

Comprehensive machine learning-based integration develops a novel prognostic model for glioblastoma

Qian Jiang,^{1,4} Xiawei Yang,^{2,4} Teng Deng,¹ Jun Yan,¹ Fangzhou Guo,¹ Ligen Mo,¹ Sanqi An,³ and Qianrong Huang¹

¹Department of Neurosurgery, Guangxi Medical University Cancer Hospital, Nanning, Guangxi, China; ²Transplant Medical Center, The Second Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China; ³Biosafety Level-3 Laboratory, Life Sciences Institute & Guangxi Collaborative Innovation Center for Biomedicine, Guangxi Medical University, Nanning, Guangxi, China

In this study, we developed a new prognostic model for glioblastoma (GBM) based on an integrated machine learning algorithm. We used univariate Cox regression analysis to identify prognostic genes by combining six GBM cohorts. Based on the prognostic genes, 10 machine learning algorithms were integrated into 117 algorithm combinations, and the artificial intelligence prognostic signature (AIPS) with the greatest average C-index was chosen. The AIPS was compared with 10 previously published models by univariate Cox analysis and the C-index. We compared the differences in prognosis, tumor immune microenvironment (TIME), and immunotherapy sensitivity between the high and low AIPS score groups. The AIPS based on the random survival forest algorithm with the highest average C-index (0.868) was selected. Compared with the previous 10 prognostic models, our AIPS has the highest C-index. The AIPS was closely linked to the clinical features of GBM. We discovered that patients in the low score group had improved prognoses, a more active TIME, and were more sensitive to immunotherapy. Finally, we verified the expression of several key genes by western blotting and immunohistochemistry. We identified an ideal prognostic signature for GBM, which might provide new insights into stratified treatment approaches for GBM patients.

INTRODUCTION

Glioblastoma (GBM) is the most prevalent and deadly primary intracranial malignancy and is characterized by significant invasion and heterogeneity.¹ The incidence of GBM is 5–8 per 100,000 people, and the median survival time from diagnosis is 15 months.^{2,3} The treatment of GBM consists of surgical resection combined with radiotherapy and temozolomide (TMZ) chemotherapy, which was developed 18 years ago.⁴ Immunotherapy has recently become the standard treatment for a growing number of solid tumors.⁵ However, the immune landscape of GBM is complicated, and GBM patients are likely to have an immunosuppressive tumor immune microenvironment (TIME), which explains why programmed cell death protein 1 (PD-1)/protein cell death ligand 1 (PD-L1) immunotherapy trials involving GBM patients were unsuccessful.⁶ In addition, the failure of targeted therapy in GBM suggests that GBM might not rely solely

on targeted therapy driven by a single pathway.⁷ Thus, it is necessary to investigate novel personalized management and multitarget therapy techniques.

Presently, the World Health Organization's (WHO's) categorization of tumors of the central nervous system is still used in the decision-making and care of GBM patients.⁸ Due to the heterogeneity of GBM and its very complicated TIME, the treatment response and prognosis of GBM patients are fairly diverse. With the development of bioinformatics technology and evidence-based medicine, many predictive gene signatures have been discovered.^{9–11} However, owing to improper machine learning techniques and a lack of thorough validation across many cohorts, the clinical use of multigenic prognostic signatures is often difficult.^{12,13} Furthermore, this information may not be fully useful since these multigenic prognostic signatures are based on the expression of genes involved in specific pathways, including those related to metabolism, hypoxia, and programmed cell death.^{14,15}

To develop a perfect signature, we constructed and validated a 79-gene artificial intelligence prognostic signature (AIPS) by combining 117 machine learning algorithms. In six multicenter cohorts and the combined cohort, the AIPS was a reliable prognostic indicator of overall survival (OS). Our AIPS with the highest C-index showed excellent and comprehensive predictive performance when compared with 10 previously published GBM/glioma models in all cohorts. Furthermore, patients in the low AIPS score group had improved prognoses, a more active TIME, and were more sensitive

Received 24 October 2023; accepted 14 June 2024;
<https://doi.org/10.1016/j.omton.2024.200838>.

⁴These authors contributed equally

Correspondence: Sanqi An, Biosafety Level-3 Laboratory, Life Sciences Institute & Guangxi Collaborative Innovation Center for Biomedicine, Guangxi Medical University, No. 22 Shuangyong Road, Qingxiu District, Nanning, Guangxi 530021, China.

E-mail: ansq@mail2.sysu.edu.cn

Correspondence: Qianrong Huang, Department of Neurosurgery, Guangxi Medical University Cancer Hospital, No. 71 Hedi Road, Qingxiu District, Nanning, Guangxi 530021, China.

E-mail: huangqianrong@gxmu.edu.cn



to immunotherapy. Finally, we verified the expression of several key genes in GBM by immunohistochemistry (IHC) and western blotting (WB).

RESULTS

Identification of OS-related genes in GBM and functional enrichment analysis

Principal-component analysis (PCA) was used to evaluate the removal of batch effects. Before the merger, the six datasets were discrete, with almost no intersection (Figure 1A). After the six datasets were merged, the differences between the datasets were reduced (Figure 1B). After all the data were fused, we obtained data for a total of 1,036 GBM patients, of which 984 had complete survival information. A total of 79 eligible OS-related genes were identified in the six cohorts by univariate Cox regression analysis. We used forest maps to visualize the OS-related genes in each cohort (Figure 1C). Subsequently, functional enrichment analysis was performed based on protective prognostic genes ($n = 31$) and prognostic risk genes ($n = 48$). Gene ontology (GO) analysis suggested that protective genes were mainly enriched in small molecule catabolic process, small molecule catabolic process, C21-steroid hormone metabolic process, cell projection membrane, focal adhesion, cell-substrate junction, testosterone dehydrogenase [NAD(P)] activity, steroid dehydrogenase activity, monocarboxylic acid binding, etc. (Figures S1A–S1C). According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, these protective genes were enriched in carbon metabolism, the sphingolipid signaling pathway, glioma, epidermal growth factor receptor tyrosine kinase inhibitor resistance, and choline metabolism in cancer pathways (Figure S1D). The results of the GO and KEGG analyses of the prognostic risk genes are shown in (Figures S1E–S1H).

Establishment and assessment of the AIPS for GBM

We next incorporated 79 OS-related genes into a machine learning program to construct an AIPS. In the combined training cohort, we built prediction models using 117 algorithms and calculated the C-index for all the cohorts. Finally, the AIPS based on the random survival forest (RSF) algorithm with the highest average C-index (0.868) was selected (Figure 2A). We calculated the AIPS risk score for each patient in all cohorts based on the genes included in the RSF model and showed the variable importance of the AIPS genes (Figure 2B). In the combined training cohort, the area under the curve (AUC) values for the AIPS predicting patient 1-, 3-, and 5-year survival were very high, all greater than 0.970 (Figure 2C). In addition, we found that the AUC value of the AIPS was greater than 0.950 at different time points (Figure 2D). Kaplan-Meier (K-M) curve analysis revealed that the prognosis of patients in the high AIPS score group was significantly worse than that of patients in the low AIPS score group (Figure 2E). Surprisingly, we observed similar results across the six validation cohorts. In all the validation cohorts, the AUC values for the AIPS were greater than 0.900 at any time point, and the high AIPS score group exhibited a worse prognosis (Figures 3A–3F).

Comparison with published signatures in GBM

With the rapid development of bioinformatics, researchers have widely reported prognostic features based on small molecule expression. We compared our AIPS with 10 published prognostic models in terms of the C-index and prognostic value to further evaluate the predictive power of the AIPS. These previously published gene prognostic signatures encompassed several biological functions, including chemotherapy sensitivity, pyroptosis, cuproptosis, and immunity. Interestingly, we found that the AIPS had a C-index of approximately 0.9 in all seven cohorts, which was much greater than that of the other 10 gene cohorts (Figures 4A–4G). We then included our AIPS, other published signatures, and clinical features in the univariate Cox regression analysis to compare their prognostic value. Notably, while four published signatures (Zhao S, Wang W, Xiao S, and Yang J) were significantly associated with GBM outcomes in most cohorts (six of seven),^{16–19} only our AIPS achieved statistical significance ($p < 0.001$) in both the validation and training cohorts (Figures 5A–5G). Overall, our model showed strong and robust predictive performance for GBM. We also analyzed the correlation of the AIPS with the clinical features of the GBM patients in the training cohort. We found that 1p19q non-codeletion and IDH wild-type patients, which both show poor prognosis, had higher AIPS scores, and a greater proportion of patients with these two subtypes were in the high AIPS score group (Figures 5H and 5I). Patients who died had higher AIPS scores than those who did not, and the proportion of cases who died was greater in the high AIPS score group (Figure 5J). These data suggested that our AIPS correlated with the clinical features of GBM.

