ The m⁶A and m⁵C modifications are the two most common methylation forms, with high abundance and stability in cells.^{2,8} These two methylations of



(legend on next page)

set variation (GSVA) pathway variation analysis with the Molecular Signatures Database (MSigDB) reference set. The preprocessed 11,668 genes from the TCGA-GBM dataset were enriched in 169 pathways. Then, we conducted differential expression pathway analysis with the “limma” package and found a total of 110 significantly different pathways (adjusted $p < 0.05$; Figures 2B and S1; Tables S2–S5). The top 20 differential pathways shown in Figure 2B suggested that the two clusters were significantly different in pathways involving cellular components and DNA stability, such as spliceosome, lysosome, cell cycle, DNA replication, and mismatch repair.

For the two distinct m^5C clusters, we estimated the relative proportion of immune matrix components in the TME for each sample via the estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) algorithm (Figures 3A–3C). The results showed that the glioma patients in m^5C cluster 2 had significantly higher stromal, immune, and ESTIMATE scores than those in m^5C cluster 1 ($p < 0.001$). Furthermore, we calculated the relative proportion of 22 types of immune cells for each sample and found that four types of cells had a relatively high proportion in these samples, with a range of $\sim 1\%$ – 40% (Figure 3D). To avoid any potential over-interpretation of the statistical results and minimize the risk of false positive signals, immunological cells with a proportion exceeding 1% were primarily presented here. Among them, the proportions of m^5C cluster 1 samples were significantly lower in resting memory CD4 T cells (Q2 [second quartile], cluster 1: 9.88%; cluster 2: 11.15%; $p < 0.0001$) and monocytes (Q2, cluster 1: 8.67%; cluster 2: 18.01%; $p < 0.0001$) and prominently higher in M0 macrophages (Q2, cluster 1: 10.10%; cluster 2: 2.14%; $p < 0.0001$; Figure 3D) and M2 macrophages (Q2, cluster 1: 38.37%; cluster 2: 37.68%; $p < 0.0001$; Figure 3D).

Construction and functional annotations of gene cluster modification patterns

To investigate the gene expression profile in the two distinct m^5C cluster modification patterns, we conducted differentially expressed gene (DEG) analysis with the “limma” package and identified 209 DEGs between the two clusters ($|\log_2FC| > 1$ (FC, stands for fold change, represents the multiple change in the expression level of the target molecule under the experimental condition relative to the control condition) and false discovery rate [FDR] < 0.05 ; Tables S2–S6; Figure S2). These genes were considered m^5C regulation-related genes and further utilized for gene cluster consensus clustering analysis. As shown in Figure S3, the 167 samples in the TCGA-GBM dataset could be divided into three distinct clusters (gene cluster A, 53; gene cluster B, 87; and gene cluster C, 27; Tables S2–S4). Then, we performed GSVA pathway variation analysis for the 3 gene clusters with pairwise comparisons, which identified 82, 114, and 62 dif-

ferential pathways for gene cluster A vs. B, B vs. C, and A vs. C comparisons, respectively (adjusted $p < 0.05$). Notably, the 3 sets of differential pathways were mainly involved in pathways for DNA stability and the immune system, such as DNA replication, mismatch repair, the cytosolic DNA sensing pathway, antigen processing and presentation, autoimmune thyroid disease, and the intestinal immune network (Figures S4 and S5; Tables S2–S7).

For the three distinct gene cluster patterns, we also estimated the relative proportion of immune matrix components in the TME for each sample and observed that glioma patients in gene cluster C had the lowest score in all three components compared with those in gene clusters A and B ($p < 0.001$; Figures S6A–S6C). The calculation of the relative proportion of 22 types of immune cells identified that the three gene cluster patterns mainly consisted of resting memory CD4 T cells, monocytes, M0 macrophages, and M2 macrophages (range, $\sim 1\%$ – 40%). Among them, gene cluster A contained the largest proportions of CD4 memory resting cells (Q2, A: 11.58%; B: 10.81%; C: 7.93%; $p < 0.0001$) and M2 macrophages (Q2, A: 39.15%; B: 38.15%; C: 36.37%; $p < 0.0001$), while gene cluster B had the largest proportions of monocytes (Q2, A: 11.23%; B: 16.79%; C: 4.91%; $p < 0.0001$). Moreover, gene cluster C contained the largest proportion of M0 macrophages (Q2, A: 6.57%; B: 2.27%; C: 27.54%; $p < 0.0001$; Figure S6D).

Construction and survival analysis of m^5C score modification patterns

For the 209 DEGs between the two distinct m^5C cluster patterns, univariate Cox regression analyses identified a set of six genes with prominent associations with disease prognostics ($p < 0.01$; *IGFBP6*, *CXCL2*, *SERPINA1*, *NEU4*, *CLEC2B*, and *CHI3L1*; Figures S7A–S7F). As shown in Figure S7, patients with a lower level of gene expression in five of the six genes, except the *NEU4* gene, had a higher survival probability. Based on the expression of the six genes and the calculated principal-component analyses (PCAs) from the TCGA-GBM test dataset, a stable m^5C score risk evaluation model was constructed with the maximum rank statistic and further utilized to classify glioma patients into m^5C -score-high and -low groups (m^5C -score-high group, 147; m^5C -score-low group, 20; Tables S2–S4). The log rank tests identified a significant difference between the two groups, and the low-score group of patients had a prominently longer overall survival (OS) time (hazard ratio [HR], 0.31; 95% confidence interval [CI], 0.16–0.62; $p = 8.77 \times 10^{-4}$; Figure 4A). We also successfully constructed the m^5C score risk system in the two validation datasets Chinese Glioma Genome Atlas (CGGA)-301 and CGGA-1018 (CGGA-301, m^5C -score-high group: 156 and m^5C -score-low group: 145; CGGA-1018, m^5C -score-high group: 536 and m^5C -score-low group: 482) and observed the consistent trend that patients with a lower score had a longer OS time (HR, 0.29; 95% CI, 0.21–0.40; $p =$

Figure 2. Patterns of m^5C methylation modification and differential pathway enrichment analysis for the two m^5C clusters

(A) Consensus clustering analysis according to 10 m^5C regulators in the TCGA-GBM dataset. (B) Top 20 differentially expressed pathways (DEPs) between the two distinct clusters. (C) Top 20 DEPs between the two m^5C score groups. The x axis represents \log_{10} (adjusted p), and the y axis shows the name of each pathway. Different colors represent opposite regulation directions of each pathway: red indicates upregulation, and blue indicates downregulation (m^5C cluster 2 and m^5C -score-low group as references, respectively, in B and C).

