Briefings in Bioinformatics, 22(5), 2021, 1–20

https://doi.org/10.1093/bib/bbab032
Problem Solving Protocol

Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients

Zihao Wang†, Yaning Wang†, Tianrui Yang, Hao Xing, Yuekun Wang, Lu Gao, Xiaopeng Guo, Bing Xing, Yu Wang and Wenbin Ma

Corresponding authors: Yu Wang, MD, Departments of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. Tel.: +86 15311860318. E-mail: ywang@pumch.cn; Wenbin Ma, MD, PhD, Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. Tel.: +86 13701364566. E-mail: mawb2001@hotmail.com
†These authors contribute equally to this work.

Abstract

Glioblastoma (GBM) is the most malignant and lethal intracranial tumor, with extremely limited treatment options. Immunotherapy has been widely studied in GBM, but none can significantly prolong the overall survival (OS) of patients without selection. Considering that GBM cancer stem cells (CSCs) play a non-negligible role in tumorigenesis and chemoradiotherapy resistance, we proposed a novel stemness-based classification of GBM and screened out certain

Zihao Wang, M.D., Ph.D., is a neurosurgeon from the Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on the multiomics analysis of intracranial tumors and translational applications in glioma.
Yaning Wang, M.D., Ph.D., is a neurosurgeon from the Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. Her researches focus on comprehensive treatment and management of intracranial malignant tumors and bioinformatics analysis.
Tianrui Yang is a Ph.D. student from the Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. Her researches focus on the immunotherapy of central nervous system malignancies.
Hao Xing is a Ph.D. student from the Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on the surgery of intracranial malignant tumors.
Yuekun Wang is a Ph.D. student from the Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. Her researches focus on malignant brain tumor oncogenesis and clinical management.
Lu Gao is a postdoc in Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on the comprehensive treatment of intracranial tumors and clinical data analysis.
Xiaopeng Guo is a postdoc in Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on the treatment and management of intracranial tumors and clinical data analysis.
Bing Xing is a professor in Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on management and treatment of intracranial tumors.
Yu Wang is an associate professor in Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on surgery, targeted therapy and immunotherapy of intracranial malignant tumors.
Wenbin Ma is a professor in Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. His researches focus on chemotherapy, molecular targeted therapy, drug resistance, tumor recurrence and gene therapy of malignant glioma and metastatic tumor.
Submitted: 26 November 2020; **Received (in revised form):** 18 January 2021

population more responsive to immunotherapy. The one-class logistic regression algorithm was used to calculate the stemness index (mRNAsi) of 518 GBM patients from The Cancer Genome Atlas (TCGA) database based on transcriptomics of GBM and pluripotent stem cells. Based on their stemness signature, GBM patients were divided into two subtypes via consensus clustering, and patients in Stemness Subtype I presented significantly better OS but poorer progression-free survival than Stemness Subtype II. Genomic variations revealed patients in Stemness Subtype I had higher somatic mutation loads and copy number alteration burdens. Additionally, two stemness subtypes had distinct tumor immune microenvironment patterns. Tumor Immune Dysfunction and Exclusion and subclass mapping analysis further demonstrated patients in Stemness Subtype I were more likely to respond to immunotherapy, especially anti-PD1 treatment. The pRRophetic algorithm also indicated patients in Stemness Subtype I were more resistant to temozolomide therapy. Finally, multiple machine learning algorithms were used to develop a 7-gene Stemness Subtype Predictor, which were further validated in two external independent GBM cohorts. This novel stemness-based classification could provide a promising prognostic predictor for GBM and may guide physicians in selecting potential responders for preferential use of immunotherapy.

Key words: stemness subgroup; mRNAsi; integrated multiomic analysis; immunotherapy; glioblastoma

Introduction

Glioma is the most common primary intracranial tumor and among them, glioblastoma (GBM) is the most malignant and invasive with a high recurrence rate [1]. The median overall survival (OS) and progression-free survival (PFS) of GBM patients is only 14–16 months and 6–7 months, separately [2]. The current standard treatment regimen for GBM is maximal surgical resection followed by concurrent chemoradiotherapy, adjuvant temozolomide (TMZ) chemotherapy and tumor-treating field if available [3]. Molecular targeted therapy and immunotherapy are also potential treatment options of GBM, but most of them are in the stage of clinical trials, and none has been found to prolong the OS of GBM patients, and only bevacizumab, which targets vascular endothelial growth factor (VEGF), was found to be able to extend PFS in GBM patients [4, 5]. Therefore, research on GBM treatment still has a long way to go.

Stemness refers to the self-renewal and differentiation potential of cells, which was first used in normal adult stem cells [6]. It has been found that in tumor tissue, there is a group of cells that have the stem cell-like features with self-renewal ability and can differentiate into different malignant cells with distinct phenotypes, named cancer stem cells (CSCs) [7]. These undifferentiated malignant cells are more likely to spread to distant sites than normal tumor cells, leading to disease progression and poor prognosis [8, 9]. Such cells also exist in GBM; they play a very important role in the initiation and treatment resistance of tumor and will also have a substantial impact on patients' response to immunotherapy [10].

Immunotherapy, a novel treatment option for malignant tumors, has been widely studied in recent years, among which the research results of immune checkpoint inhibitor (ICI) are particularly satisfactory. Immune checkpoints are a kind of immunosuppressive molecules that mainly regulate the immune response of T cells to avoid damage and destruction of normal tissues. The activation of immune checkpoints is one of the main causes of immune tolerance during the development of tumors. ICIs can avoid the inhibition of the human body's anti-tumor immune response, and then attack the tumor and inhibit its growth, among which anti-PD-1/PD-L1 is the most effective. This drug has already made great achievements in melanoma, colon cancer and non-small cell lung cancer [11, 12]. And the success in other cancers has also brought this treatment to the attention of neuro-oncologists. Gliomas, located in the central nervous system, a relative highly immunosuppressive microenvironment, are what researchers call 'cold tumors'. Although in recent years, a number of researches focused on immunotherapy in GBM provide the theoretical basis [13, 14], the results of some large scale phase III clinical trials, including CHECKMATE 143 for recurrent GBM, CHECKMATE 498 for newly diagnosed O6-methylguanine-DNA-methyltransferase (*MGMT*) promoter unmethylated GBM and CHECKMATE 548 for newly diagnosed *MGMT* promoter methylated GBM suggest that anti-PD-1/PD-L1 could not extend survival of GBM patients without selection [15, 16]. In addition to the immunosuppressive microenvironment, low tumor mutation burden (TMB) of GBM, especially newly diagnosed GBM, is another important reason for its failure to immunotherapy [17]. It was found that patients with high TMB may benefit from PD-1 blockade [18]. Therefore, how to screen out patients with high TMB more effectively, so as to identify patients who can benefit from immunotherapy is an urgent problem to be solved.

In this study, transcriptome analysis was performed on GBM patients to evaluate their stemness index. Subsequently, based on distinct stemness features, GBM patients were divided into two subtypes with distinct survival outcomes, functional annotations, and clinical features. Then, integrated analysis was used to analyze the differences of the genomic variations, tumor microenvironment and immunogenomic patterns of patients between two stemness subtypes, different benefits of immunotherapy and TMZ were also identified. Furthermore, the Stemness Subtype predictor that could quickly distinguish these two subtypes in GBM patients was constructed by multiple machine learning algorithms and validated in two independent GBM cohorts. Which provided possible means for screening patients who are more likely to have a positive response to immunotherapy. Our study aimed to facilitate individualized survival prediction and better treatment options for both physicians and GBM patients according to the novel stemness-based molecular classification.