Functional enrichment analysis based on the AIPS

Considering the stable prognostic value of our AIPS, we further explored the biological functions associated with it. We analyzed the correlation between the AIPS and all genes in the training cohort and selected the top 50 genes with significant positive (Figure S2A) and negative (Figure S2B) correlations. Next, GO and KEGG analyses were performed based on these 100 significantly related genes. The GO results showed that these genes were mainly related to the functions of extracellular activities and components, and they were also involved in regulating mechanisms such as blood vessel formation and hypoxia (Figures S2C and S2E). According to the KEGG analysis, these genes were significantly enriched in focal adhesion and extracellular matrix (ECM)-receptor interaction, which are associated with cell migration, and many cancer-related pathways, such as the PI3K-Akt, HIF-1, TNF, and p53 signaling pathways, were also identified (Figure S2F). The Reactome pathway analysis by gene set enrichment analysis (GSEA) showed significant enrichment for collagen formation, ECM organization, and cell cycle checkpoints in the high AIPS score group (Figure S3A). In addition, we conducted a comprehensive gene set variation analysis (GSVA), including HALLMARK, KEGG, Reactome, Biocarta, and WikiPathways analyses, to explore the pathways that differed between the high and low AIPS score groups (Figures S3B–S3F). Overall, the high AIPS score patients showed more active cancer-related features, especially aggressive GBM features.

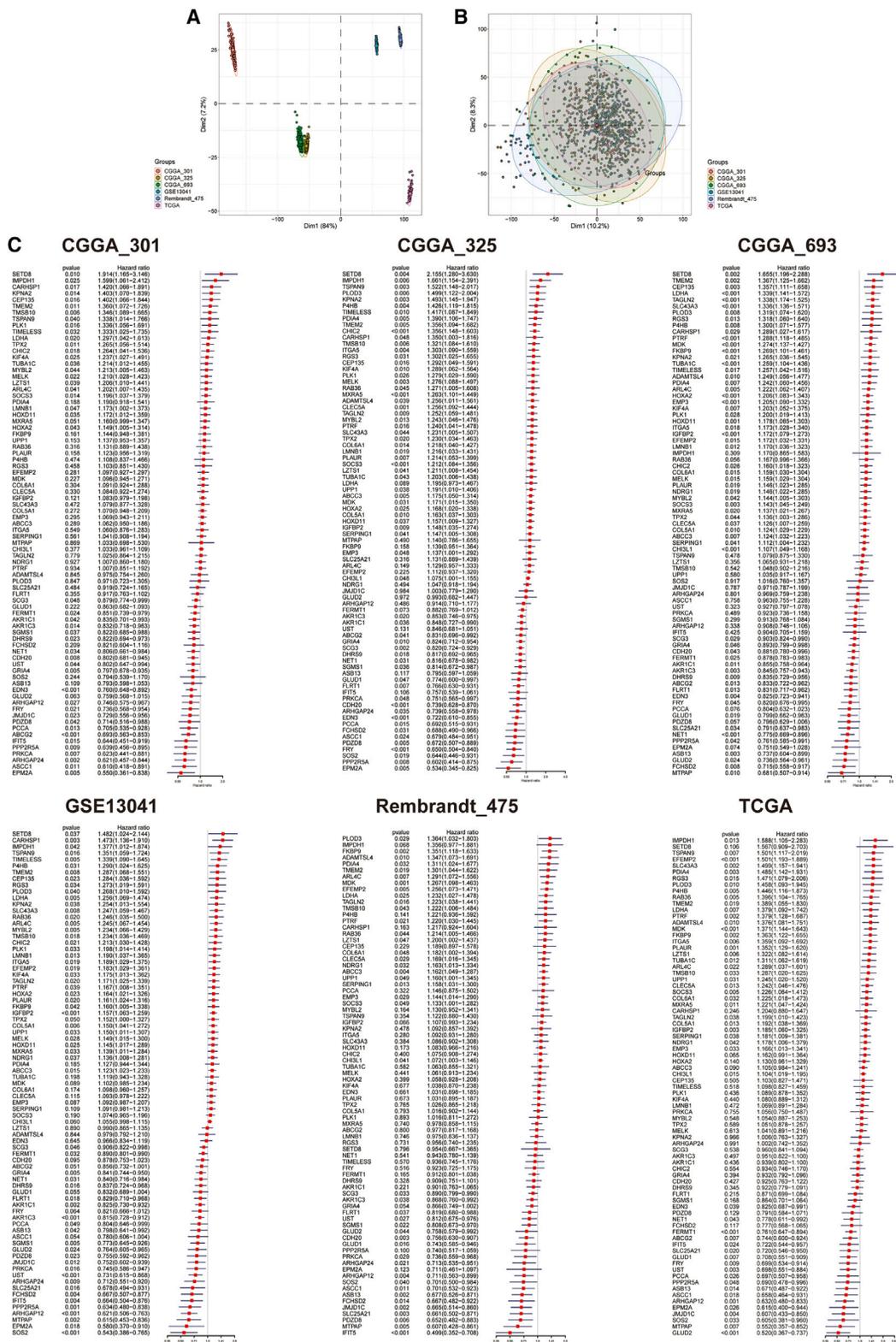
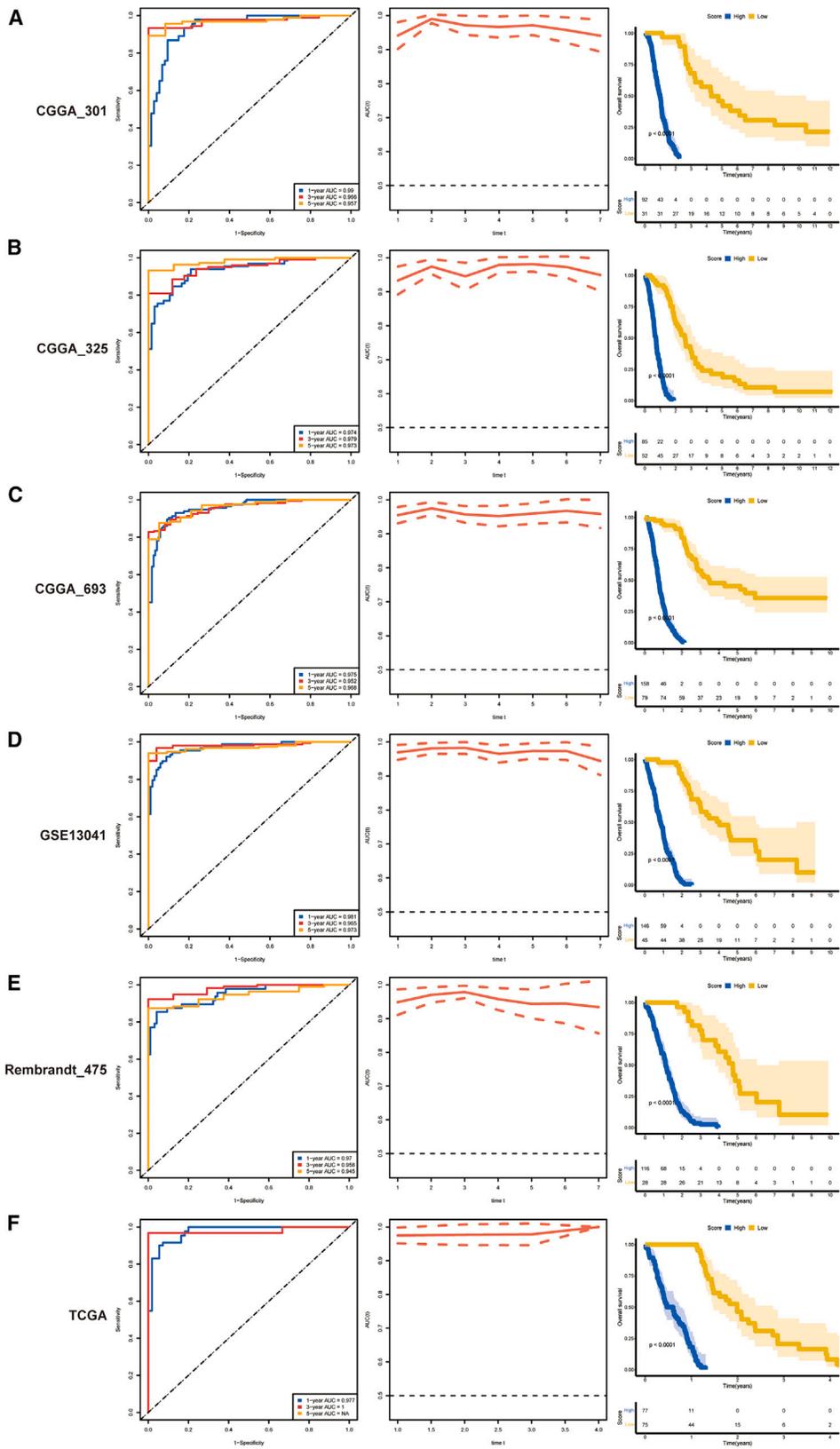


Figure 1. Merging of datasets and identification of prognostic genes

(A) PCA before batch effect removal. (B) PCA after batch effect removal and merging. (C) Results of univariate Cox analysis of 79 prognostic genes in each dataset.