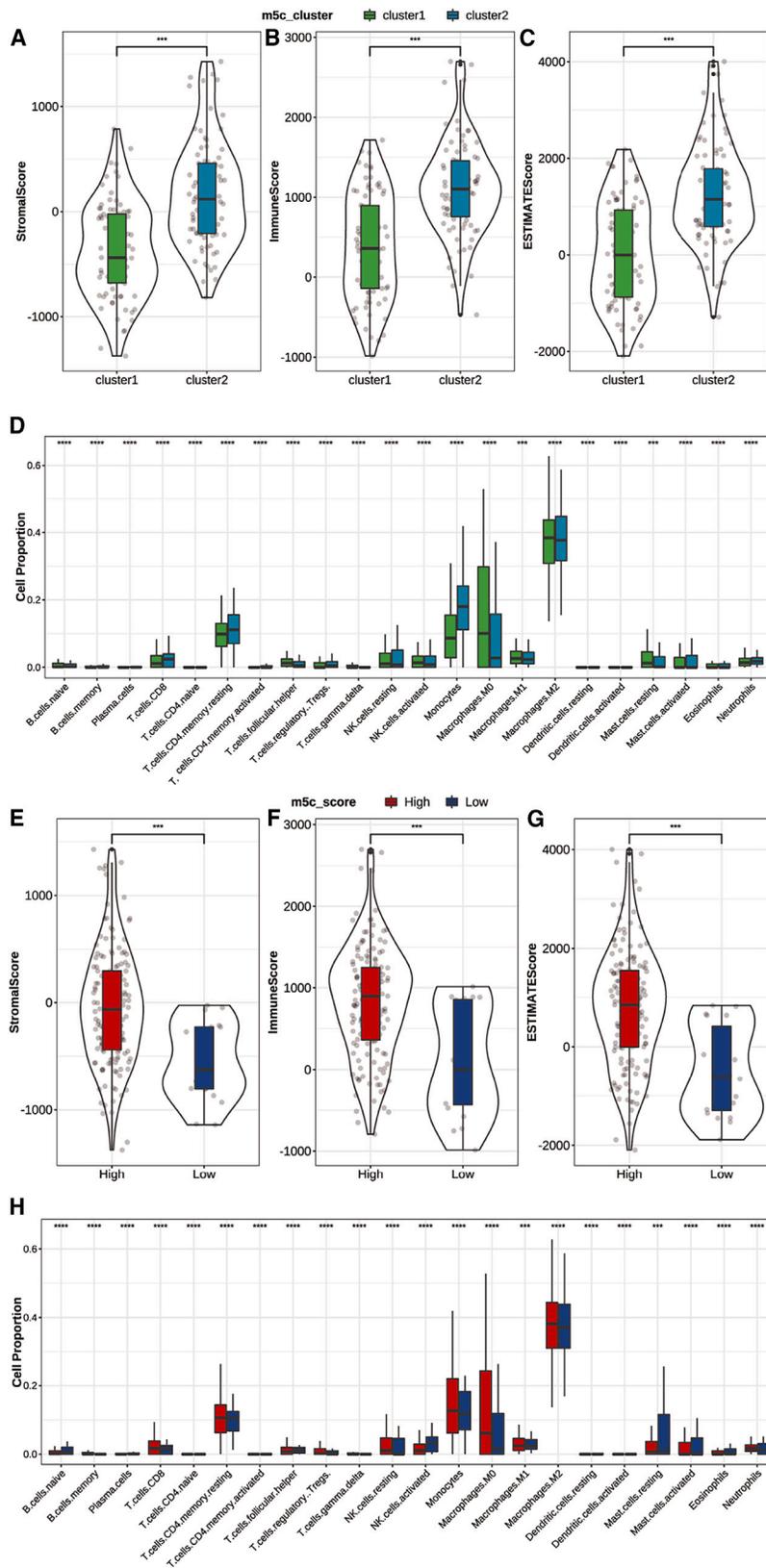


Figure 3. TME cell infiltration characteristics

(A–D) TME cell infiltration characteristics in two distinct m⁵C cluster patterns. Shown are comparison of the relative scores of immune matrix components (A–C) and comparison of the relative proportions of 22 types of TME-infiltrating cells for each sample (D). ns, p ≥ 0.05; *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. (E–H) TME cell infiltration characteristics in the high- and low-m⁵C-score groups. Shown are comparison of the relative scores of immune matrix components between the two groups (E) and comparison of the relative proportions of 22 types of TME-infiltrating cells for each sample (F). ns, p ≥ 0.05; *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

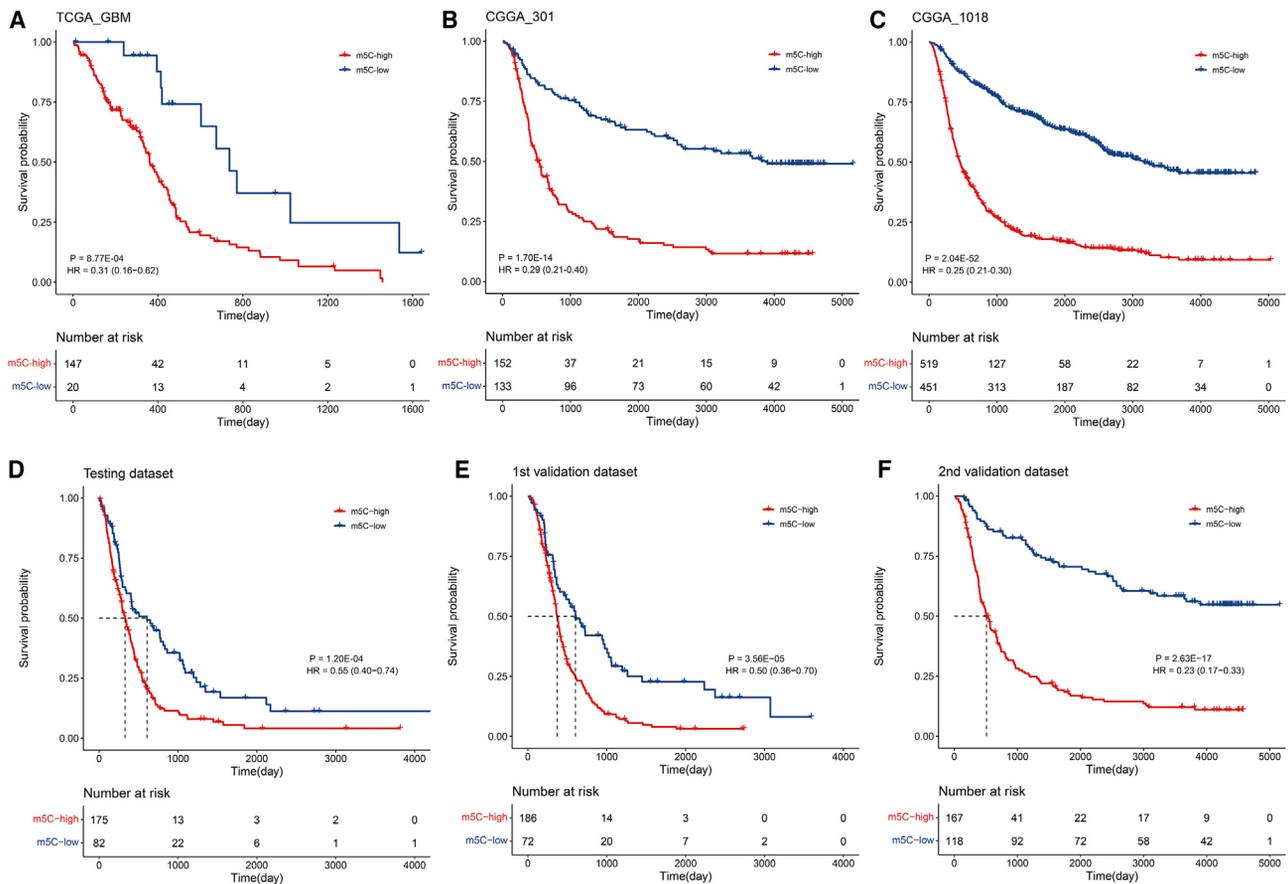


Figure 4. Comparison of overall survival (OS) time between the high- and low-m⁵C-score groups

(A) OS curves drawn by the Kaplan-Meier method in the TCGA-GBM test dataset. (B) OS curves drawn by the Kaplan-Meier method in the CGGA-301 validation dataset. (C) OS curves drawn by the Kaplan-Meier method in the CGGA-1018 validation dataset. (D) OS curves drawn by the Kaplan-Meier method in the testing dataset. (E) OS curves drawn by the Kaplan-Meier method in the first validation dataset. (F) OS curves drawn by the Kaplan-Meier method in the second validation dataset. The x axis represents the survival time (days), and the y axis shows the survival probability (percentage).