Methods

Patient population and multiomic data acquisition

The gene expression profiles of pluripotent stem cells (PSCs), including induced PSCs (iPSCs) and embryonic stem cells (ESCs), were collected by the Progenitor Cell Biology Consortium (PCBC, https://progenitorcells.org/) and were obtained via the synapser

package (version 0.6.61) in R software. In addition, the level-3 gene expression data based on the AffyU133a platform and corresponding clinical and follow-up information of The Cancer Genome Atlas (TCGA) GBM patients were downloaded from the UCSC Xena (https://xenabrowser.net/); and the RNA sequencing (RNA-Seq) data based on the Illumina HiSeq platform were also obtained from the Chinese Glioma Genome Atlas (CGGA) database (http://www.cgga.org.cn). After excluding samples without complete clinical information, we finally enrolled 518 GBM patients from TCGA and 350 patients from CGGA database. In order to make the gene expression profiling comparable between different platforms, the Trans Per Million (TPM) values of RNA-Seq, robust multichip analysis (RMA)-processed values of miacroarry, and qRT-PCR data were log2 transformed and then normalized with the scale method by using the limma package in R [19, 20]. The demographics and follow-up data of the 868 GBM patients are displayed in Table 1. Furthermore, the somatic mutation data (MAF format) of 390 GBM patients based on the whole exome sequencing platform were also downloaded from the TCGA database. The mutation types and frequencies of genes were analyzed and visualized by using maftools and the GenVisR package in R [21, 22]. TMB, a potential biomarker for immunotherapy response, was defined as the total number of nonsynonymous mutations in the coding region per megabase [23]. Additionally, the copy number alteration (CNA) data of 518 GBM patients were obtained from the TCGA dataset, and significant amplifications or deletions in the whole genome were identified by GISTIC 2.0 [24]. To better exhibit the gain/loss alterations of chromosomes, CNA summary plots were visualized by Circos plots using the RCircos package in R [25]. The CNA burden was defined as the total number of genes with copy number changes in each sample [23].

Sample collection of the Peking Union Medical College Hospital (PUMCH) cohort

From January 2016 to October 2019, a total of 38 freshly frozen surgically resected GBM specimens with complete clinical and prognostic information were collected from patients without preoperative treatment at PUMCH. The diagnoses were confirmed by histopathology. The OS and PFS data were obtained through electronic medical records or telephone follow-up. The demographic and follow-up data of the 38 GBM patients are displayed in Table 1.

This study was approved by the Institutional Ethics Committee of PUMCH (Ethic code: JS-2012) in accordance with the ethical standards of the Institutional Ethics Committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent forms were signed and obtained from all participants by the TCGA and CGGA member institutions and PUMCH.

Calculation of the gene expression-based stemness index (mRNAsi)

The one-class logistic regression (OCLR) algorithm was used to calculate the stemness index based on gene expression profiles of normal PSCs, including iPSCs and ESCs [6]. We obtained 78 stem cell samples with 8087 protein-coding genes for each sample, and the expression of mRNAs was mean-centered. The stemness signature was generated via the OCLR algorithm by utilizing the gelnet (version 1.2.1) package in R [26]. Then, we calculated the Spearman correlations between the weight vectors of the stemness signature and mRNA expressions of GBM samples. Finally, the stemness index was mapped to the range of 0 to 1 using a linear transformation that subtracted the minimum and divided by the maximum of the Spearman correlation coefficient [6]. The stemness index generated by gene expression profiles was defined as mRNAsi.

The tumor immune microenvironment (TIME) patterns and immunogenomic features of GBM

Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) was employed to evaluate the tumor microenvironment and predict the tumor purity and abundances of intratumoral stromal and immune cells based on the gene expression profiles of GBM samples [27]. ESTIMATE can generate four types of scores: immune score (positively reflecting the abundance of immune cells), stromal score (positively reflecting the abundance of stromal cells), ESTIMATE score (positively reflecting nontumor composites), and tumor purity. In addition, CIBERSORT, a deconvolution algorithm based on linear support vector regression, was further utilized to quantify the compositions of 22 types of tumor-infiltrating immune cells (TIICs) based on the gene expression profiles of GBM samples [28]. He et al. introduced 29 immune signatures, representing the overall immune activity of tumors, containing the types, functions and molecular pathways of TIICs [29]. The enrichment levels of those immune gene sets were quantified by single-sample gene set enrichment analysis (ssGSEA) [30]. Then, based on the ssGSEA scores of the 29 immune signatures, unsupervised hierarchical clustering was performed to classify the GBM patients into different clusters, termed by immune subtypes [29].

Differential analysis of the high and low mRNAsi groups

The GBM patients were stratified into high and low mRNAsi groups based on the median value of the stemness index [31]. Kaplan–Meier (K-M) survival analysis was performed to evaluate the OS and PFS of patients with high and low mRNAsi, and the survival differences were evaluated by the two-sided log-rank test.

The differentially expressed genes (DEGs) between the high and low mRNAsi groups were screened by using the limma package in R [20]. Adjusted P-values were applied to correct the false positive results using the default Benjamini-Hochberg false discovery rate (FDR) method. FDR < 0.01 and $|\text{fold change (FC)}| > 2$ were considered the cutoff values for determining DEGs. Then, functional annotation and pathway enrichment analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, were performed on the DEGs by using the WebGestaltR package in R [32–34]. FDR < 0.05 was considered statistically significant.

Identification of the stemness-based molecular classification of GBM patients

Unsupervised consensus clustering, based on the k-means machine learning algorithm, was used to explore a novel molecular classification of GBM patients based on the expression of DEGs by using the ConsensusClusterPlus package in R [35]. The clustering procedure, with 1000 iterations, was performed by sampling 80% of the data in each iteration. The

Table 1. Demographics and clinicopathological features of GBM patients in the TCGA, CGGA and PUMCH cohort

Variables	TCGA cohort			CGGA cohort			PUMCH cohort		
	Total ($n=518$)	Stemness Subtype I ($n=233$)	Stemness subtype II ($n=285$)	Total ($n=350$)	Stemness subtype I ($n=58$)	Stemness subtype II ($n=292$)	Total ($n=38$)	Stemness subtype I ($n=28$)	Stemness subtype II ($n=10$)
Age (years)	57.5 ± 14.6	55.9 ± 16.4	58.9 ± 12.8	48.1 ± 13.3	47.4 ± 13.2	51.4 ± 13.3	57.5 ± 12.9	55.4 ± 13.0	63.6 ± 11.3
Gender									
Male	314	141	173	211	34	177	18	14	4
Female	204	92	112	139	24	115	20	14	6
OS status									
Alive	82	43	39	64	10	54	0	0	0
Dead	436	190	246	286	48	238	38	28	10
Progression									
Yes	430	186	244	209	37	172	23	18	5
No	88	47	41	141	21	120	15	10	5
Pretreatment KPS									
<80	101	46	55	NA	NA	NA	16	13	3
>=80	286	131	155	NA	NA	NA	22	15	7
NA	131	56	75	NA	NA	NA	0	0	0
Surgery									
Biopsy only	67	35	32	NA	NA	NA	4	3	1
Tumor resection	451	198	253	NA	NA	NA	34	25	9
Chemotherapy									
TMZ	266	118	148	61 (No)	8	53	17	11	6
TMZ + BEV	102	44	58	269 (Yes)	46	223	16	13	3
Others (No TMZ)	55	25	30	–	–	–	0	0	0
No or NA	95	46	49	20 (NA)	4	16	5	4	1
Radiotherapy									
No	106	48	58	48	4	44	10	6	4
Yes	412	185	227	283	49	234	28	22	6
NA	0	0	0	19	5	14	0	0	0
Standard chemoradiotherapy									
No	312	137	175	84	12	72	20	15	5
Yes	194	87	107	246	42	204	18	13	5
NA	12	9	3	20	4	16	0	0	0
IDH mutation status									
Wild type	463	197	266	271	39	232	24	15	9
Mutant	55	36	19	79	19	60	14	13	1
MGMT promoter status									
Methylated	229	110	119	169	25	144	8	5	3
Unmethylated	289	123	166	181	33	148	30	23	7
G-CIMP status									
Non G-CIMP	472	191	281	NA	NA	NA	NA	NA	NA
G-CIMP	46	42	4	NA	NA	NA	NA	NA	NA
TCGA subtype									
Classical	143	49	94	74	13	61	NA	NA	NA
Mesenchymal	152	24	128	122	10	112	NA	NA	NA
Neural	87	48	39	81	17	64	NA	NA	NA
Proneural	136	112	24	73	18	55	NA	NA	NA
1p/19q status									
Non-codeletion	NA	NA	NA	323	58	265	33	27	6
Codeletion	NA	NA	NA	17	0	17	5	1	4
NA	NA	NA	NA	10	0	10	0	0	0

'Others (No TMZ)' in pharmacotherapy included PCV, PCV + BEV and other drugs, including avastin, carmustine and irinotecan.
Abbreviations: GBM, glioblastoma; NA, not available; KPS, Karnofsky performance score; TMZ, temozolomide; BEV, bevacizumab; PCV, procarbazine lomustine vincristine.

optimal number of clusters was comprehensively determined by the relative change in the area under the cumulative distribution function (CDF) curves, the proportion of ambiguous clustering (PAC) algorithm, and the consensus heatmap [36]. We performed comparisons of the clinicopathological parameters within different clusters to further explore the associations between the stemness subtypes and the clinical features of GBM patients. Then, K-M survival analysis was performed to evaluate the prognosis of patients in different stemness subtype groups. Univariate and multivariate Cox regression analyses were performed to determine whether the performance of the stemness subtype in predicting prognosis could be independent of other clinicopathological variables [37].