(legend on next page)

Association of the AIPS with the TIME and genome alterations

Next, we systematically evaluated the correlation between the AIPS and the TIME using eight algorithms. Several immune cells, such as macrophages, fibroblasts, and cancer-associated fibroblasts (CAFs), were highly infiltrated in the high AIPS score group (Figure S4A). Correlation analysis revealed that these immune cells were significantly positively correlated with the AIPS (Figure S4B). Using the estimate algorithm, we found that compared with those in the low AIPS score group, the immune score, stroma score, and ESTIMATE score were greater in the high AIPS score group, but the tumor purity was lower (Figure S4A). In addition, the expression levels of chemokines and receptors, such as interleukin (IL)6, CXCR4, IFNGR2, IL10RB, VEGFA, and IL4R, were generally greater in the high AIPS score group (Figure S5A), and these chemokines were significantly positively correlated with the AIPS score (Figure S5B). We analyzed the somatic mutation data and identified the top 10 genes with mutation rates in both subgroups (Figures 6A and 6B). By comparison, we found that the mutation frequencies of five genes (IDH1, ADAMTS12, CYP4B1, PCDH18, and SNTG2) were significantly greater in the low AIPS score group than in the high AIPS score group (Figure 6C). Currently, IDH1 mutation is recognized as an important molecular marker for the prognosis of GBM²⁰; thus, the remaining four gene mutations may serve as potential prognostic markers for GBM.

Immunotherapy prediction and drug sensitivity

A previous study suggested that patients with TP53 mutations are more susceptible to being treated with immune checkpoint inhibitors (ICIs).²¹ We found that the low AIPS score group (45%) seemed to have more TP53 mutations than did the high AIPS score group (26%). Thus, we further explored the ability of the AIPS to predict immunotherapy response. The tumor immune dysfunction and exclusion (TIDE) algorithm is a widely used and reliable method for predicting immunotherapy response (Figure 6D).²² In the training cohort, we found that the low AIPS score patients had lower TIDE scores and a higher proportion of treatment response than the high AIPS score patients (Figures 6E and 6F). These data suggested that patients with high AIPS scores based on our AIPS had greater immunosuppression and immune escape and that patients with low AIPS scores might be potential candidates for immunotherapy.²² Moreover, we calculated the half maximal inhibitory concentration (IC50) of chemotherapy drugs to investigate the relationship between the AIPS and chemotherapy resistance. We identified several potentially sensitive chemotherapeutic agents for low (Figure 6G) and high AIPS score (Figure 6H) GBM patients.

Expression patterns and single-cell analysis of the top 12 genes of variable importance

Given the robust predictive power of the AIPS, we further analyzed the top 12 genes in the RSF model, which may play a critical role in GBM. Using the GEPIA website, we found that the mRNA expres-

sion levels of CHI3L1, FKBP9, HOXD11, IMPDH1, MDK, MYBL2, SETD8, TMEM2, and UST were significantly upregulated in GBM, while the expression levels of EPM2A and FLRT1 were significantly decreased in GBM (Figure 7A). EDN3 expression in GBM tissues was lower than that in normal tissues, but the difference did not reach statistical significance (Figure 7A). To the best of our knowledge, five (CHI3L1, IMPDH1, MDK, MYBL2, and SETD8) of the top 12 genes have received less research attention in GBM. In this study, IHC and WB revealed that all five genes were upregulated in GBM tissues compared with adjacent tissues, which is consistent with the findings of our bioinformatics research (Figures 7B and 7C). We further explored the percentage and distribution of these 12 genes in GBM through the GSCA website. The results showed that IMPDH1, FKBP9, EDN3, MYBL2, EPM2A, and UST had high copy number variation (CNV) proportions in GBM (Figure 7D). Among them, IMPDH1, FKBP9, EDN3, and MYBL2 mainly exhibited heterozygous amplification, while EPM2A and UST exhibited heterozygous deletion (Figure 7D). Correlation analysis revealed that the expression of IMPDH1, FKBP9, EPM2A, and FLRT1 was significantly positively correlated with the proportion of CNV (Figure 7E), and the methylation of CHI3L1, FKBP9, IMPDH1, MDK, and FLRT1 significantly inhibited their expression in GBM (Figure 7F). In addition, we investigated the expression of 12 key genes in immune cells using glioma/GBM single-cell data. As shown in (Figure S6), these genes were expressed to varying degrees in immune cells.

DISCUSSION

GBM is the most aggressive and heterogeneous primary central nervous system tumor.^{23,24} Although tumor-treating field therapy has enriched the comprehensive treatment of GBM in recent years, the median survival time has only been extended to 19.4 months in a phase III trial.²⁵ Thus, we created a 79-gene AIPS with 117 machine learning algorithm integrations in the combined training cohort to determine useful GBM biomarkers. Through the use of time, receiver operating characteristic (ROC), and survival analyses, the predictive power of the AIPS was confirmed across several independent cohorts. In addition, the AIPS was also compared with 10 previously developed prognostic signatures for GBM/glioma based on the C-index. The results showed that the AIPS exhibited the best predictive performance compared with the other prognostic models in all independent cohorts. Finally, based on the AIPS, we also distinguished between two different groupings. A greater tumor mutation burden, a more favorable prognosis, a significantly active TIME, and a stronger immunotherapy response were detected in the low AIPS group. Therefore, in clinical practice, our AIPS may be a reliable biomarker for stratified treatment of GBM.

Traditional WHO tumor categorization no longer accurately reflects the demands of neurosurgeons for optimum GBM indicators due to

Figure 3. The predictive power of the AIPS was evaluated in validation cohorts

ROC curves for 1, 3, and 5 years, AUC values at different times, and survival curves in the (A) CGGA_301, (B) CGGA_325, (C) CGGA_693, (D) GSE13041, (E) Rembrandt_475, and (F) TCGA validation cohorts.