1.70×10^{-14} ; Figure 4B; HR, 0.25; 95% CI, 0.21–0.30; $p = 2.04 \times 10^{-52}$; Figure 4C). For another m⁵C score risk model that utilized the combined WHO grade 4 samples from the TCGA-GBM and CGGA-1018 datasets as the testing dataset, the m⁵C-score-low patients were found to have a longer OS time in the testing dataset, the first validation dataset, and the second validation dataset, respectively (HR, 0.55; 95% CI, 0.40–0.74, $p = 1.20 \times 10^{-4}$; Figure 4D; HR, 0.50; 95% CI, 0.36–0.70; $p = 3.56 \times 10^{-5}$; Figure 4E; HR, 0.23; 95% CI, 0.17–0.33; $p = 2.63 \times 10^{-17}$; Figure 4F). These findings support the robustness of the two different scoring systems.

We performed GSVA for the two groups and identified 75 differential biological pathways (adjusted $p < 0.05$; Tables S2–S8; Figures 2C and S8). These pathways were mainly involved in immune-related functions (cytokine-cytokine receptor interaction, antigen processing and presentation, Toll-like receptor signaling pathway, and intestinal immune network) and cellular components (lysosome, ribosome, and cell adhesion molecules). Following the findings, we applied Pearson's correlation coefficients to evaluate the relationship

of the constructed m⁵C score with a set of 10 immune checkpoint-related genes, which were successfully matched in the TCGA-GBM dataset. As shown in Figure S9, the constructed m⁵C score patterns had significant positive correlations with a total of 8 target genes (range of R value, 0.25–0.71; $p < 0.0001$) and had a significant inverse correlation with the *CD200* gene ($R = -0.35$, $p < 0.0001$).

For the high and low m⁵C score groups, we further estimated the relative proportion of immune matrix components in the TME for each sample and found that glioma patients with a high score had significantly higher stromal, immune, and ESTIMATE scores than those with a low score ($p < 0.001$; Figures 3E–3G). The calculation of the relative proportion of 22 types of immune cells implied that the high-m⁵C-score group contained a higher proportion of monocytes (Q2, high group: 12.69%; low group: 11.75%, $p < 0.0001$), M0 macrophages (Q2, high group: 6.15%; low group: 1.50%, $p < 0.0001$), and M2 macrophages (Q2, high group: 38.15%; low group: 37.12%, $p < 0.0001$), while the low-m⁵C-cluster group had a higher proportion of activated natural

killer (NK) cells (Q2, high group: 1.09%; low group: 3.38%, $p < 0.0001$; Figure 3H).

Comparison of clinical characteristics between distinct m⁵C modification patterns

In comparison with clinical characteristics and those well-known molecular subtypes for the 3 sets of distinct m⁵C modification patterns in the TCGA-GBM dataset (m⁵C cluster, gene cluster, and m⁵C score), we observed that patients in different groups were sex matched. Interestingly, we found significant differences in the history of neoadjuvant treatment and in the IDH mutation status between the two m⁵C cluster patterns ($p = 0.04$ and 0.01 , respectively; Figures 5E and 5F) as well as the IDH mutation status between the high- and low-m⁵C-score groups ($p = 9.98 \times 10^{-12}$; Figure 5N), implying that the constructed m⁵C modification patterns and m⁵C risk groups were highly consistent with the well-known IDH mutation status for glioma patients (Figure 6A). However, we did not find any significant difference in clinical characteristics or molecular subtypes among the three m⁵C regulation-related gene cluster patterns (Figure S10).

For the two validation datasets, the patients in the low- and high-m⁵C-score groups were also sex matched to those in the test dataset (Figures 7A and 7I). In the CGGA-301 dataset, we observed that patients with a high m⁵C score had a substantially higher proportion of disease primary rate and WHO grade 4 ($p = 8.26 \times 10^{-22}$; Figure 7E) as well as a higher proportion of IDH mutation status and 1p19q codeletion status ($p = 2.13 \times 10^{-17}$, Figure 7F; $p = 1.18 \times 10^{-4}$, Figure 7G). Additionally, the high-m⁵C-score group also had a higher rate of chemotherapy ($p = 4.15 \times 10^{-3}$; Figure 7B). For another validation dataset, CGGA-1018, we also observed a similar phenomenon for the chemotherapy rate, disease primary rate, WHO 4 grade, IDH mutation status, and 1p19q codeletion status in patients with a high m⁵C score (Figures 7J and 7L–7O). In addition, the high-m⁵C-score group had a lower rate of MGMT (a gene located on chromosome 10q26, encodes a DNA repair protein responsible for removing alkyl groups from the O6 position of guanine, a crucial site for DNA alkylation) promoter methylation status ($p = 0.02$; Figure 7P). These findings suggested that glioma patients with a higher m⁵C score in the two validation datasets usually had worse well-known molecular subtypes and a higher WHO grade level.

DISCUSSION

The present study successfully constructed, for the first time, a relatively reliable scoring system to quantify the m⁵C modification pattern in individual glioma patients by integrating all putative m⁵C regulators from the test dataset TCGA-GBM, which was further confirmed by more glioma patients in the other two validation datasets. Notably, the glioma patients in all 3 datasets with a higher m⁵C score were consistently found to have a lower OS probability (HR < 0.31, $p < 8.77 \times 10^{-4}$), suggesting the robust stability and universality of the constructed scoring system. In addition, we also observed that this scoring system had a strong correlation with immune heterogeneity and the well-known molecular subtypes for glioma patients.

Helpful for differentiation of immunologic heterogeneity for glioma patients with the constructed m⁵C clusters and scoring system

The current study employed the ESTIMATE algorithm, which obtained significantly higher scores in stromal, immune, and ESTIMATE in patients with m⁵C cluster 2 (Figure 3A) and with a high m⁵C score (Figure 3E). Moreover, we also observed statistically significant differences in the proportions of 22 types of immune cells in different m⁵C cluster and m⁵C score groups, implying an essential role of m⁵C modification patterns in mediating individual glioma infiltration characteristics. Additionally, the identified differential pathways were mainly involved in immune-related functions. Taken together, these findings suggested that the constructed m⁵C clusters and scoring system may be helpful for differentiation of immunologic heterogeneity for glioma patients. High-grade glioma patients have an abysmal prognosis even when undergoing a combination of therapies, including operation, radiation, chemotherapy, hormonotherapy, and immunotherapy.^{10,11} The significant heterogeneity of gliomas in terms of the composition of the immune microenvironment and gene mutations could account for their varied biological behavior.¹² Impaired regulation of the immune response and immune evasion could cause tumorigenesis, invasion, and metastasis. Monocytes, eosinophils, and neutrophils are part of the innate immune system.^{12,13} The adaptive immune system is essential in recognizing the antigen of tumors and providing helper and killer functions for tumors but often fails to establish immune memory.^{14,15} In addition, activated dendritic cells (DCs) can present antigens to CD4⁺ T cells and recruit monocytes or macrophages to the tumor site.¹⁶

The brain immune response is mediated mainly by myeloid cells.¹⁷ Thus, macrophages and T cell-dependent immune responses play crucial roles in tumor behavior. Specific CD4⁺ T cells are able to determine whether the disease course is monophasic or relapsing. A higher percentage of CD8⁺ cytotoxic T and NK cells is associated with an enhanced antitumoral immune response.^{10,18} Moreover, activated macrophages (M1) are supposed to induce antitumoral responses through proinflammatory activity, whereas M2 macrophages are thought to be protumoral.^{15,19} According to our statistical results for the proportions of 22 types of immune cells, primarily focusing on immunological cells with a proportion exceeding 1%, we observed a higher percentage of M2 macrophages and a lower percentage of memory CD4 resting T cells in patients with the constructed high m⁵C score, which could partially explain the shorter survival duration for this group of patients.