Gene set variation analysis (GSVA)

GSVA was utilized to evaluate the most significantly enriched molecular pathways of the stemness subtypes by using the GSVA package in R [30]. Differential analysis of the enrichment scores of KEGG pathways between the two stemness subtypes was performed by the limma package in R [20]. The KEGG pathways with $|\log2FC| > 0.1$ and $FDR < 0.05$ were considered the most differentially enriched molecular pathways between the two stemness subtypes [38].

Prediction of immunotherapy and TMZ response

TIDE, short for The Tumor Immune Dysfunction and Exclusion (http://tide.dfci.harvard.edu/), was developed on the basis of two primary mechanisms of tumor immune evasion: the induction of T cell dysfunction in tumors with high infiltration of cytotoxic T lymphocytes (CTLs) and the prevention of T cell infiltration in tumors with low CTL levels [39]. The TIDE algorithm can determine the signatures of T cell dysfunction by testing how the expression of each gene in tumors interacts with the CTL infiltration level to influence patient survival and response to immunotherapy [39]. Hence, the clinical response to immunotherapy of GBM patients can be predicted by the TIDE algorithm using the gene expression profiles of GBM samples. Then, an unsupervised subclass mapping (https://cloud.genepattern.org/gp/) method was also used to predict the response to ICI therapy of the different stemness subtypes [40]. $FDR < 0.05$ was considered the threshold for a significant response or nonresponse to anti-PD1 and anti-CTLA4 therapy. Furthermore, we used the pRRophetic package to predict the chemotherapeutic response to TMZ, which was determined by the half maximal inhibitory concentration (IC50) of each GBM sample based on the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) [41, 42]. After integrating the expression profiles of cell lines (training set) and GBM samples (test set), the IC50 of TMZ in GBM patients was estimated by ridge regression analysis, and the prediction accuracy was evaluated by 10-fold cross-validation.

Connectivity Map (CMap) analysis

The CMap database (https://clue.io/) was used to explore potential compounds targeting the molecular pathways and genes associated with the stemness subtypes of GBM patients [43]. It can not only predict drugs based on gene expression signatures but also reveal the mode of action (MoA) of compounds targeting corresponding molecular pathways. The DEGs between the high and low mRNAsi groups were employed to query the CMap database, and the most significantly highly expressed genes of each stemness subtype were considered potential targets of compounds. The enrichment scores of compounds were calculated, and compounds with a negative enrichment score and $P < 0.05$ were considered potential therapeutic drugs for each stemness subtype.

Construction and validation of the Stemness Subtype Predictor by multiple machine learning methods

The 518 GBM patients were randomly classified into training ($N = 376$) and testing ($N = 142$) sets at a ratio of 7:3. First, in the training set, least absolute shrinkage and selection operator (LASSO) regression, support vector machine (SVM), Random Forest and Boruta (RFB), and extreme gradient boosting (XGBoost) analyses were performed to select the most important group-relevant features by calculating the importance score for each variable via the glmnet, rms, e1071, caret, randomForest, Boruta, and XGBoost packages in R [44–47]. The expression of the stemness-associated DEGs was selected as the input variable (independent variables), and the status of stemness subtypes was selected as the outcome (binary dependent variables, 0 or 1). The performance of the four machine learning algorithms for feature selection in the training set was evaluated by receiver operating characteristic (ROC) curves, and the areas under the ROC curve (AUCs) were subsequently compared. Afterwards, the intersecting genes among the LASSO, SVM, RFB and XGBoost analyses were considered the most critical stemness subtype-related genes and were visualized by a Venn diagram. Finally, multivariate logistic regression analysis was performed on the critical genes for constructing the predictive model, which was termed by 'Stemness Subtype Predictor' [48]. The performance of the Stemness Subtype Predictor was investigated by the ROC curve to determine the optimal cutoff values in discriminating different subtypes, as well as the AUC, sensitivity, specificity, and accuracy. Finally, the predictive performance of the Stemness Subtype Predictor was also validated in the test set in a similar way.

Quantitative real-time polymerase chain reaction (qRT-PCR) analysis of the PUMCH cohort

The primers used for qRT-PCR are summarized in Supplementary Table 1. Total RNA was extracted from 38 freshly frozen GBM specimens by SuPerfecTRITM Reagent (Pufei Biotech, Cat. No. 3101–100) and reverse transcribed using the M-MLV RT Reagent Kit (Promega, Cat. No. M1705) according to the manufacturer's protocol. The SYBR Master Mixture Kit (Takara, Cat. No. DRR041B) was utilized to amplify the resulting cDNA, and qRT-PCR was performed using the LightCycler 480 II Real-Time PCR Detection System (Roche, Switzerland) according to the manufacturers' instructions. Each experiment was conducted at least three times. The expression of target gene was calculated using the $2^{-\Delta Ct}$ method relative to the geometric mean of three housekeeping genes (*GAPDH*, *Beta-Actin*, and *U1*). Then, the expression levels of genes were normalized by using the scale function in R [49].

In addition, we investigated the most common biomarkers of GBM, including tumor protein P53 (*TP53*), isocitrate dehydrogenase 1/2 (*IDH1/2*), v-raf murine sarcoma viral oncogene homolog B1 (*BRAF*), gene of phosphate and tension homology deleted on chromosome ten (*PTEN*), epidermal growth factor receptor (*EGFR*), telomerase reverse transcriptase (*TERT*) promoter, alpha-thalassemia chromatin remodeler (*ATRX*) mutation, 1p19q codeletion status, and *MGMT* promoter methylation

status. *TP53*, *IDH1/2*, *BRAF*, *PTEN*, *EGFR*, *TERT* promoter, and *ATRX* mutation status were measured by direct sequencing [50–52], 1p/19q codeletion status was determined by FISH [53], and *MGMT* promoter methylation status was measured by pyrosequencing as described before [54].

Statistical analysis

The independent Student's t test for continuous data and the $\chi 2$ test for categorical data were utilized for pairwise comparisons between groups. The Mann–Whitney *U* test was used to compare categorical variables and non-normally distributed variables between two groups. The Kruskal–Wallis test was used to compare multiple groups. Correlations between normally distributed variables were assessed with Pearson's correlation test, while correlations between non-normally distributed variables were assessed with Spearman's correlation test. A *P*-value <0.05 and |correlation coefficient (R)| >0.3 were considered significantly correlated. The statistical analyses in this study were performed by using SPSS 22.0 (SPSS, Inc., Chicago, Illinois, USA) and R 3.6.1 software. A two-tailed P-value <0.05 was considered statistically significant. Odds ratios (ORs), hazard ratios (HRs) and 95% confidence intervals (CIs) are reported if necessary.

Results

Associations between the stemness index and clinical features

The overall workflow of the present study is displayed in Supplementary Figure 1. By using the OCLR algorithm, the stemness index based on the gene expression profiles of 518 GBM patients was calculated and then ranked from low to high to explore the associations between mRNAsi and clinical features (Figure 1A and C). As shown in Figure 1B, patients aged <40 years had significantly higher mRNAsi scores than older patients, while the stemness index scores of the 40–59, 60–79 and ≥ 80 groups did not differ significantly among the groups. Regarding the TCGA molecular subtypes, the proneural subtype had the highest mRNAsi scores, followed by the neural, classical, and mesenchymal subtypes (all $P < 0.05$). In addition, patients with progression ($P < 0.001$), those with glioma-CpG island methylator phenotype (G-CIMP) status ($P < 0.001$), and those with *MGMT* promoter methylation ($P = 0.004$) demonstrated significantly higher mRNAsi scores (Figure 1B). Combined with somatic mutation data, we found no significant correlation between mRNAsi and TMB. However, regarding the most common biomarkers of GBM, the stemness index was significantly lower in *PTEN*-mutant samples than in *PTEN*-wild-type samples ($P = 0.037$) but significantly higher in *TP53*-mutant ($P < 0.001$), *IDH*-mutant ($P = 0.004$) and *ATRX*-mutant ($P < 0.001$) samples than in wild-type samples (Figure 1D).