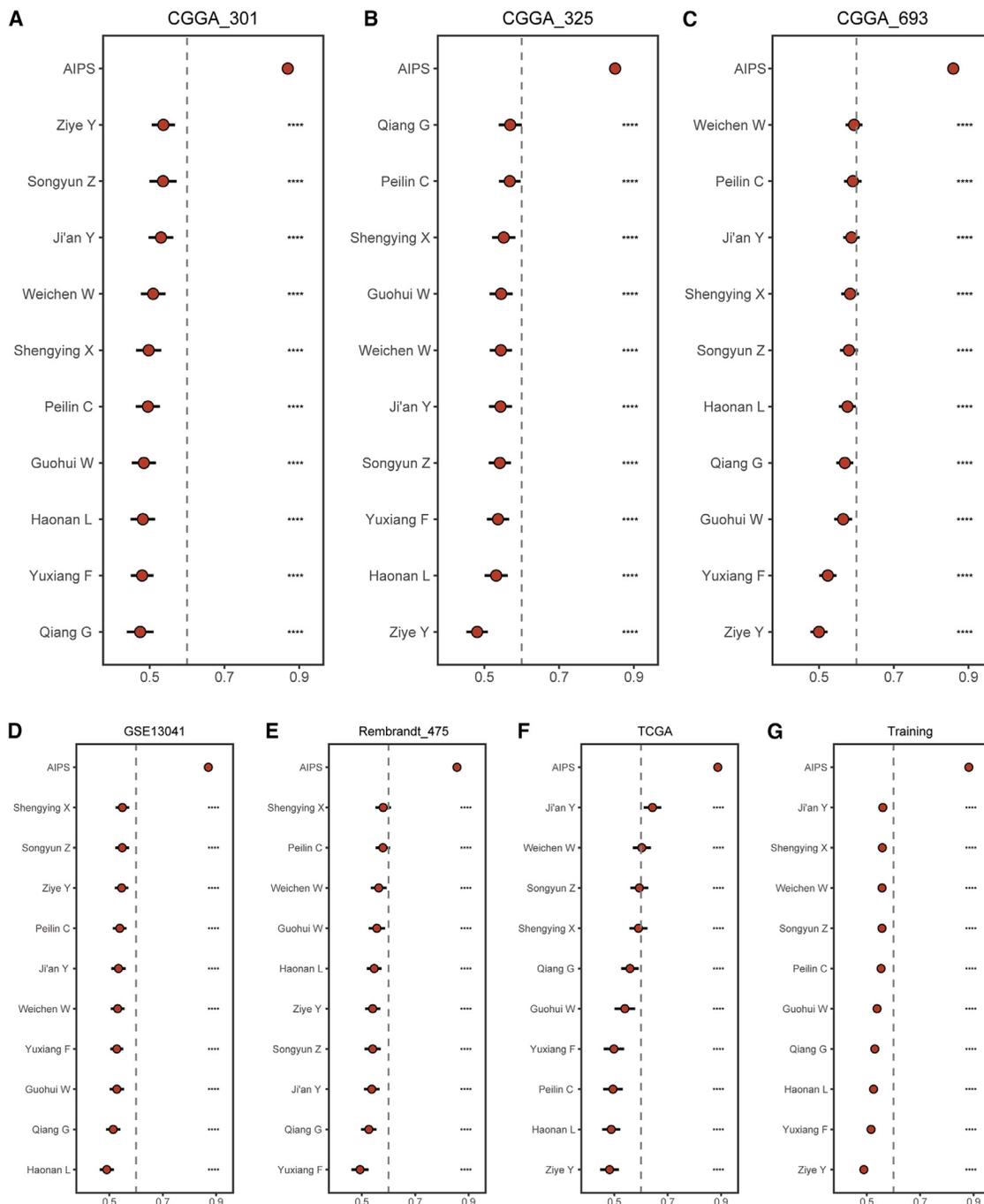


Figure 4. Comparison of our AIPS with the 10 previous models in the C-index

The C-index of our model was higher than that of previous models in (A) CGGA_301, (B) CGGA_325, (C) CGGA_693, (D) GSE13041, (E) Rembrandt_475, (F) TCGA validation, and (G) training cohorts, **** $p < 0.0001$.

advancements in medical therapy. Numerous prognostic signatures for GBM have been identified in recent years based on metabolism, hypoxia, and programmed cell death; however, these signatures are all dependent on certain pathways and have limitations.^{14,15} To fully represent the importance of diverse pathways in the onset and pro-

gression of GBM, we incorporated the intersecting genes of six distinct GBM cohorts—which were not based on genes associated with particular pathways—in our investigation. Additionally, in earlier studies, researchers preferred to build signatures based on a single algorithm.^{17–19,26} In this study, for the purpose of building a

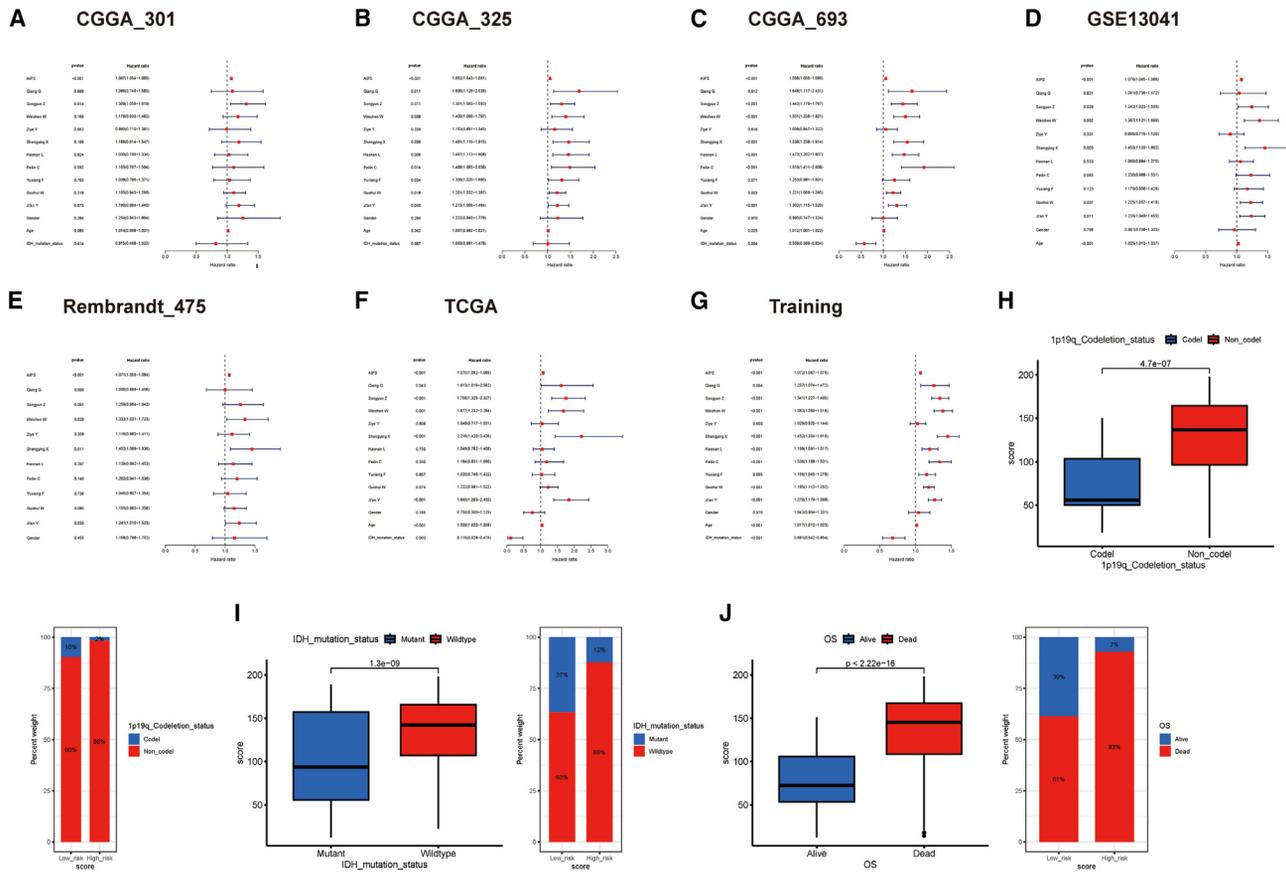


Figure 5. Univariate Cox regression analysis and clinical correlation analysis

(A–G) Prognostic role of AIPS, previously published models, and clinical features in validation and training cohorts. (H–J) Relationship between AIPS and 1p19q status, IDH status, and survival status of GBM patients in the training cohort.

signature, we merged 10 common machine learning arithmetic permutations into 117 algorithm combinations, which might compensate for the deficiencies of a single method. Among the 117 algorithm combinations, RSF was found to be the best signature with the greatest C-index (0.868).

Multiple enrichment analysis was used to clarify the possible biological mechanism and significant signaling pathways implicated in the AIPS. The enrichment analysis of protective and risk genes in our AIPS revealed the existence of a wide range of biological processes and signaling pathways, including those related to metabolic reprogramming, programmed cell death, and drug research information, in contrast to the model genes identified in previous studies, which were based on specific biological pathways. These findings suggest that the AIPS can accurately capture the biological data involved in GBM development. In addition, we noted that the AIPS-related genes were mainly enriched in extracellular structures, ECM tissues, focal adhesion, and ECM-receptor interactions, which are biological processes related to cancer migration and invasion^{27,28} This finding suggested that tumors with higher AIPS scores had a greater proportion of malignant components. According to the pathway analysis, the

high AIPS group was mostly enriched in immune-related, invasion, and migration pathways, which helps to explain why its immunological microenvironment is more suppressive and its biological activity is more malignant. The low AIPS group exhibited enrichment of metabolism-related pathways, suggesting that the AIPS is involved in the metabolism of GBM substances and may influence the formation and incidence of GBM through the regulation of AIPS-related metabolic enzymes. These findings, to some extent, reveal the reason for the significant difference in the prognosis of GBM patients in the two subgroups.

Despite recent advances in the treatment of solid tumors with immunotherapy, no phase III clinical studies of immunotherapy have been successful in improving patient outcomes for GBM owing to the existence of a highly immune-suppressive TIME.^{29,30} The TIME has a major influence on the development, progression, and treatment resistance of tumors, according to earlier research.³¹ In our investigation, immunosuppression was present in the high AIPS group, but relative immune activation was observed in the low AIPS group. In addition, we discovered that there were substantial differences in tumor-infiltrating immune cells between the high AIPS score group and