Strong correlation of the constructed m⁵C clusters and scoring system with the well-known molecular subtypes and grade levels

The two m⁵C modification patterns (m⁵C cluster and m⁵C score) were found to be associated with the patients' therapy history (chemotherapy, neoadjuvant treatment, radiotherapy, or hormonotherapy) in the test and validation datasets. Moreover, the patients with a high m⁵C score in the validation datasets showed significantly fewer IDH mutations and 1p19q codeletion as well as a substantially

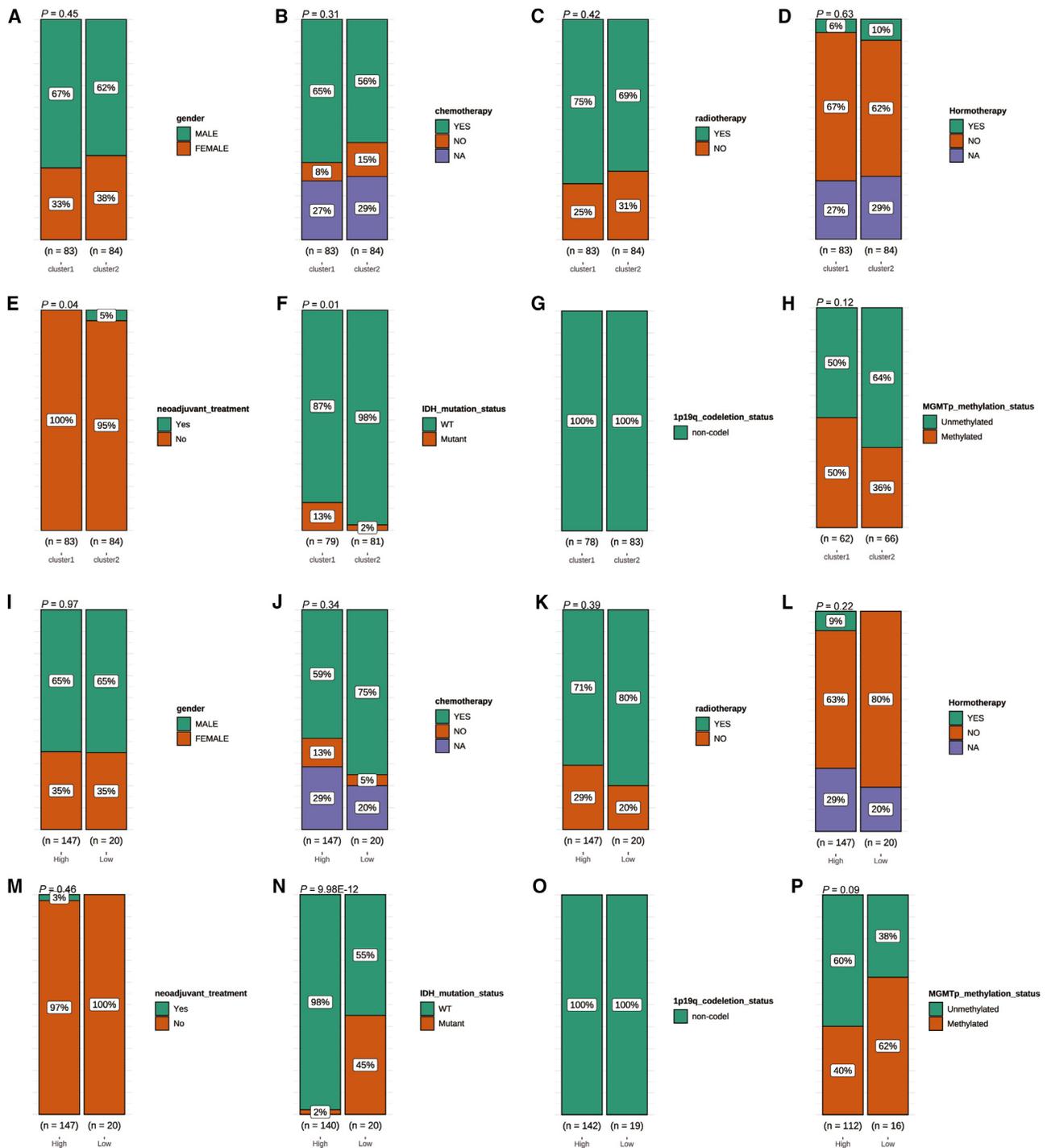


Figure 5. Comparison of clinical characteristics between the two m⁵C cluster modification patterns/m⁵C score groups in the TCGA-GBM dataset

(A–H) Comparison of clinical characteristics between the two m⁵C cluster modification patterns/m⁵C score groups in the TCGA-GBM dataset. (I–P) Comparison of clinical characteristics between the two in the TCGA-GBM dataset. Significant: $p < 0.05$; non-significant: $p \geq 0.05$.

higher proportion in WHO high-level grade, suggesting much poorer outcomes for this group of patients. These findings demonstrated the efficacy of m⁵C modification models in predicting the immunologic

response to glioma and the potential efficacy of immunotherapy. It has been reported that mutations in the IDH gene are frequently seen in infiltrating grade 2 and 3 gliomas of adults as well as secondary