Associations between the stemness index and TIME patterns

First, the enrichment levels of the 29 immune signatures, representing the overall immune activity of GBM, were quantified by ssGSEA, and 518 GBM patients were classified into three immune subtypes by utilizing unsupervised hierarchical clustering. The high-immunity group, containing 21 (4.1%) patients, was defined as having 'immune hot' tumors due to having the highest enrichment scores; the low-immunity group, containing 265 (51.2%) patients, was defined as having 'immune cold' tumors due to having the lowest enrichment scores; and the medium immunity group, containing 232 (44.8%) patients, was defined as having 'immune altered' tumors, indicating the potential to transform into hot or cold tumors (Figure 2A). Then, the TIME patterns were evaluated by the ESTIMATE and CIBERSORT algorithms. The stemness index was significantly negatively correlated with the immune, stromal, and ESTIMATE scores, indicating that the infiltration levels of immune and stromal cells decrease with the elevation of the stemness of GBM (Figure 2B). However, no significant correlation between mRNAsi and tumor purity was observed. Additionally, the immune and stromal scores were both significantly the highest in high-immunity tumors, indicating high abundances of immune cells and stromal cells, followed by medium- and low-immunity tumors (Figure 2C). In contrast, mRNAsi decreased while tumor purity increased from the high-immunity group to the low-immunity group. Then, the abundances of 22 types of immune cells were quantified by the CIBERSORT algorithm (Figure 2D). As shown in Figure 2E, the stemness index was significantly positively correlated with T cell subsets [including follicular helper ($R = 0.63$, $P < 0.001$), naive CD4 ($R = 0.38$, $P = 0.002$), and memory activated CD4 ($R = 0.32$, $P = 0.041$) T cells], activated natural killer (NK) cells ($R = 0.49$, $P < 0.001$), memory B cells ($R = 0.48$, $P < 0.001$), and plasma cells ($R = 0.35$, $P < 0.001$); meanwhile the stemness index was significantly negatively correlated with M2 macrophages ($R = -0.47$, $P < 0.001$) and monocytes ($R = -0.32$, $P = 0.047$).

mRNAsi correlated with contradictory OS and PFS outcomes of GBM patients

According to the median value of the stemness index, the 518 GBM patients were stratified into high and low mRNAsi groups. K-M survival analysis indicated that the high mRNAsi group presented significantly better OS (HR = 0.803, log-rank $P = 2.215 \times 10^{-2}$) and poorer PFS (HR = 1.284, log-rank $P = 9.752 \times 10^{-3}$) than the low mRNAsi group (Figure 3A and B). The median survival time (MST) of the high and low mRNAsi groups was very close for both OS (1.21 and 1.13 years) and PFS (0.54 and 0.63 years). Subgroup analysis also demonstrated that when patients were stratified by different clinical variables, high mRNAsi remained an indicator of favorable OS and unfavorable PFS (Supplementary Figure 2A and B).

Considering the interesting prognostic difference between the high and low mRNAsi groups, we further performed differential expression analysis between the two groups. A total of 130 DEGs were identified, including 41 upregulated and 89 downregulated genes in the high mRNAsi group (Figure 3C, Supplementary Table 2). Then, functional enrichment analysis was performed by WebGestaltR. There were 49 significantly enriched biological processes (BPs), including chromatid/chromosome segregation, nuclear division, and cell division; 94 significantly enriched cellular components (CC), including extracellular matrix and chromosome; 44 significantly enriched molecular functions (MFs), including glycosaminoglycan binding and peptidase regulator activity; and 37 significantly enriched KEGG pathways, including the focal adhesion, cell cycle, and p53 signaling pathways (Figure 3D).

Furthermore, a total of 35 out of 130 DEGs (26.9%) had a mutation frequency > 1%, and most (80.0%, 28/35) were downregulated in the high mRNAsi group (Figure 3E). The differential analysis of copy number variations between the two groups revealed that compared with the low mRNAsi group, 40 (30.8%) genes had significant amplifications and 82 (63.1%) genes had significant deletions in the high mRNAsi group (Figure 3F).

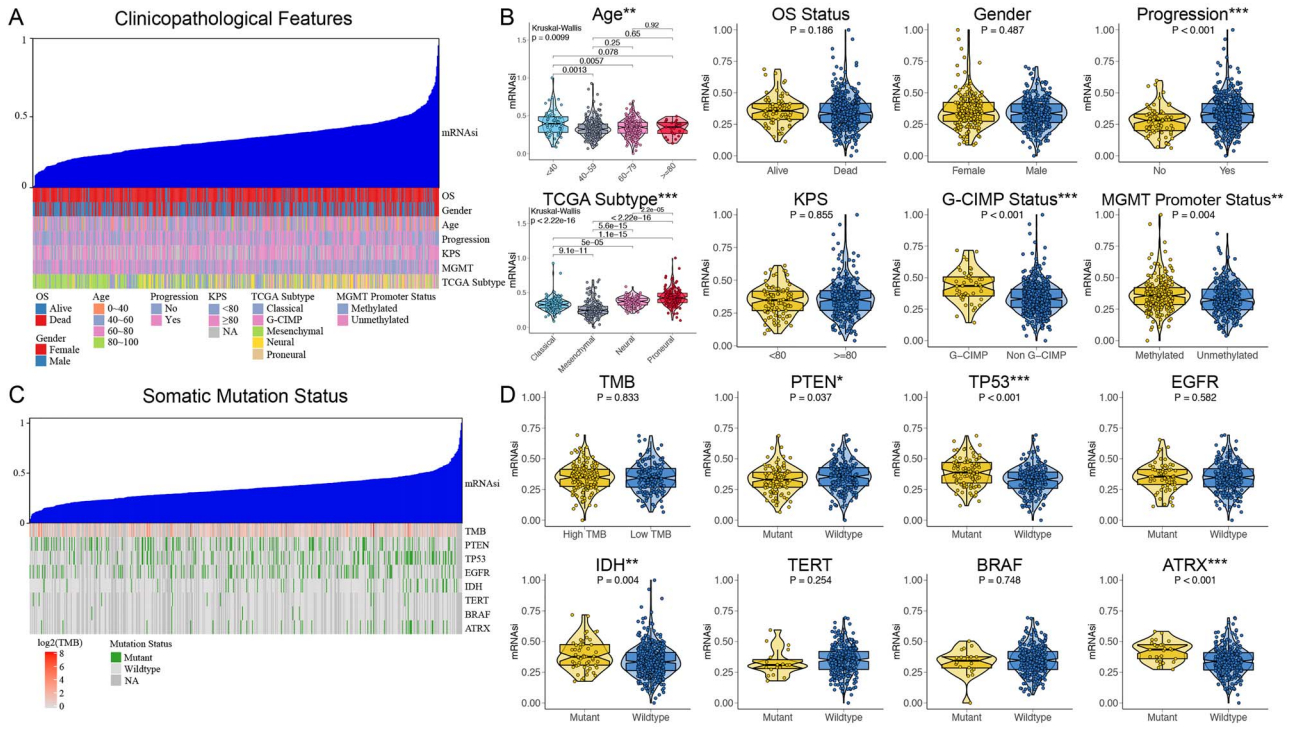

Figure 1. The clinical and molecular features associated with the stemness index (mRNAsi) in GBM patients. (**A**) An overview of the association between mRNAsi and clinicalpathological features of patients. Columns represented samples ranked by mRNAsi from low to high (top row), and rows represent known clinical and molecular characteristics associated with mRNAsi. (**B**) Violin plots of mRNAsi in individual samples of GBM patients, stratified by age, OS status, gender, progression, TCGA subtype, KPS, G-CIMP status and *MGMT* promoter status. (**C**) An overview of the association between mRNAsi and somatic mutation status of the most popular biomarkers of GBM and TMB. (**D**) Violin plots of mRNAsi in individual samples of GBM patients, stratified by TMB, *PTEN*, *TP53*, *EGFR*, *IDH*, *TERT* promoter, *BRAF* and *ATRX* mutation status. * means $P < 0.05$, ** means $P < 0.01$, and *** means $P < 0.001$.