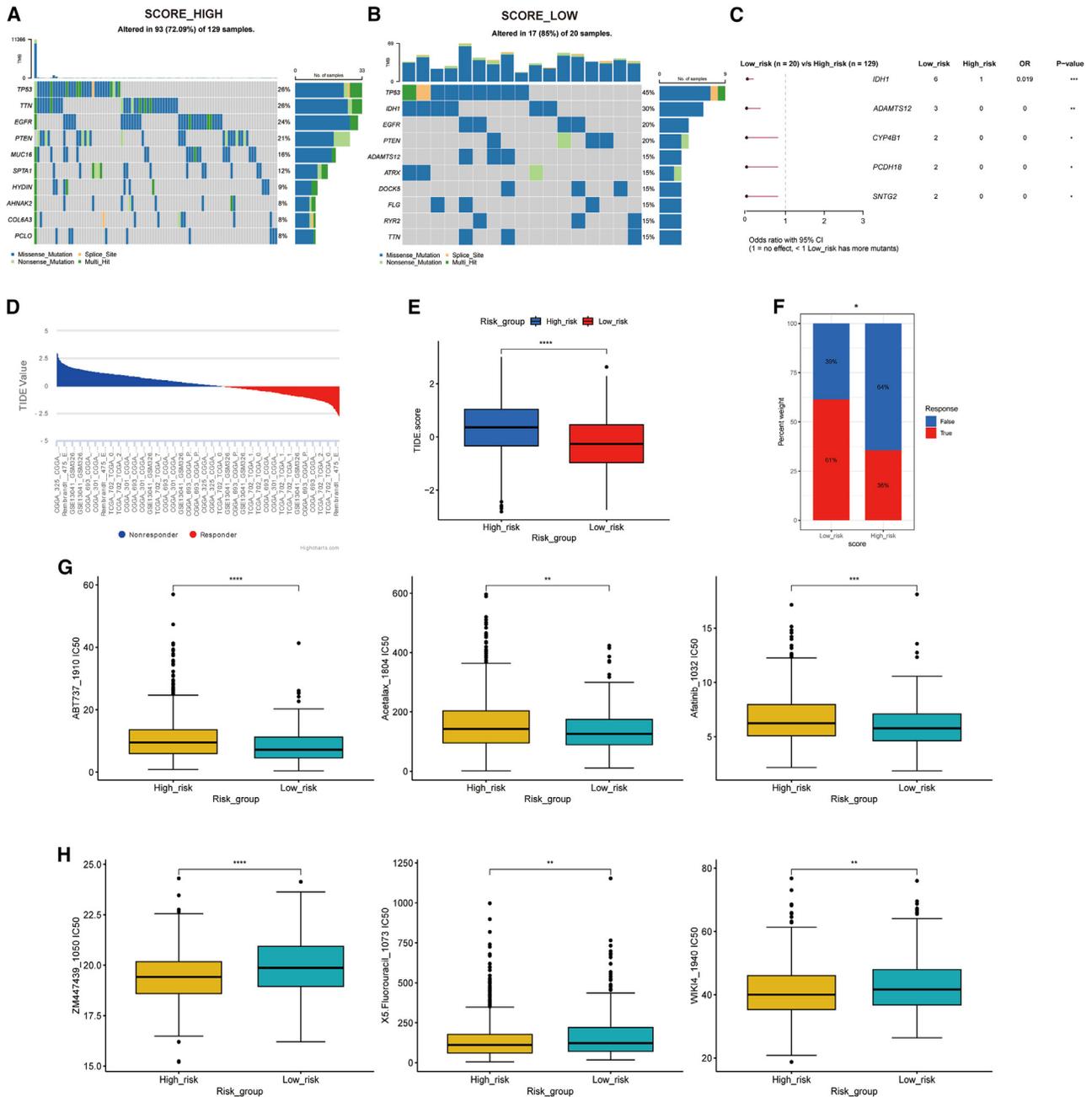
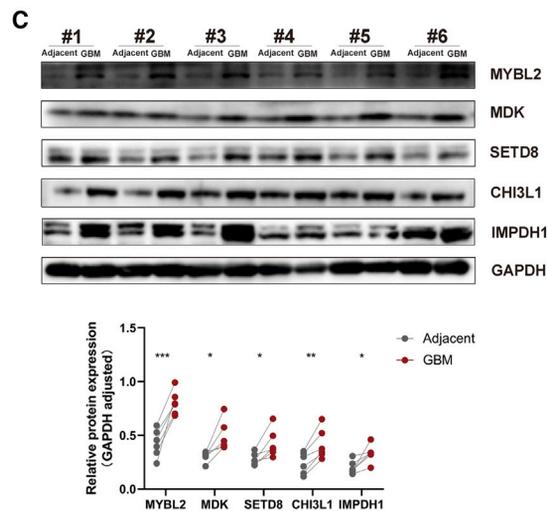
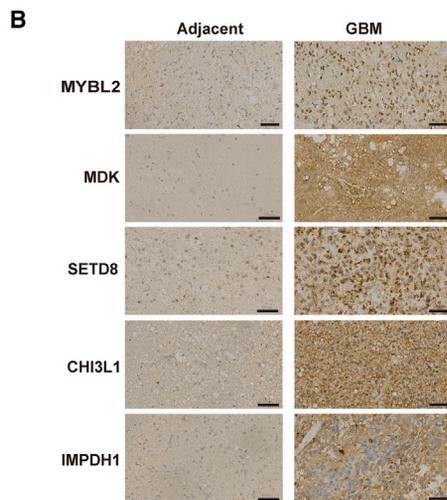
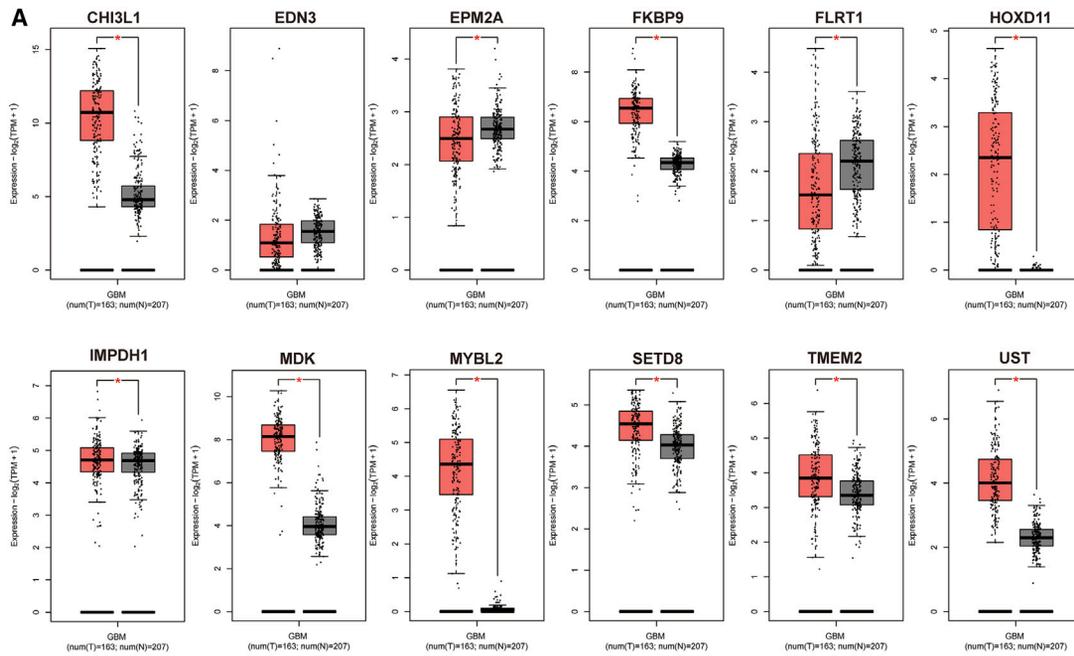


Figure 6. Genomic alterations, immunotherapy prediction, and drug sensitivity analyses

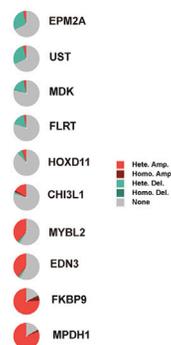
(A and B) Waterfall plots showed the top 10 genes with mutation rates in high- and low-AIPS score subgroups. (C) Genes with different mutation rates between two groups. (D) Identification of TIDE scores for GBM patients in the training cohort. (E) TIDE scores were compared between the two subgroups. (F) The immunotherapy response rates of the two subgroups were compared. Screening of sensitive drugs for (G) low-AIPS score and (H) high-AIPS score groups by IC50 analysis, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

low AIPS score group. These results implied that the genes involved in the construction of the AIPS were important for the development of GBM and the makeup of the TIME. CAFs secrete immunomodulatory chemicals, physically interact with immune cells, and modify the ECM to promote the growth of cancer cells and immune escape.³²

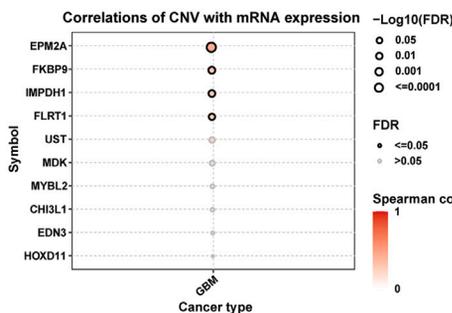
Additionally, CAFs, the most crucial matrix elements, have the ability to dramatically promote the aggressiveness and development of cancer.³³ In our study, increased CAF infiltration was a manifestation of poor prognosis in the high AIPS score group and might indicate that CAFs promote the growth of GBM. Tumor-associated macrophages



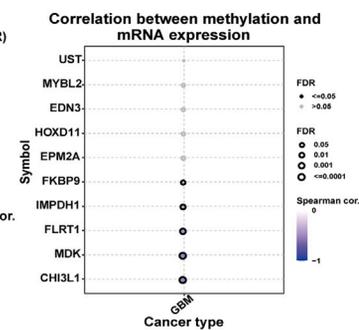
D CNV percentage in GBM



E



F



(legend on next page)