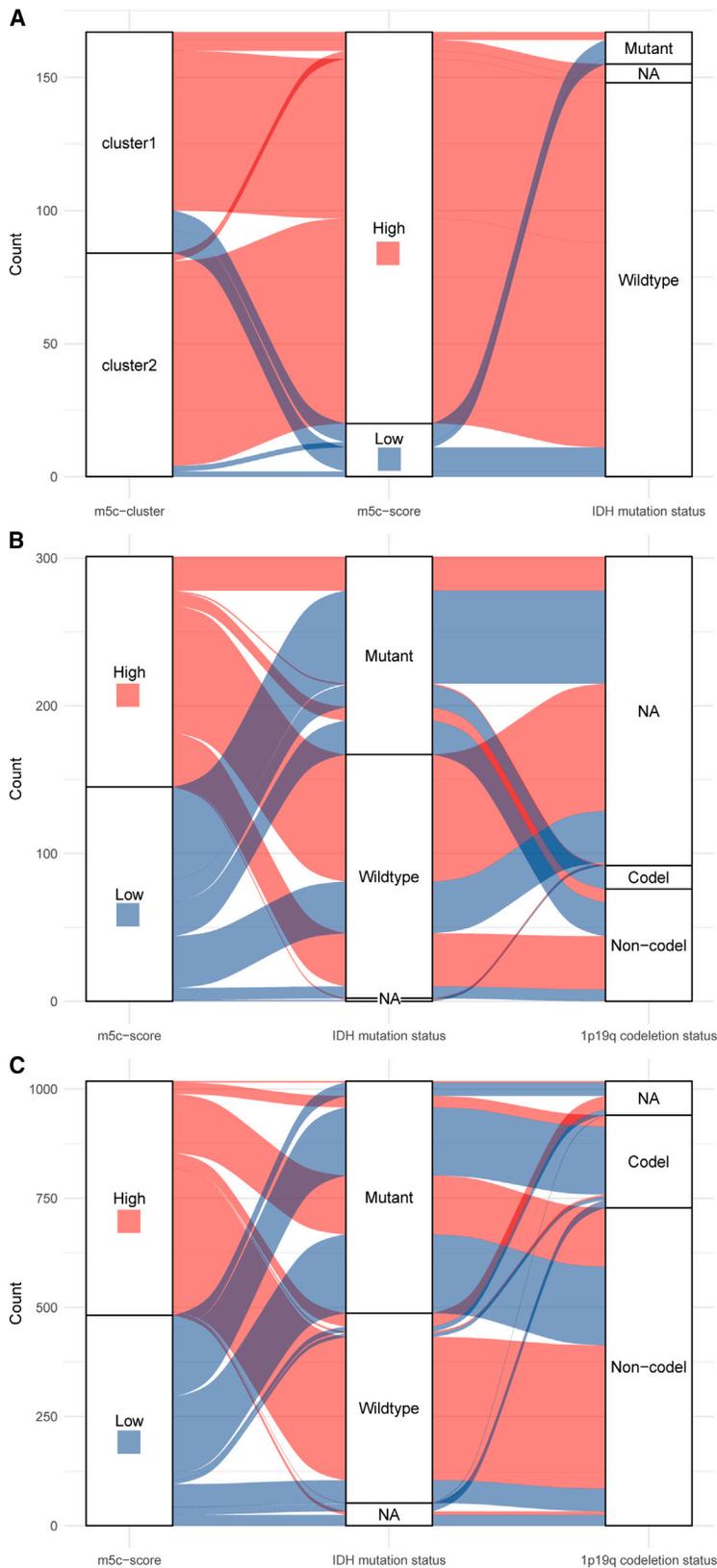


Figure 6. Sankey diagrams showing the changes in m⁵C cluster, m⁵C score, the well-known IDH mutation, and 1p19q codeletion status

(A) Sankey diagram for TCGA-GBM. Left: m⁵C cluster. Center: m⁵C score. Right: IDH mutation status. (B) Sankey diagram for CGGA-301. Left: m⁵C score. Center: IDH mutation. Right: 1p19q codeletion status. (C) Sankey diagram for CGGA-1018. Left: m⁵C score. Center: IDH mutation. Right: 1p19q codeletion status.

GBMs and are significant factors in discriminating the biologic class.²⁰ Glioma patients with IDH and 1p19q codeletion have a longer survival duration.² Mutations in the IDH gene could lead to low IDH enzyme activity, preventing the efficient conversion of 2-oxoglutarate to R-2-hydroxyglutarate (R-2-HG). This conversion inhibits enzymes that regulate transcription and metabolism in nuclear, cytoplasmic, and mitochondrial biochemistry.²¹ Moreover, patients with the 1p/19q codeletion have been reported to be sensitive to chemotherapy.^{22–24}

Limitations

There are several limitations in the present study. The first is that all patients in the test dataset TCGA-GBM were classified as WHO grade 4, while less than 50% of patients in the two validation datasets were classified as grade 4 glioma, which may induce a statistical bias to the constructed m⁵C scoring system from the relatively pure test dataset. However, we observed a similar trend where patients with a higher m⁵C score had a shorter OS time in all 3 datasets, which, in turn, implies the robustness in stability and universality of the system. Second, we could not fully include all 14 m⁵C-related regulators to construct the m⁵C cluster patterns, as 4 of them were not matched in the test dataset, which may result in missing values for the 4 regulators contributing to the cluster patterns. In addition, m⁵C and m⁶A are the two most common methylation forms, with high abundance in eukaryotic cells.² Including the two methylation forms simultaneously in one system may substantially improve the accuracy and efficacy of the prediction system.

Conclusions

In summary, the present study proposed a new molecular classification method for glioma patients, based on the m⁵C score methylation scoring system, that had strong correlations with the immune cell infiltration, therapeutic response, survival duration, and WHO grade of the patients.

MATERIALS AND METHODS

Flowchart of the study

In this study, we first identified two distinct m⁵C cluster modification patterns according to the expression of 10 m⁵C regulators in the preprocessed 11,668 genes from the TCGA-GBM test dataset. We then recognized three distinct gene cluster patterns based on DEGs between the two m⁵C cluster patterns (see flowchart in Figure 1). Among these DEGs, we further identified a set of six genes with a prominent association with disease prognostics and then established a set of m⁵C score systems to quantify the m⁵C modification pattern in individual patients. Next, we analyzed the TME cell infiltration characteristics and clinical characteristics in high- and low-m⁵C-score patients.

Glioma dataset source and preprocessing

By systematically searching the TCGA and the CGGA databases, we obtained three gene expression datasets and patients' full clinical annotation information, including their basic information, therapy methods, molecular subtypes, and survival data (Tables 1 and S2,

T1-T3). All of them were glioma samples and were named TCGA-GBM, CGGA-1018 (RNA-seq_1018), and CGGA-301 (mRNA-array_301), respectively. To maintain the consistency of gene expression data for further analyses, RNA sequencing data (FPKM value) (fragments per kilobase of transcript per million mapped reads, is a normalized measure of gene expression, accounting for both gene length and total read count) were downloaded from all three datasets for further analyses and comparisons. For TCGA-GBM, expression data were obtained from the Genomic Data Commons (GDC) using the R package TCGA Biolinks.²⁵ After that, genes with FPKM values of less than 1 in over 20% of the samples and FPKM values equal to 0 in over 50% of the samples were removed from further analysis (Table S1). For CGGA-1018 (Table S2) and CGGA-301 (Table S3), the FPKM data had already been preprocessed and were directly downloaded from the CGGA database (<http://www.cgga.org.cn>).

Unsupervised clustering for 14 m⁵C regulators

A total of 14 m⁵C regulators or related genes were systematically searched from various reports and the literature: 11 writers, 1 eraser, and 2 readers (*NSUN1*, *NSUN2*, *NSUN3*, *NSUN4*, *NSUN5*, *NSUN6*, *NSUN7*, *DNMT1*, *DNMT2*, *DNMT3A*, *DNMT3B*, *TET2*, *ALYREF*, and *YBX1*).^{7,9,26–36} Ten of them were detected from the test dataset (TCGA-GBM) and finally utilized to construct the m⁵C modification patterns. Unsupervised clustering analysis was employed to identify distinct patterns (m⁵C cluster) based on the gene expression of the 10 regulators and to classify glioma patients into different distinct patterns. The “ConsensusClusterPlus” R package was applied here to perform clustering analysis for identifying distinct patterns with the following parameters: the “partitioning around medoids” (PAM) algorithm for clustering analysis to determine the number of clusters, Euclidean distance for calculating pattern distance, 80% of total sample for resampling, and 100 repetitions to guarantee the stability of classification.³⁷

Identification of DEGs between distinct m⁵C patterns and construction of DEG-based clusters

After obtaining stable consensus clusters, the empirical Bayesian approach of the “limma” R package was utilized to determine DEGs between different modification patterns.³⁸ The Benjamini-Hochberg method was applied to control the FDR.³⁹ Genes with |log₂FC| greater than 1 and FDR less than 0.05 were considered m⁵C regulation-related DEGs, which were visualized by the R package ggplot2. Based on those DEGs, unsupervised clustering analysis was also applied to identify distinct gene patterns (gene cluster) and classify glioma patients for further analysis. The “pheatmap” R package was utilized to classify three different gene clusters with the Ward.D algorithm.⁴⁰

Construction and validation of the m⁵C score prognostic risk model

With those identified DEGs between the two m⁵C cluster patterns, we constructed a set of scoring risk systems so that we could quantify the m⁵C modification patterns of individual tumors and eventually evaluate the prognostic risk score of individual patients with glioma (m⁵C