Identification of two stemness subtypes with distinct survival outcomes, functional annotations, and clinical features

Unsupervised consensus clustering was used to explore a novel molecular classification of GBM patients based on the expression patterns of 130 stemness-based DEGs. According to the relative change in the area under the CDF curve, the PAC algorithm, and the consensus heatmap, the optimal number of clusters was determined to be two (k value = 2) (Figure 4A–C). Hence, all GBM patients were categorized into two groups, which were termed Stemness Subtype I (233 patients, 45.0%) and Stemness Subtype II (285 patients, 55.0%) (Figure 4D). K-M survival analysis indicated that GBM patients in the Stemness Subtype I group presented significantly better OS (HR = 0.606, log-rank $P = 1.987 \times 10^{-3}$) and poorer PFS (HR = 1.349, log-rank $P = 2.499 \times 10^{-7}$) than those in the Stemness Subtype II group (Figure 4E and F). The median OS time of Stemness Subtype I group patients was longer than that of Stemness Subtype II group patients (1.21 versus 1.05 years), whereas the median PFS time of Stemness Subtype I group patients was markedly shorter than that of Stemness Subtype II group patients (0.48 versus 0.71 years). Subgroup analysis also demonstrated that when patients were stratified by different clinical variables, Stemness Subtype I remained an indicator of favorable OS and unfavorable PFS in GBM patients (Supplementary Figure 2C and D). In addition, we performed univariate and multivariate Cox regression analyses to evaluate the prognostic significance of the stemness subtypes combined with various clinicopathological variables. In the TCGA cohort, univariate and subsequent multivariate analyses indicated that the stemness subtype was significantly associated with OS (Table 2) and PFS (Table 3), suggesting that the stemness subtype was an independent prognostic factor for predicting the OS and PFS of GBM patients.

GSVA was performed to explore the molecular pathways and underlying mechanisms related to the stemness subtypes of GBM. A total of 36 differentially enriched molecular pathways were identified, including 4 pathways positively correlated with Stemness Subtype I and 32 pathways positively correlated with Stemness Subtype II (Figure 4G). Stemness Subtype I tumors mainly correlated with the cell cycle, DNA replication, spliceosome, and terpenoid backbone biosynthesis, whereas Stemness Subtype II tumors mainly correlated with molecular pathways related to tumorigenesis (e.g., apoptosis and focal adhesion), metabolism (e.g., glycosaminoglycan degradation), and immune responses (e.g., B cell receptor signaling pathway, antigen processing and presentation, and leukocyte migration).

Subsequently, the demographics and clinicopathological features of GBM patients in the Stemness Subtype I and II groups were compared (Table 1). As shown in Figure 5A, patients in the Stemness Subtype I group were significantly younger in age at diagnosis than those in the Stemness Subtype II group (55.9 ± 16.4 versus 58.9 ± 12.8 years, $P = 0.022$). The composition of the TCGA molecular subtype in the Stemness Subtype I group [proneural (48.1%), classical (21.0%), neural (20.6%) and mesenchymal (10.3%)] was significantly different from that in the Stemness Subtype II group [mesenchymal (44.9%), classical (33.0%), neural (13.7%) and proneural (8.4%)] ($P < 0.001$). Moreover, the proportion of patients with G-CIMP status was significantly higher in the Stemness Subtype I group than in the Stemness

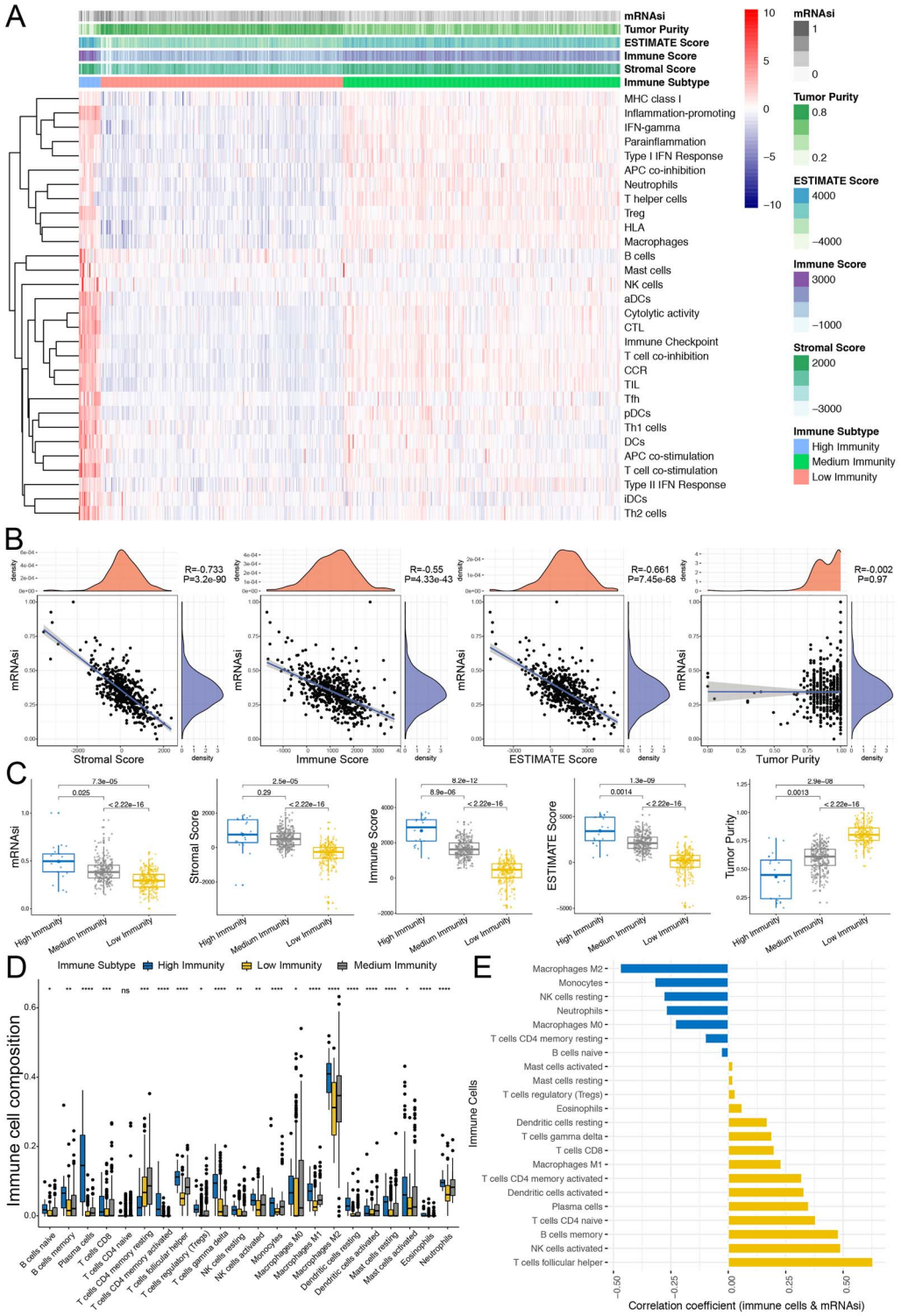

Figure 2. The tumor immune microenvironment patterns and immunogenomic features of GBM associated with the mRNAsi. (**A**) The immune subtypes of GBM patients were categorized on the basis of the overall immune activity of GBM. Most of the patients were classified as low immunity ('immune cold' tumor) group, followed by medium immunity ('immune altered') and high immunity ('immune hot') group. (**B**) Correlation analysis between mRNAsi and the stromal score, immune score, ESTIMATE score and tumor purity evaluated by ESTIMATE algorithm. (**C**) Comparisons of mRNAsi, the infiltration level of stromal and immune cells, the ESTIMATE score and tumor purity in different immune subtypes by boxplots. (**D**) Comparisons of the abundances of 22 immune cells in three immune subtypes. 'ns' means $P > 0.05$, * means $P < 0.05$, ** means $P < 0.01$, *** means $P < 0.001$ and **** means $P < 0.0001$. (**E**) Correlation analysis between immune cells and mRNAsi. Yellow bars meant correlation coefficient > 0, and blue bars meant correlation coefficient < 0.