(TAMs) are crucial for the genesis, invasion, metastasis, and immunosuppressive microenvironment of tumors.³⁴ In addition, substantial TAM infiltration is believed to decrease susceptibility to radiation, chemotherapy, and targeted treatments and is often linked to a poor prognosis in a number of tumor types.³⁵ Furthermore, Tang et al. discovered that Smad3 might encourage TAMs to differentiate into CAFs with protumor effects.³³ In our study, TAMs demonstrated greater infiltration and poorer prognosis in the high AIPS score group and reduced infiltration and improved prognosis in the low AIPS score group, which was consistent with previous research. According to earlier research, B lymphocytes are the main immune cells that participate in the adaptive immune response against tumors.³⁶ Wataru Kida et al. suggested that the use of PD-1/PD-L1 blockade in hypopharyngeal cancer can increase the activation of B lymphocytes.³⁷ Moreover, loss of B cells in melanoma was linked to a poor response to ICIs.³⁸ Additionally, B lymphocytes are linked to a favorable prognosis for a number of cancers, such as ovarian, breast, and colorectal cancer.^{39–41} These results imply that B lymphocytes may be a potent tool for antitumor treatment and may provide a novel direction for immunotherapy in this regard. In our study, we observed a significant increase in B cells in the low AIPS score group. Therefore, the use of ICIs in the low AIPS score group could produce a stronger immune response. According to a growing number of studies, patients with gliomas who have CD4+ or CD8+ T cells have a better prognosis than those with gliomas with CD4– or CD8– T cells.^{42,43} Fan et al. noted that glioma development may be inhibited by promoting T cell proliferation and memory T cell responses.⁴⁴ In our study, we observed that CD4+ T cell infiltration was not significantly different between the high and low AIPS score groups, while CD8+ T cell infiltration was significantly greater in the low AIPS score group than in the high AIPS score group, which partly indicates that the low AIPS score group may be more sensitive to immunotherapy. One of the most promising immunotherapies for GBM is PD-1 checkpoint inhibition, yet the outcomes of clinical studies have revealed that patient efficacy varies substantially.⁴⁵ Determining which GBM patients are likely to respond to immunotherapy is a crucial challenge, according to the findings of many prior clinical trials.⁴⁶ In this study, the TIDE algorithm was used to determine that the GBM patients in the low AIPS score group had a better response to immunotherapy, while the GBM patients in the high AIPS score group displayed immunosuppression and poor immunotherapy efficacy. These data further suggest that our AIPS is helpful in identifying potential immunotherapy-sensitive GBM patients.

Modern medicine requires clinicians to provide accurate and individualized treatment for patients. As a result, we developed specific medications for the GBM patients in the high AIPS score group using the “OncoPredict” R package. According to our research, ZM447439-1050, X5.Fluorouracil-1073, and WIKI4-1940 all demonstrated

considerably greater sensitivity in the high AIPS score group compared with the low AIPS score group. The sensitivity of ZM447439-1050, X5.Fluorouracil-1073, and WIKI4-1940 in the GBM patients in the high AIPS score group should be confirmed in further clinical studies. There are limitations to our study. First, the samples used in this investigation were all retrospective, and prospective studies are required to verify the capacity of the AIPS for prediction. Second, more research must be conducted on the biological traits linked to the AIPS. Third, additional *in vivo* and *in vitro* research is required to investigate the mechanisms of numerous AIPS genes in GBM.

In summary, based on 117 machine learning algorithms, this study developed a reliable gene signature for predicting the prognosis of GBM patients. The AIPS not only showed excellent predictive performance in patients but also classified GBM into different states at the immune and genomic levels.

MATERIALS AND METHODS

GBM data collection and preprocessing

Datasets were collected from the Chinese Glioma Genome Atlas (CGGA) and Gene Expression Omnibus (GEO). Then, 1,036 GBM samples from six cohorts (CGGA_301 [$n = 124$], CGGA_325 [$n = 139$], CGGA_693 [$n = 249$], GSE13041 [$n = 191$], Rembrandt_475 [$n = 181$], and The Cancer Genome Atlas [TCGA] [$n = 152$]) were obtained for analysis. We combined the data from the six datasets and removed batch effects using the “limma” and “sva” R packages to reduce differences between the datasets due to experimental errors, allowing downstream analyses to consider only biological differences. By applying the “FactoMineR” and “factoextra” packages, we used PCA to evaluate batch removal.

Univariate Cox regression analysis and enrichment analysis

Univariate Cox regression analysis was performed for all six cohorts. OS-related genes were selected according to the following criteria: (1) $p < 0.05$ was achieved in at least four cohorts, and (2) the hazard ratio was consistently greater than 1 or less than 1. Then, we used the “clusterProfiler” R package to carry out GO and KEGG analyses for risk and protective OS-related genes, respectively.

Establishment and validation of the AIPS prognostic signature

To establish a comprehensive AIPS for GBM, we used 10 machine learning algorithms, including random forest (RSF), step Cox, least absolute shrinkage and selection operator (LASSO), gradient boosting machine (GBM), CoxBoost, partial least-squares regression for Cox (plsRcox), elastic network (Enet), ridge, survival support vector machine (Survival-SVM), and supervised principal component (Super PC) algorithms. Then, we arranged and combined the 10 machine learning algorithms into 117 algorithm integrations. In this study,

Figure 7. The expression patterns of the top 12 genes in the RSF model

(A) The expression levels of these genes in GBM and normal tissue were compared using the GEPIA website. (B and C) IHC and WB were used to evaluate the expression of CHI3L1, IMPDH1, MDK, MYBL2, and SETD8 in GBM and adjacent tissue. (D) Evaluation of CNV percentage. (E) Correlation analysis of CNV with mRNA expression. (F) Correlation between methylation and mRNA expression, * $p < 0.05$.

the merged cohort was used as the training cohort. We used 117 algorithm integrations to construct signatures based on 79 prognostic genes in the training cohort. The consistency index (C-index) values were calculated for all seven cohorts (CGGA_301, CGGA_325, CGGA_693, GSE13041, Rembrandt_475, TCGA, and training cohorts), and the model with the highest average C-index was regarded as the best prognostic signature. The predictive efficacy of the AIPS was evaluated in six test cohorts and the training cohort. The optimal cutoff scores for the high and low AIPS subgroups were calculated through the “survminer” package. The K-M curve and time-dependent receiver operating characteristic (tdROC) curve were used to evaluate the data. The “survminer” package was used to calculate the optimal cutoff score for the high and low AIPS score groups.

Comparison with published signatures and clinical correlation analysis

The AIPS was compared with 10 previously published GBM/glioma models using six GBM test cohorts and the training cohort. We calculated the risk scores of the seven cohorts based on the expression of signature genes and coefficients reported previously in the literature and evaluated their ability to predict the prognosis of GBM patients according to the C-index. In addition, we further compared their prognostic efficacy in all cohorts by univariate Cox regression analysis. We also explored the association of the AIPS with common clinical features.

Functional enrichment analysis of the AIPS

To explore the underlying biological mechanisms of the AIPS in GBM, the top 50 genes positively and negatively correlated with the AIPS were identified in the training cohort by Pearson analysis. Next, GO and KEGG analyses were performed based on these 100 significantly related genes. GSEA (Reactome pathway) was conducted to compare the high and low AIPS score groups using the “clusterProfiler” R package. The enrichment gene sets achieved a $p.adjust < 0.05$, which was considered to indicate statistical significance. The HALLMARK, KEGG, Reactome, Biocarta, and WikiPathways gene sets were obtained from the MSigDB database, and the terms associated with the AIPS score for each GBM patient were identified by the “GSVA” package in R.

Evaluation of the TIME and genome alterations

We explored the relationship between the AIPS and immune cells in the TIME. Immune cell evaluation was carried out using the “IOBR” package and was based on the following eight methods: MCPcounter, EPIC, xCell, CIBERSORT, IPS, quanTIseq, ESTIMATE, and TIMER.⁴⁷ The relationship between immune cell levels and AIPS scores was evaluated using Pearson correlation analysis. GBM mutation data (TCGA-GBM) was downloaded from the TCGA database. The assessment of somatic variation data was carried out via the “maftools” package.

Immunotherapy prediction and drug sensitivity

We evaluated the ability of the AIPS to predict the response to tumor immunotherapy using the TIDE algorithm. The TIDE scores and

immunotherapy responses of the GBM patients in the training group were obtained from the TIDE database. The IC50 is a common measure of a cell’s sensitivity to drugs and can help identify potentially sensitive drugs; a higher IC50 indicates a lower sensitivity to treatment. The “oncoPredict” R package was used to predict the IC50 value of each sample for multiple cancer drugs.⁴⁸ We identified potentially sensitive drugs by comparing the difference in IC50 values between the high and low AIPS score groups.