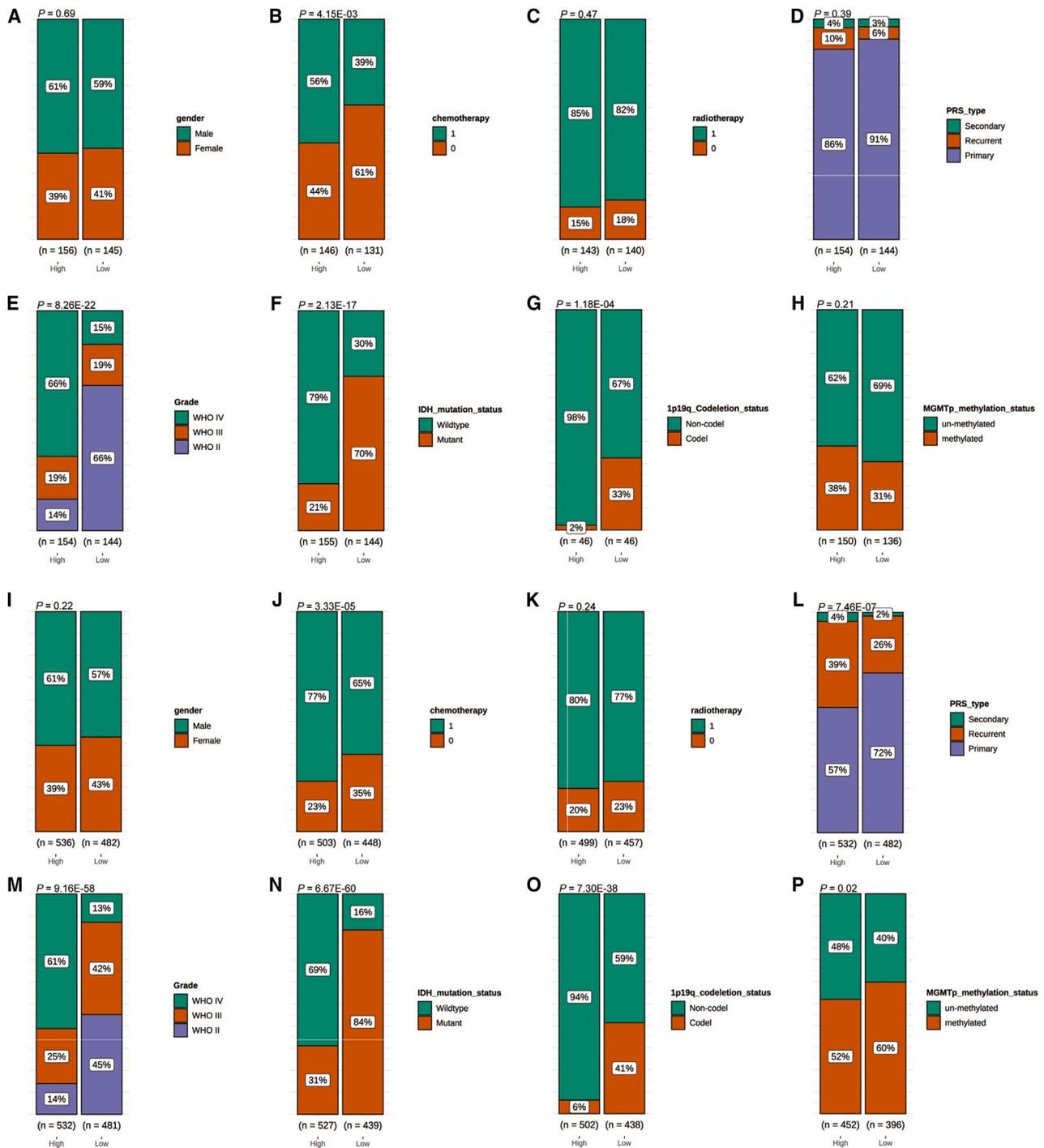


Figure 7. Comparison of clinical characteristics and molecular phenotypes between the two m⁵C score groups in the CGGA-301 and CGGA-1018 validation datasets

(A–H) Comparison of clinical characteristics and molecular phenotypes between the two m⁵C score groups in the CGGA-301 validation dataset. (I–P) Comparison of clinical characteristics and molecular phenotypes between the two m⁵C score groups in the CGGA-1018 validation dataset. Significant: $p < 0.05$; non-significant: $p \geq 0.05$.

Table 1. Full clinical annotation information for the three sets of glioma datasets

	TCGA-GBM	CGGA_1018	CGGA_301
Platform	Illumina RNAseq	Illumina HiSeq	Agilent Technologies Whole Human Genome (array)
Number of genes	11,668	23,271	19,416
Number of patients	167	1,018	301
Sex	female: 59 male: 108	female: 417 male: 601	female: 121 male: 180
Grade	WHO 4: 167	WHO 2: 291 WHO 3: 334 WHO 4: 388 N/A: 5	WHO 2: 117 WHO 3: 57 WHO 4: 124 N/A: 3
PRS type	N/A	primary: 651 recurrent: 333 secondary: 30 N/A: 4	primary: 264 recurrent: 23 secondary: 11 N/A: 3
Chemotherapy	yes: 101 no: 20 N/A: 46	yes: 679 no: 272 N/A: 67	yes: 133 no: 144 N/A: 24
Radiotherapy	yes: 120 no: 47	yes: 754 no: 202 N/A: 62	yes: 237 no: 46 N/A: 18
IDH mutation status	mutant: 12 wild type: 160 N/A: 7	mutant: 531 wild type: 435 N/A: 52	mutant: 134 wild type: 165 N/A: 2
1p19q codeletion status	code1: 0 not code1: 161 N/A: 6	code1: 212 not code1: 78 N/A: 728	code1: 16 not code1: 76 N/A: 209
MGMTp methylation status	methylated: 55 unmethylated: 73 N/A: 39	methylated: 472 unmethylated: 170 N/A: 113	methylated: 55 unmethylated: 187 N/A: 15
Hormonotherapy	yes: 13 no: 108 N/A: 46	N/A	N/A
Neoadjuvant treatment	yes: 4 no: 163	N/A	N/A
New tumor type	locoregional disease: 2 progression of disease: 75 Recurrence: 21 unknown: 69	N/A	N/A
Survival data	OS	OS	OS

TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas; N/A, not available; OS, overall survival.

score). Briefly, the risk scoring system was established and validated through four main steps as follows.

- (1) The univariate Cox regression model was applied for the prognostic analysis to select the DEGs that had a significant association with disease prognostics. We set the significance threshold at a stricter level with a p value of 0.01, which could ensure a strong association between the selected DEGs and disease prognostics.
- (2) With those prognostic-associated DEGs, PCA was performed to construct the m^5C score risking system. In brief, all principal components were calculated based on the expression data of the associated DEGs, and then the number of components having

a cumulative explained variance over 80% was selected to act as a signature score in the test dataset (TCGA-GBM). The method had a strong ability to ensure that the constructed risk model mainly stood for the original expression data and statistically explained enough variance of disease. The m^5C score for the i^{th} patient (S_i) was calculated by the equations we developed as follows:

$$A = VS^T$$

$$m^5C - score = \sum_{n=1}^j A_{ij}$$

where V is the feature vector ($n = 1$) or the feature vector matrix ($n > 1$), S is the gene expression matrix (row for each gene and column for each sample), and A is the n dimension matrix (row for each sample i and column for each principal component j).

- (3) Based on the constructed m^5C score for individual patients, the “surv-cutpoint” function in the “survminer” R package was employed to dichotomize the score, which could divide all the patients into m^5C -high and m^5C -low groups according to the maximally selected log rank statistics to decrease the batch effect of calculation. The applied function was designed to repeatedly test all potential cut points to find the maximum rank statistic. After that, the classical functions in the “survival” R package were adopted to determine the significance of differences by log rank tests and generate the 3-year OS curves via the Kaplan-Meier method.
- (4) The effectiveness of the constructed m^5C -score system was validated by using the other two validation datasets of glioma samples (CGGA-1018 and CGGA-301). In the two datasets, PCA was conducted, and the number of PCs having a cumulative explained variance over 80% was selected to act as a signature score, which could keep a consistent standard between the test and validation datasets. We also identified the significance of differences for the high and low m^5C score groups.