Table 2. Univariate and multivariate cox proportional hazards analysis of clinicopathological variables based on overall survival (OS) in the TCGA GBM training cohort, CGGA and PUMCH validation cohort

For OS variables	TCGA training cohort ($n = 518$)				CGGA validation cohort ($n = 350$)				PUMCH validation cohort ($n = 38$)			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age	1.032(1.025–1.040)	1.98e-17	1.024(1.015–1.032)	8.87e-09	1.078(1.048–1.108)	1.63e-05	1.044(1.035–1.052)	3.42e-02	1.017(1.009–1.025)	2.96e-03	1.010(0.988–1.034)	0.37
Gender	1.160(0.956–1.408)	0.13	–	–	1.063(0.837–1.350)	0.61	–	–	1.192(0.608–2.337)	0.75	–	–
Surgery	1.141(0.869–1.499)	0.34	–	–	NA		NA		2.554(0.849–7.684)	0.095	–	–
KPS	1.031(0.871–1.221)	0.72	–	–	NA		NA		1.106(0.571–2.141)	0.77	–	–
Chemotherapy	1.288(1.179–1.407)	2.03e-08	1.134(1.033–1.245)	8.12e-03	1.783(1.345–2.364)	5.83e-03	1.642(1.204–2.240)	1.73e-03	1.555(1.073–2.442)	2.29e-05	1.671(1.383–2.853)	3.79e-03
Radiotherapy	3.437(2.730–4.327)	7.92e-26	2.481(1.917–3.212)	5.10e-12	1.557(1.146–2.114)	4.62e-05	1.307(1.064–1.549)	3.09e-02	2.937(1.374–6.278)	5.43e-03	4.506(1.69–11.479)	1.60e-03
Standard chemoradiotherapy	1.166(0.976–1.395)	0.091	–	–	1.209(0.907–1.512)	0.23	–	–	1.198(0.620–2.316)	0.59	–	–
IDH mutation status	2.354(1.695–3.271)	3.33e-07	1.680(1.176–2.402)	4.40e-03	1.255(1.216–1.294)	1.16e-08	1.345(1.059–1.630)	4.38e-02	2.264(0.683–7.503)	0.18	–	–
MGMT promoter methylation status	0.655(0.542–0.793)	1.44e-05	0.647(0.533–0.785)	9.95e-06	0.810(0.641–0.969)	2.64e-03	0.777(0.607–0.994)	3.27e-02	0.261(0.097–0.701)	7.68e-03	0.366(0.115–0.859)	4.16e-03
G-CIMP status	0.329(0.224–0.482)	1.30e-08	0.658(0.419–1.033)	0.069	NA		NA		NA		NA	
TCGA subtype	0.927(0.856–1.005)	0.065	–	–	0.915(0.875–0.955)	1.09e-02	0.909(0.814–1.015)	0.089	NA		NA	
1p/19q codeletion status	NA		NA		0.913(0.662–1.259)	0.58	–	–	0.883(0.337–.311)	0.799	–	–
Stemness subtype	1.326(1.097–1.604)	3.58e-04	1.129(1.090–1.169)	2.24e-03	1.347(1.308–1.387)	6.02e-05	1.483(1.073–2.048)	1.69e-03	2.289(1.672–4.452)	2.08e-07	2.702(1.154–4.155)	3.58e-03

Abbreviations: OS, overall survival; GBM, glioblastoma; NA, not available; HR, hazard ratio; CI, confidence interval; KPS, Karnofsky performance score; TMZ, temozolomide; BEV, bevacizumab; PCV, procarbazine lomustine vincristine. 'Others (No TMZ)' in pharmacotherapy included PCV, PCV + BEV and other drugs, including avastin, carmustine and irinotecan.
For continuous variables, they are centered and scaled (standard deviation set to one) in the models, and HR is the relative hazard when increasing the variable one standard deviation.
All statistical tests were two sided. Bold type means $P < 0.05$.

Table 3. Univariate and multivariate cox proportional hazards analysis of clinicopathological variables based on progression free survival (PFS) in the TCGA GBM training cohort, CGGA and PUMCH validation cohort

For PFS variables	TCGA training cohort ($n = 518$)				CGGA validation cohort ($n = 350$)				PUMCH validation cohort ($n = 38$)			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age	1.018(1.011–1.026)	4.50e-07	1.007(0.9995–1.015)	0.065	1.008(0.919–1.097)	0.11	–	–	0.977(0.952–1.221)	0.073	–	–
Gender	1.197(0.986–1.452)	0.068	–	–	1.009(0.720–1.414)	0.96	–	–	1.078(0.928–2.505)	0.33	–	–
Surgery	1.233(0.931–1.632)	0.14	–	–	NA		NA		1.157(0.283–3.502)	0.57	–	–
KPS	0.989(0.848–1.154)	0.89	–	–	NA		NA		0.676(0.308–1.484)	0.33	–	–
Chemotherapy	1.063(0.974–1.160)	0.17	–	–	1.052(0.686–1.615)	0.82	–	–	1.190(0.405–3.498)	0.75	–	–
Radiotherapy	1.661(1.295–2.131)	6.58e-05	1.357(1.051–1.752)	1.91e-03	1.240(1.189–1.292)	3.34e-05	1.242(1.151–1.332)	3.07e-02	2.858(1.397–5.846)	4.04e-03	2.493(1.264–3.983)	2.04e-02
Standard chemoradiotherapy	0.949(0.798–1.128)	0.55	–	–	0.879(0.681–1.079)	0.15	–	–	0.744(0.336–1.644)	0.46	–	–
IDH mutation status	2.287(1.638–3.192)	1.18e-06	1.469(1.020–2.114)	3.86e-02	1.261(1.140–1.381)	1.98e-03	1.246(1.154–1.339)	2.82e-02	2.495(1.720–4.379)	4.66e-05	2.607(2.329–2.885)	2.98e-03
MGMT promoter methylation status	0.769(0.637–0.928)	6.16e-05	0.784(0.646–0.950)	1.31e-02	0.630(0.351–0.938)	3.02e-04	0.699(0.406–0.988)	4.02e-02	0.616(0.244–0.891)	3.05e-05	0.330(0.194–0.471)	1.37e-02
G-CIMP status	0.306(0.207–0.450)	2.08e-02	0.731(0.420–1.171)	0.286	NA		NA		NA		NA	
TCGA subtype	0.904(0.834–0.981)	1.48e-02	1.090(0.988–1.203)	0.115	1.177(1.006–1.377)	0.042	1.162(0.989–1.365)	0.068	NA		NA	
1p/19q codeletion status	NA		NA		1.051 (0.684–1.615)	0.82	–	–	1.286(0.376–4.398)	0.69	–	–
Stemness subtype	0.606(0.500–0.735)	3.25e-07	0.684(0.555–0.844)	3.93e-04	0.848(0.726–0.983)	2.39e-05	0.481(0.304–0.762)	1.80e-03	0.317(0.107–0.639)	3.81e-03	0.252(0.075–0.846)	2.56e-02

Abbreviations: PFS, progression free survival; GBM, glioblastoma; NA, not available; HR, hazard ratio; CI, confidence interval; KPS, Karnofsky performance score; TMZ, temozolomide; BEV, bevacizumab; PCV, procarbazine lomustine vincristine.
'Others (No TMZ)' in pharmacotherapy included PCV, PCV + BEV, and other drugs, including avastin, carmustine and irinotecan.
For continuous variables, they are centered and scaled (standard deviation set to one) in the models, and HR is the relative hazard when increasing the variable one standard deviation.
All statistical tests were two sided. Bold type means $P < 0.05$.