Expression patterns and single-cell analysis

The top 12 genes were selected according to their variable importance in RSF for subsequent analysis. The differences in the expression of 12 key genes between GBM and normal tissues were assessed via the GEPIA website. We explored the percentage and distribution of these 12 genes in GBM through the GSCA website. The TISCH2 website was used to analyze the expression of key genes in immune cells at the single-cell level.

Immunohistochemical staining

Differences in the expression of CHI3L1, IMPDH1, MDK, MYBL2, and SETD8 between GBM and adjacent normal tissues were verified in clinical samples. All specimens, including six GBM tissues and the relevant adjacent tissues, were obtained from the Department of Neurosurgery, Guangxi Medical University Cancer Hospital between May 2020 and August 2023. The information of clinical patients is shown in [Table S1](#). All GBM and adjacent tissue samples were approved by the Ethics Committee of the Guangxi Medical University Cancer Hospital and informed consent was obtained from the patients. The GBM and adjacent tissues were fixed with paraformaldehyde, dehydrated, embedded in paraffin, and sectioned for routine IHC staining. The slides were incubated with the following primary antibodies: MYBL2 polyclonal antibody (1:50, Proteintech, 18896-1-AP), MDK polyclonal antibody (1:200, Proteintech, 28546-1-AP), SETD8 polyclonal antibody (1:200, Proteintech, 14063-1-AP), CHI3L1 polyclonal antibody (1:200, Proteintech, 12036-1-AP), and IMPDH1 polyclonal antibody (1:200, Proteintech, 22092-1-AP). Finally, all the IHC-stained sections were photographed using a microscope.

Western blotting

Protease inhibitors and NP40 (Beyotime) lysates were used to lyse the tissues, and a BCA Protein Assay Kit (Thermo Fisher Scientific) was used to measure the protein concentration. Protein was electrophoresed on a 10% SDS-PAGE gel, transferred to a PVDF membrane, blocked for 2 h with 5% skim milk, and incubated with primary antibodies, including MYBL2 polyclonal antibody (1:1,000, Proteintech, 18896-1-AP), MDK polyclonal antibody (1:1,000, Proteintech, 28546-1-AP), SETD8 polyclonal antibody (1:1,000, Proteintech, 14063-1-AP), CHI3L1 polyclonal antibody (1:500, Proteintech, 12036-1-AP), and IMPDH1 polyclonal antibody (1:1,000, Proteintech, 22092-1-AP), overnight at 4°C, followed by incubation with secondary antibodies for 2 h at room temperature. Then, three further washes with TBST solution were performed. Finally, we

employed an enhanced chemiluminescence (ECL) developer (Abbkine) to detect the antibody signal.

Statistical analysis

All data cleaning, analysis, and plotting were performed in R4.2.2 software. The C-indices of the different groups were compared through the “Compare C” R package. Normally distributed variables were analyzed through t tests, and nonnormally distributed data were analyzed through the Wilcoxon rank-sum test. The comparison of categorical variables between the two groups was performed using the chi-square test. The difference in prognosis between the two groups was assessed by K-M survival analysis and the log rank test. The rest of the statistical methods have been described above. $p < 0.05$ was regarded as statistically significant.

DATA AND CODE AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. The data that support the results of the current study are available from the Chinese Glioma Genome Atlas (CGGA, <http://www.cgga.org.cn/>), Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>), The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) websites, Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu/>), GEPIA (<http://gepia.cancer-pku.cn/index.html>), GSCA (<http://bioinfo.life.hust.edu.cn/GSCA/#/>), and TISCH2 (<http://tisch.comp-genomics.org/home/>) websites.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omton.2024.200838>.

ACKNOWLEDGMENTS

This research was supported by the Youth Program of the Scientific Research Foundation of Guangxi Medical University Cancer Hospital (YQJ2022-10). Approval of the research protocol by an Institutional Review Board: The Ethics Committee of the Guangxi Medical University Cancer Hospital (Approval Number: KY2023820).

AUTHOR CONTRIBUTIONS

Q.J. and X.Y. drafted the manuscript and performed experiments. Q.J., S.A., and Q.H. were responsible for bioinformatics analysis. T.D. and J.Y. performed data collection and supervised this project. F.G. and L.M. provided technical support. S.A. and Q.H. designed this study, provided methodology, and revised the manuscript. All authors read and approved the submitted version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Kim, J.-H., Jeong, K., Li, J., Murphy, J.M., Vukadin, L., Stone, J.K., Richard, A., Tran, J., Gillespie, G.Y., Flemington, E.K., et al. (2021). SON drives oncogenic RNA splicing in glioblastoma by regulating PTBP1/PTBP2 switching and RBFOX2 activity. *Nat. Commun.* *12*, 5551.
- Yuan, Y., Wang, L.H., Zhao, X.X., Wang, J., Zhang, M.S., Ma, Q.H., Wei, S., Yan, Z.X., Cheng, Y., Chen, X.Q., et al. (2022). The E3 ubiquitin ligase HUWE1 acts through the N-Myc-DLL1-NOTCH1 signaling axis to suppress glioblastoma progression. *Cancer Commun.* *42*, 868–886.
- Batool, S.M., Muralidharan, K., Hsia, T., Falotico, S., Gambin, A.S., Rosenfeld, Y.B., Khanna, S.K., Balaj, L., and Carter, B.S. (2022). Highly Sensitive *EGFRvIII* Detection in Circulating Extracellular Vesicle RNA of Glioma Patients. *Clin. Cancer Res.* *28*, 4070–4082.
- Bikfalvi, A., Da Costa, C.A., Avril, T., Barnier, J.-V., Bauchet, L., Brisson, L., Cartron, P.F., Castel, H., Chevet, E., Chneiweiss, H., et al. (2023). Challenges in glioblastoma research: focus on the tumor microenvironment. *Trends Cancer* *9*, 9–27.
- Fraza, A., Rethacker, L., Jeudy, G., Colombo, M., Pasmant, E., Avril, M.-F., Toubert, A., Moins-Teisserenc, H., Roelens, M., Dalac, S., et al. (2020). BRAF inhibitor resistance of melanoma cells triggers increased susceptibility to natural killer cell-mediated lysis. *J. Immunother. Cancer* *8*, e000275.
- Kreatsoulas, D., Bolyard, C., Wu, B.X., Cam, H., Giglio, P., and Li, Z. (2022). Translational landscape of glioblastoma immunotherapy for physicians: guiding clinical practice with basic scientific evidence. *J. Hematol. Oncol.* *15*, 80.
- Li, S., Wang, C., Chen, J., Lan, Y., Zhang, W., Kang, Z., Zheng, Y., Zhang, R., Yu, J., and Li, W. (2023). Signaling pathways in brain tumors and therapeutic interventions. *Signal Transduct. Targeted Ther.* *8*, 8.
- Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* *131*, 803–820.
- Zhou, M., Deng, Y., Fu, Y., Liang, R., Liu, Y., and Liao, Q. (2023). A new prognostic model for glioblastoma multiforme based on coagulation-related genes. *Transl. Cancer Res.* *12*, 2898–2910.
- Xie, P., Yan, H., Gao, Y., Li, X., Zhou, D.-B., and Liu, Z.-Q. (2022). Construction of m6A-Related lncRNA Prognostic Signature Model and Immunomodulatory Effect in Glioblastoma Multiforme. *Front. Oncol.* *12*, 920926.
- Huang, K., Rao, C., Li, Q., Lu, J., Zhu, Z., Wang, C., Tu, M., Shen, C., Zheng, S., Chen, X., and Lv, F. (2022). Construction and validation of a glioblastoma prognostic model based on immune-related genes. *Front. Neurol.* *13*, 902402.
- Li, H., He, J., Li, M., Li, K., Pu, X., and Guo, Y. (2022). Immune landscape-based machine-learning-assisted subclassification, prognosis, and immunotherapy prediction for glioblastoma. *Front. Immunol.* *13*, 1027631.
- Lockett, P.H., Olufawo, M., Lamichhane, B., Park, K.Y., Dierker, D., Verastegui, G.T., Yang, P., Kim, A.H., Chheda, M.G., Snyder, A.Z., et al. (2023). Predicting survival in glioblastoma with multimodal neuroimaging and machine learning. *J. Neuro Oncol.* *164*, 309–320.
- Chen, W., Lei, C., Wang, Y., Guo, D., Zhang, S., Wang, X., Zhang, Z., Wang, Y., and Ma, W. (2023). Prognostic Prediction Model for Glioblastoma: A Ferroptosis-Related Gene Prediction Model and Independent External Validation. *J. Clin. Med.* *12*, 1341.
- Lei, C., Chen, W., Wang, Y., Zhao, B., Liu, P., Kong, Z., Liu, D., Dai, C., Wang, Y., Wang, Y., and Ma, W. (2021). Prognostic Prediction Model for Glioblastoma: A Metabolic Gene Signature and Independent External Validation. *J. Cancer* *12*, 3796–3808.
- Zhao, S., Ji, W., Shen, Y., Fan, Y., Huang, H., Huang, J., Lai, G., Yuan, K., and Cheng, C. (2022). Expression of hub genes of endothelial cells in glioblastoma-A prognostic model for GBM patients integrating single-cell RNA sequencing and bulk RNA sequencing. *BMC Cancer* *22*, 1274.
- Wang, W., Lu, Z., Wang, M., Liu, Z., Wu, B., Yang, C., Huan, H., and Gong, P. (2022). The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma. *Front. Immunol.* *13*, 998236.
- Xiao, S., Yan, Z., Zeng, F., Lu, Y., Qiu, J., and Zhu, X. (2022). Identification of a pyroptosis-related prognosis gene signature and its relationship with an immune microenvironment in gliomas. *Medicine (Baltimore)* *101*, e29391.