In addition to employing the methods above, we also utilized another approach to construct and validate the risk model with the three datasets. Notably, the CGGA-301 dataset (Agilent Technologies Whole Human Genome-GPL16022) was conducted on a different platform from the one used for the TCGA-GBM and CGGA-1018 datasets (Illumina HiSeq). Therefore, we initially combined the WHO grade 4 samples from the TCGA-GBM and CGGA-1018 datasets as the testing dataset and the first validation dataset after removing batch effects. Additionally, the grade 4 samples in the CGGA-301 dataset were considered as the second validation dataset. First, we removed low-expression genes with FPKM values of less than 1 in over 20% of the samples and FPKM values equal to 0 in over 50% of samples in the former two datasets, including all samples in the TCGA-GBM dataset and all WHO grade 4 samples in the CGGA-1018 dataset. To minimize potential heterogeneity between the two datasets, we further utilized the “Combat” package in R language programming to remove any potential batch effects.⁴¹ Outlier samples were also removed after performing a PCA. The remaining samples were equally divided into two parts in a random way, named the training dataset and the first validation dataset, respectively. Subsequently, we constructed the prognostic risk-scoring system from the training dataset following the flowchart in Figure 1 and validated the efficacy of the score system in the first and second validation datasets, respectively.

GSVA and functional annotations

To further explore the difference in biological processes between the distinct m^5C modification patterns (m^5C cluster, gene cluster, and m^5C score), the “GSVA” R package was applied to perform GSVA.

With the samples of an expression dataset, GSVA is usually applied to estimate the variation of a gene set in pathways and biological processes activity.⁴² A reference set of “c2.cp.kegg.v7.4.symbols” was downloaded from the public MSigDB to perform GSVA and obtain an enrichment scoring matrix. After that, the “limma” R package was employed to identify differentially expressed pathways and processes between those distinct patterns. An adjusted $p < 0.05$ was considered statistically significant.

Estimation of glioma TME cell infiltration

The ESTIMATE algorithm in the “estimate” R package was utilized to quantify the relative proportion of immune matrix components in the TME for each glioma sample.⁴³ Based on the gene expression of each glioma sample, the CIBERSORT (a method for deconvoluting the cell composition of complex tissues from their gene expression profiles) algorithm was applied to estimate the relative abundance of each TME-infiltrating cell for each sample. The gene expression matrix LM22 was downloaded from <http://cibersort.stanford.edu/>. The matrix contains 547 genes that could distinguish 22 human hematopoietic cell phenotypes, including seven T cell types, naive and memory B cells, plasma cells, NK cells, and myeloid subsets.

Statistical analysis

A t test and one-way ANOVA were used to assess the difference of components in two or multiple immune groups, respectively. For distinct m^5C modification patterns (m^5C cluster, gene cluster, and m^5C score), the two types of analysis methods were also utilized to identify potential differences in clinical characteristics, including sex, chemotherapy, neoadjuvant therapy, hormone therapy, radiotherapy, tumor progression, IDH molecular subtype, 1P/19Q codeletion and MGMT promoter methylation status. p or post hoc $p < 0.05$ was considered statistically significant. The immuno-component diagram and parallel bar graph for multiple pathways were visualized by the “ggplot2” package, while a histogram for various clinical characteristics was drawn by the function “ggbarstats” in the “ggstatsplot” package. Additionally, Pearson’s correlation coefficients were calculated to evaluate the relationship of the constructed m^5C score with a set of 36 immune checkpoint-related genes. Ten of them (*TNFRSF14*, *PLEKHG5*, *LGALS9*, *PDCD1LG2*, *NRP1*, *CD86*, *CD44*, *CD40*, and *CD200*) were matched in the TCGA-GBM dataset, and correlation analyses were conducted, which were further visualized by the “geom_tile” function in the “ggplot2” package. All data processing was conducted in the R programming language (v.4.0.1).

DATA AND CODE AVAILABILITY

The datasets analyzed during the current study are available in TCGA (<https://portal.gdc.cancer.gov/>) and CGGA (<http://www.cgga.org.cn/>; CGGA-1018, http://www.cgga.org.cn/download?file=download/20200506/CGGA.mRNAseq_693.RSEM-genes.20200506.txt.zip&type=mRNAseq_693&time=20200506, http://www.cgga.org.cn/download?file=download/20200506/CGGA.mRNAseq_325.RSEM-genes.20200506.txt.zip&type=mRNAseq_325&time=20200506; CGGA-301, http://www.cgga.org.cn/download?file=download/20200506/CGGA.mRNA_array_301_gene_level.20200506.txt.zip&type=mRNA_array_301_gene_level&time=20200506).

The data generated based on the public database are available from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omton.2024.200790>.

ACKNOWLEDGMENTS

The authors thank the TCGA and CGGA for the datasets available. This study was supported by the National Natural Science Foundation of China (grant 81873776 to X.G.) and the Guangdong Basic and Applied Basic Research Foundation (grants 2021A1515011681 and 2023A1515010495 to X.G.). We also thank the staff and technicians at Zhujiang Hospital for kind assistance and support.