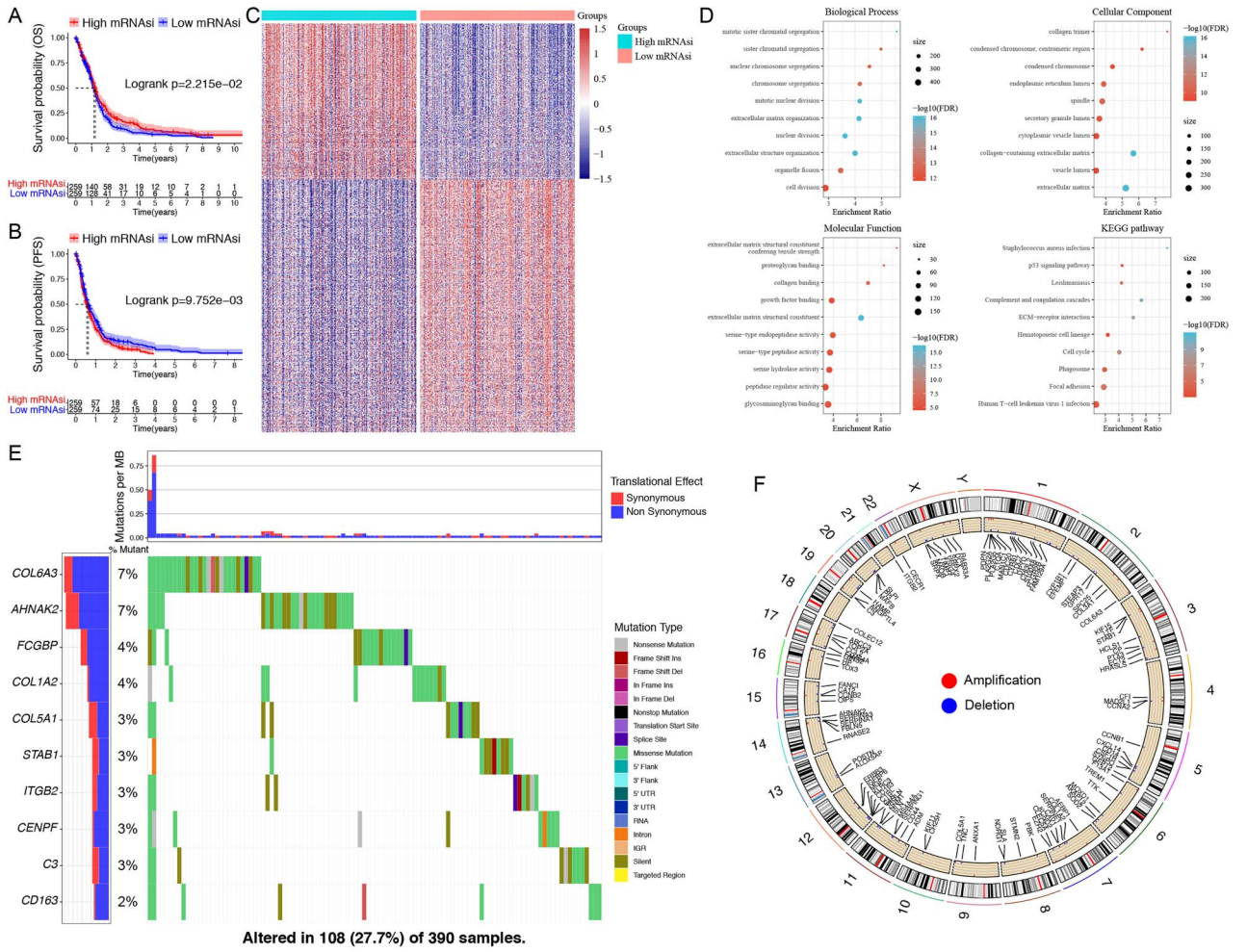

Figure 3. Survival analysis and differential expression analysis between high and low mRNAsi groups. Kaplan–Meier survival analyses indicated significantly better OS (**A**) and poorer PFS (**B**) in the high mRNAsi groups. (**C**) The heatmap showed the expression levels of DEGs between two groups. Red represented high expressions, and blue represented low expressions of genes. (**D**) The functional enrichment analyses of DEGs, including significantly enriched biological processes, cellular components, molecular functions and KEGG pathways. (**E**) Oncoplot of 10 most frequently mutated DEGs, which were altered in 108 GBM samples. (**F**) The differential analysis of copy number variations between two groups was visualized by Circos plot, which revealed that compared with the low mRNAsi group, 40 (30.8%) genes were significantly amplified, and 82 (63.1%) were significantly deleted in the high mRNAsi group. Red dots represented amplifications, blue dots represented deletions and black dots represented no significant CNAs.

Subtype II group (18.0% versus 1.4%, $P < 0.001$). However, OS status, progression status, sex, Karnofsky performance status (KPS) score, and *MGMT* promoter methylation status did not differ between the two subtypes. Finally, the stemness index of patients with Stemness Subtype I (0.43 ± 0.12) was significantly higher than that of patients with Stemness Subtype II (0.27 ± 0.08), with $P < 2.0 \times 10^{-16}$ (Figure 5A). This demonstrated that patients in the Stemness Subtype I group had higher levels of neoplastic stemness, which suggested stronger potential for the self-renewal, differentiation, and proliferation of tumor cells and might explain their poorer PFS.

Stemness Subtype I possessed a higher CNA burden and TMB

Previous studies have investigated the potential value of genomic alterations in regulating tumor immunity and immune infiltration patterns [55, 56]. Hence, CNA analysis and somatic mutation analysis were performed to explore the distinct genomic variations in the different stemness subtypes. As shown in Figure 5B, patients with Stemness Subtype I tended to bear a greater burden of copy number amplifications ($P = 1.8 \times 10^{-6}$) and deletions ($P = 0.042$) than those with Stemness Subtype II. Somatic mutation analysis revealed that each stemness subtype possessed specific top mutated genes (Figure 5C and D). *TP53* (42%) was the most frequently mutated gene in Stemness Subtype I, whereas *PTEN* (39%) was the most frequently mutated gene in Stemness Subtype II. However, *ATRX* and *IDH1*, with mutation frequencies <10%, were not observed in the top 10 mutated genes of Stemness Subtype II. Regarding the most common biomarkers of GBM, the proportion of patients with *TP53* mutations in the Stemness Subtype I group (41.6%) was significantly higher than that in the Stemness Subtype II group (22.3%, $P < 0.001$). The same situations were also observed for *IDH* and *ATRX*, indicating that the mutation frequencies of *IDH* (I versus II, 15.5 versus 6.7%; $P = 0.001$) and *ATRX* (I versus II, 16.8% versus 2.3%; $P < 0.001$) in the Stemness Subtype I group were significantly higher (Figure 5E). No significant difference was found regarding the mutation frequencies of *BRAF*, *PTEN*, *EGFR* and *TERT* between the two subtypes (Figure 5E). In addition,