19. Yang, J., Wang, L., Xu, Z., Wu, L., Liu, B., Wang, J., Tian, D., Xiong, X., and Chen, Q. (2020). Integrated Analysis to Evaluate the Prognostic Value of Signature mRNAs in Glioblastoma Multiforme. *Front. Genet.* *11*, 253.
20. Molinaro, A.M., Taylor, J.W., Wiencke, J.K., and Wrensch, M.R. (2019). Genetic and molecular epidemiology of adult diffuse glioma. *Nat. Rev. Neurol.* *15*, 405–417.
21. Wang, Z., Wang, C., Lin, S., and Yu, X. (2021). Effect of TTN Mutations on Immune Microenvironment and Efficacy of Immunotherapy in Lung Adenocarcinoma Patients. *Front. Oncol.* *11*, 725292.
22. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* *24*, 1550–1558.
23. Wang, X., Guo, G., Guan, H., Yu, Y., Lu, J., and Yu, J. (2019). Challenges and potential of PD-1/PD-L1 checkpoint blockade immunotherapy for glioblastoma. *J. Exp. Clin. Cancer Res.* *38*, 87.
24. Le Joncour, V., Filppu, P., Hyvönen, M., Holopainen, M., Turunen, S.P., Sihto, H., Burghardt, L., Joensuu, H., Tynneninen, O., Jääskeläinen, J., et al. (2019). Vulnerability of invasive glioblastoma cells to lysosomal membrane destabilization. *EMBO Mol. Med.* *11*, e9034.
25. Chen, D.G., Zhu, B., Lv, S.Q., Zhu, H., Tang, J., Huang, C., Li, Q., Zhou, P., Wang, D.L., and Li, G.H. (2017). Inhibition of EGFR1 inhibits glioma proliferation by targeting CCND1 promoter. *J. Exp. Clin. Cancer Res.* *36*, 186.
26. Cong, P., Wu, T., Huang, X., Liang, H., Gao, X., Tian, L., Li, W., Chen, A., Wan, H., He, M., et al. (2021). Identification of the Role and Clinical Prognostic Value of Target Genes of m6A RNA Methylation Regulators in Glioma. *Front. Cell Dev. Biol.* *9*, 709022.
27. Tu, Z., Shu, L., Li, J., Wu, L., Tao, C., Ye, M., Zhu, X., and Huang, K. (2020). A Novel Signature Constructed by RNA-Binding Protein Coding Genes to Improve Overall Survival Prediction of Glioma Patients. *Front. Cell Dev. Biol.* *8*, 588368.
28. Giussani, M., Triulzi, T., Sozzi, G., and Tagliabue, E. (2019). Tumor Extracellular Matrix Remodeling: New Perspectives as a Circulating Tool in the Diagnosis and Prognosis of Solid Tumors. *Cells* *8*, 81.
29. Frederico, S.C., Hancock, J.C., Bretschneider, E.E.S., Ratnam, N.M., Gilbert, M.R., and Terabe, M. (2021). Making a Cold Tumor Hot: The Role of Vaccines in the Treatment of Glioblastoma. *Front. Oncol.* *11*, 672508.
30. Candi, E., Agostini, M., Melino, G., and Bernassola, F. (2014). How the TP53 Family Proteins TP63 and TP73 Contribute to Tumorigenesis: Regulators and Effectors. *Hum. Mutat.* *35*, 702–714.
31. Hinshaw, D.C., and Shevde, L.A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* *79*, 4557–4566.
32. Caligiuri, G., and Tuveson, D.A. (2023). Activated fibroblasts in cancer: Perspectives and challenges. *Cancer Cell* *41*, 434–449.
33. Tang, P.C.T., Chung, J.Y.F., Xue, V.W.W., Xiao, J., Meng, X.M., Huang, X.R., Zhou, S., Chan, A.S.W., Tsang, A.C.M., Cheng, A.S.L., et al. (2022). Smad3 Promotes Cancer-Associated Fibroblasts Generation via Macrophage–Myofibroblast Transition. *Adv. Sci.* *9*, 2101235.
34. Zhang, X., Ji, L., and Li, M.O. (2023). Control of tumor-associated macrophage responses by nutrient acquisition and metabolism. *Immunity* *56*, 14–31.
35. Xiang, X., Wang, J., Lu, D., and Xu, X. (2021). Targeting tumor-associated macrophages to synergize tumor immunotherapy. *Signal Transduct. Targeted Ther.* *6*, 75.
36. Liu, Z., Huang, K., He, Y., Hao, S., Wei, Z., and Peng, T. (2023). A pan-cancer analysis of the expression and prognostic significance of PDRG1. *Ann. Transl. Med.* *11*, 36.
37. Kida, W., Nakaya, M., Ito, A., Kozai, Y., and Bingo, M. (2022). A Case of Acquired Factor V Inhibitor Following Nivolumab Administration. *Cureus* *14*, e21670.
38. Zou, S., Zhang, Y., Zhang, L., Wang, D., and Xu, S. (2022). Construction and validation of a prognostic risk model of angiogenesis factors in skin cutaneous melanoma. *Aging* *14*, 1529–1548.
39. Shi, J.Y., Bi, Y.Y., Yu, B.F., Wang, Q.F., Teng, D., and Wu, D.N. (2021). Alternative Splicing Events in Tumor Immune Infiltration in Colorectal Cancer. *Front. Oncol.* *11*, 583547.
40. Xu, X., Wu, Y., Jia, G., Zhu, Q., Li, D., and Xie, K. (2023). A signature based on glycosyltransferase genes provides a promising tool for the prediction of prognosis and immunotherapy responsiveness in ovarian cancer. *J. Ovarian Res.* *16*, 5.
41. Cui, S., Feng, J., Tang, X., Lou, S., Guo, W., Xiao, X., Li, S., Chen, X., Huan, Y., Zhou, Y., and Xiao, L. (2023). The prognostic value of tumor mutation burden (TMB) and its relationship with immune infiltration in breast cancer patients. *Eur. J. Med. Res.* *28*, 90.
42. Yang, M.-C., Wu, D., Sun, H., Wang, L.-K., and Chen, X.-F. (2022). A Metabolic Plasticity-Based Signature for Molecular Classification and Prognosis of Lower-Grade Glioma. *Brain Sci.* *12*, 1138.
43. Zhang, F., Cai, H.-B., Liu, H.-Z., Gao, S., Wang, B., Hu, Y.-C., Cheng, H.-W., Liu, J.-X., Gao, Y., and Hong, W.-M. (2022). High Expression of C1SD2 in Relation to Adverse Outcome and Abnormal Immune Cell Infiltration in Glioma. *Dis. Markers* *2022*, 8133505–8133525.
44. Fan, X., Lu, H., Cui, Y., Hou, X., Huang, C., and Liu, G. (2018). Overexpression of p53 delivered using recombinant NDV induces apoptosis in glioma cells by regulating the apoptotic signaling pathway. *Exp. Ther. Med.* *15*, 4522–4530.
45. Zhang, Z., Liu, L., Ma, C., Cui, X., Lam, R.H.W., and Chen, W. (2021). An In Silico Glioblastoma Microenvironment Model Dissects the Immunological Mechanisms of Resistance to PD-1 Checkpoint Blockade Immunotherapy. *Small Methods* *5*, 2100197.
46. Arrieta, V.A., Dmello, C., McGrail, D.J., Brat, D.J., Lee-Chang, C., Heimberger, A.B., Chand, D., Stupp, R., and Sonabend, A.M. (2023). Immune checkpoint blockade in glioblastoma: from tumor heterogeneity to personalized treatment. *J. Clin. Invest.* *133*, e163447.
47. Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., Zhou, R., Qiu, W., Huang, N., Sun, L., et al. (2021). IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* *12*, 687975.
48. Maeser, D., Gruener, R.F., and Huang, R.S. (2021). oncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Briefings Bioinf.* *22*, bbab260.