AUTHOR CONTRIBUTIONS

X.G. and S.R. conceived and designed the study. Y.W., X.G., X.C., and R.L. performed the experiments and data curation. S.Z., K.W., W.L., H.H., and H.X. conducted the statistical analysis and provided general support for the study. X.G., Y.W., X.C., X.G., and S.R. were responsible for writing and revising the original draft. All authors have read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol.* *23*, 1231–1251.
- Lee, S.C. (2018). Diffuse Gliomas for Nonneuropathologists The New Integrated Molecular Diagnostics. *Arch. Pathol. Lab Med.* *142*, 804–814.
- Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., and Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* *114*, 97–109.
- Domingues, P., González-Tablas, M., Otero, Á., Pascual, D., Miranda, D., Ruiz, L., Sousa, P., Ciudad, J., Gonçalves, J.M., Lopes, M.C., et al. (2016). Tumor infiltrating immune cells in gliomas and meningiomas. *Brain Behav. Immun.* *53*, 1–15.
- Franceschi, E., Hofer, S., Brandes, A.A., Frappaz, D., Kortmann, R.D., Bromberg, J., Dangouloff-Ros, V., Boddaert, N., Hattingen, E., Wiestler, B., et al. (2019). EANO-EURACAN clinical practice guideline for diagnosis, treatment, and follow-up of post-pubertal and adult patients with medulloblastoma. *Lancet Oncol.* *20*, e715–e728.
- Rushing, E.J., and Wesseling, P. (2015). Towards an integrated morphological and molecular WHO diagnosis of central nervous system tumors: a paradigm shift. *Curr. Opin. Neurol.* *28*, 628–632.
- Dong, Z., and Cui, H. (2020). The Emerging Roles of RNA Modifications in Glioblastoma. *Cancers* *12*.
- Cusenza, V.Y., Tameni, A., Neri, A., and Frazzi, R. (2023). The lncRNA epigenetics: The significance of m6A and m5C lncRNA modifications in cancer. *Front. Oncol.* *13*, 1063636.
- Lopez-Bertoni, H., Johnson, A., Rui, Y., Lal, B., Sall, S., Malloy, M., Coulter, J.B., Lugo-Fagundo, M., Shudir, S., Khela, H., et al. (2022). Sox2 induces glioblastoma cell stemness and tumor propagation by repressing TET2 and deregulating 5hmC and 5mC DNA modifications. *Signal Transduct. Target. Ther.* *7*, 37.
- Mitchell, D.A., Fecci, P.E., and Sampson, J.H. (2003). Adoptive immunotherapy for malignant glioma. *Cancer J.* *9*, 157–166.
- Wang, G., and Wang, W. (2022). Advanced Cell Therapies for Glioblastoma. *Front. Immunol.* *13*, 904133.
- Romani, M., Pistillo, M.P., Carosio, R., Morabito, A., and Banelli, B. (2018). Immune Checkpoints and Innovative Therapies in Glioblastoma. *Front. Oncol.* *8*, 464.
- Wang, S., and van de Pavert, S.A. (2022). Innate Lymphoid Cells in the Central Nervous System. *Front. Immunol.* *13*, 837250.
- Gieryng, A., Pszczolkowska, D., Walentynowicz, K.A., Rajan, W.D., and Kaminska, B. (2017). Immune microenvironment of gliomas. *Lab. Invest.* *97*, 498–518.
- Wei, J., Chen, P., Gupta, P., Ott, M., Zamlar, D., Kassab, C., Bhat, K.P., Curran, M.A., de Groot, J.F., and Heimberger, A.B. (2020). Immune biology of glioma-associated macrophages and microglia: functional and therapeutic implications. *Neuro Oncol.* *22*, 180–194.
- Wang, M., Zhou, Z., Wang, X., Zhang, C., and Jiang, X. (2022). Natural killer cell awakening: unleash cancer-immunity cycle against glioblastoma. *Cell Death Dis.* *13*, 588.
- Dhodapkar, K.M., Banerjee, D., and Steinman, R.M. (2005). Harnessing the immune system against human glioma. *Ann. N. Y. Acad. Sci.* *1062*, 13–21.
- Stathopoulos, A., Samuelson, C., Milbouw, G., Hermanne, J.P., Schijns, V.E.J.C., and Chen, T.C. (2008). Therapeutic vaccination against malignant gliomas based on all-ecognition and syngeneic tumor antigens: proof of principle in two strains of rat. *Vaccine* *26*, 1764–1772.
- Tong, N., He, Z., Ma, Y., Wang, Z., Huang, Z., Cao, H., Xu, L., Zou, Y., Wang, W., Yi, C., et al. (2021). Tumor Associated Macrophages, as the Dominant Immune Cells, Are an Indispensable Target for Immunologically Cold Tumor-Glioma Therapy? *Front. Cell Dev. Biol.* *9*, 706286.
- Appin, C.L., and Brat, D.J. (2015). Biomarker-driven diagnosis of diffuse gliomas. *Mol. Aspects Med.* *45*, 87–96.
- Hvinden, I.C., Cadoux-Hudson, T., Schofield, C.J., and McCullagh, J.S.O. (2021). Metabolic adaptations in cancers expressing isocitrate dehydrogenase mutations. *Cell Rep. Med.* *2*, 100469.
- Zhao, J., Ma, W., and Zhao, H. (2014). Loss of heterozygosity 1p/19q and survival in glioma: a meta-analysis. *Neuro Oncol.* *16*, 103–112.
- Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G.W., Aldape, K.D., Yung, W.K.A., Salama, S.R., Cooper, L.A.D., Rheinbay, E., Miller, C.R., Vitucci, M., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* *372*, 2481–2498.
- Jenkins, R.B., Blair, H., Ballman, K.V., Giannini, C., Arusell, R.M., Law, M., Flynn, H., Passe, S., Felten, S., Brown, P.D., et al. (2006). A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. *Cancer Res.* *66*, 9852–9861.
- Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* *44*, e71.
- Han, B., Meng, X., Wu, P., Li, Z., Li, S., Zhang, Y., Zha, C., Ye, Q., Jiang, C., Cai, J., and Jiang, T. (2020). ATRX/EZH2 complex epigenetically regulates FADD/PARP1 axis, contributing to TMZ resistance in glioma. *Theranostics* *10*, 3351–3365.
- Gupta, M.K., Polisetty, R.V., Sharma, R., Ganesh, R.A., Gowda, H., Purohit, A.K., Ankathi, P., Prasad, K., Mariswamappa, K., Lakshminantha, A., et al. (2019). Altered transcriptional regulatory proteins in glioblastoma and YBX1 as a potential regulator of tumor invasion. *Sci. Rep.* *9*, 10986.
- Jovčevska, I., Zupanec, N., Urlep, Ž., Vranič, A., Matos, B., Stokin, C.L., Muylldermans, S., Myers, M.P., Buzdin, A.A., Petrov, I., and Komel, R. (2017). Differentially expressed proteins in glioblastoma multiforme identified with a nanobody-based anti-proteome approach and confirmed by OncoFinder as possible tumor-class predictive biomarker candidates. *Oncotarget* *8*, 44141–44158.
- Nakano, S., Suzuki, T., Kawarada, L., Iwata, H., Asano, K., and Suzuki, T. (2016). NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA(Met). *Nat. Chem. Biol.* *12*, 546–551.
- Zhang, Y., Jiang, X., Wu, Z., Hu, D., Jia, J., Guo, J., Tang, T., Yao, J., Liu, H., and Tang, H. (2020). Long Noncoding RNA LINC00467 Promotes Glioma Progression through Inhibiting P53 Expression via Binding to DNMT1. *J. Cancer* *11*, 2935–2944.

31. Zhou, D., Wan, Y., Xie, D., Wang, Y., Wei, J., Yan, Q., Lu, P., Mo, L., Xie, J., Yang, S., and Qi, X. (2015). DNMT1 mediates chemosensitivity by reducing methylation of miRNA-20a promoter in glioma cells. *Exp. Mol. Med.* *47*, e182.
32. Fomchenko, E.I., Erson-Omay, E.Z., Zhao, A., Bindra, R.S., Huttner, A., Fulbright, R.K., and Moliterno, J. (2019). DNMT3A co-mutation in an IDH1-mutant glioblastoma. *Cold Spring Harb. Mol. Case Stud.* *5*, a004119.
33. Zhou, J., Vincent, K., Findlay, S., Choi, D., Godbout, R., Postovit, L.M., and Fu, Y. (2018). 61 Functional characterization of ribosomal RNA methyltransferase NSUN5 in glioblastoma. *Can. J. Neurol. Sci.* *45*, S10–S11.
34. Narsia, N., Ramagiri, P., Ehrmann, J., and Kolar, Z. (2017). Transcriptome analysis reveals distinct gene expression profiles in astrocytoma grades II-IV. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech. Repub.* *161*, 261–271.
35. Gu, X., Gong, H., Shen, L., and Gu, Q. (2018). MicroRNA-129-5p inhibits human glioma cell proliferation and induces cell cycle arrest by directly targeting DNMT3A. *Am. J. Transl. Res.* *10*, 2834–2847.
36. Janin, M., Ortiz-Barahona, V., de Moura, M.C., Martínez-Cardús, A., Llinàs-Arias, P., Soler, M., Nachmani, D., Pelletier, J., Schumann, U., Calleja-Cervantes, M.E., et al. (2019). Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. *Acta Neuropathol.* *138*, 1053–1074.
37. Kaufman, L., and Rousseeuw, P.J. (1990). *Partitioning Around Medoids (Program PAM)* (Hoboken, New Jersey: John Wiley & Sons, Inc.), pp. 68–125.
38. Smyth, G.K. (2005). *Limma: Linear Models for Microarray Data* (Springer), pp. 397–420.
39. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B* *57*, 289–300.
40. Hu, K. (2021). Become Competent in Generating RNA-Seq Heat Maps in One Day for Novices Without Prior R Experience. *Methods Mol. Biol.* *2239*, 269–303.
41. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* *28*, 882–883.
42. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* *14*, 7.
43. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* *4*, 2612.