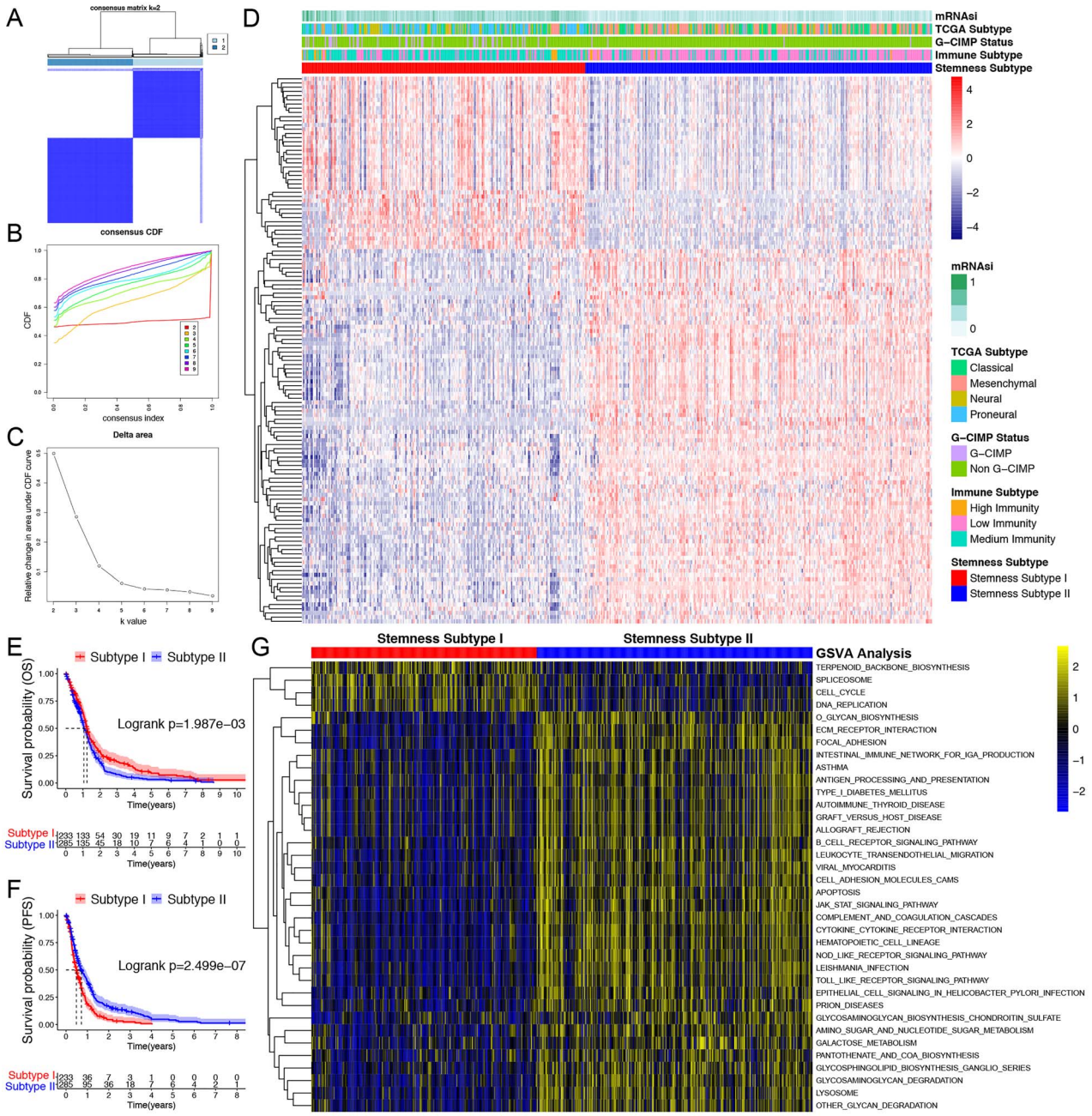

Figure 4. Identification of two stemness subtypes with distinct survival outcomes and functional annotations. (**A**) Consensus clustering matrix for k = 2, which was the optimal cluster number. (**B**) CDF curves of the consensus score from k = 2 to 9. (**C**) The relative change in the area under the CDF curve from k = 2 to 9. (**D**) The heatmap of the expression patterns of 130 DEGs, with red indicating high expressions and blue indicating low expressions. The upper columns were consisted of mRNAsi, and four molecular classification methods of GBM. Kaplan–Meier survival analysis exhibited significantly better OS (**E**) and poorer PFS (**F**) in patients with Stemness Subtype I. (**G**) Heatmap illustrated the enrichment scores of 36 differentially enriched molecular pathways evaluated by GSVA analysis between Stemness Subtype I and II. Yellow represented high enrichment scores, and blue represented low enrichment scores.

patients in the Stemness Subtype I group had significantly higher TMB than those in the Stemness Subtype II group ($P = 4.5 \times 10^{-13}$). All these findings could suggest underlying differences in the immunotherapy response of the two stemness subtypes.

The stemness subtypes had distinct TIME and immunogenomic patterns

First, the ESTIMATE algorithm was performed to reveal the compositions of the TIME of the two stemness subtypes. Compared with that in Stemness Subtype II, the immune, stromal, and ESTIMATE scores were significantly lower and the tumor purity score was significantly higher in Stemness Subtype I (all $P < 2.2 \times 10^{-16}$), demonstrating low abundances of immune and stromal cells and high tumor purity in Stemness Subtype I tumors (Figure 6A). Afterwards, CIBERSORT was utilized to quantify the infiltration abundances of TIICs in GBM. Most CD4+ and CD8+ T cell subsets, NK cells, monocytes, macrophages, and neutrophils were more abundant in Stemness Subtype II (Figure 6B). However, only plasma cells, follicular helper

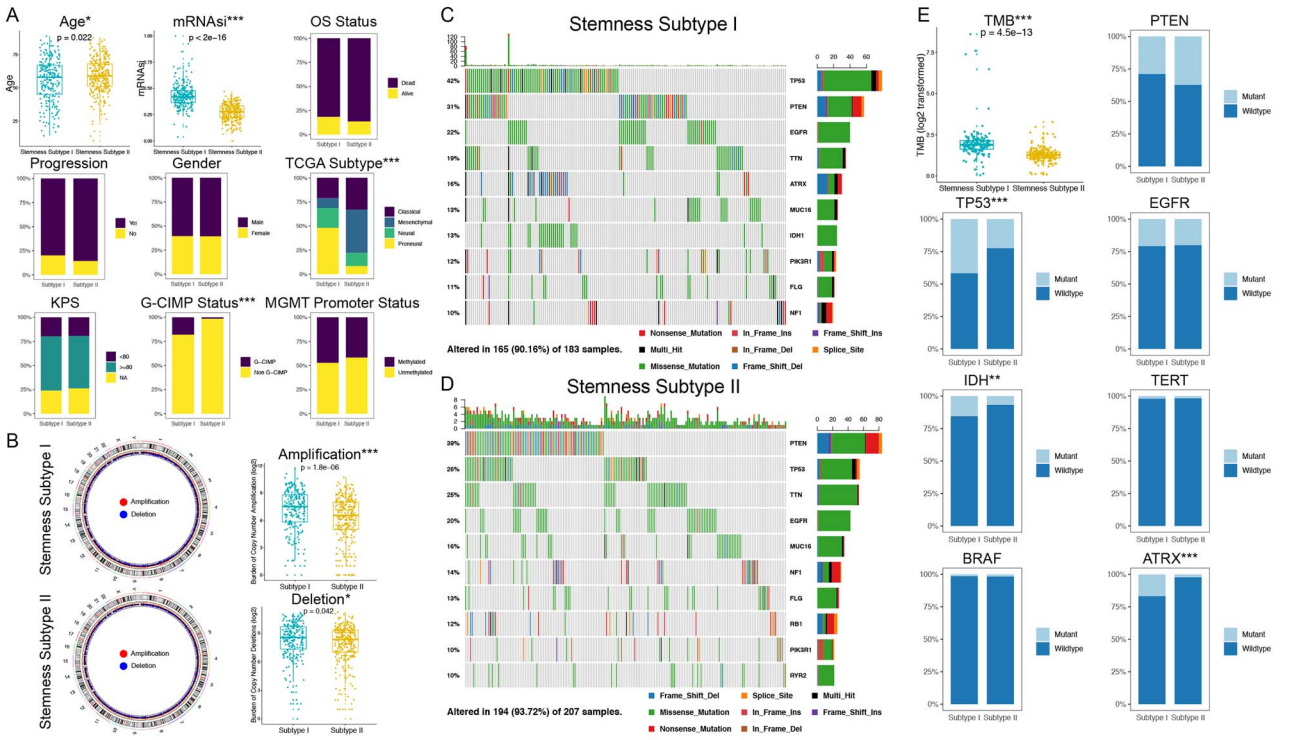

Figure 5. Comparisons of clinicalpathological and somatic variations between Stemness Subtype I and II. (**A**) Comparisons of age, mRNAsi, OS status, progression status, gender, TCGA subtype, KPS, G-CIMP status and *MGMT* promoter status between two subtypes. (**B**) Left panel: Circos plots of each stemness subtype revealing the amplifications and deletions of chromosomes, with red dots representing amplifications, blue dots representing deletions, and black dots representing no significant CNAs. Right panel: Boxplots inhibited more burdens of copy number amplifications and deletions in Stemness Subtype I. Waterfall plots showed the top 10 mutated in Stemness Subtype I (**C**) and II (**D**). (**E**) The comparisons of TMB, mutation status of *PTEN*, *TP53*, *EGFR*, *IDH*, *TERT* promoter, *BRAF* and *ATRX* between Stemness Subtype I and II. * means $P < 0.05$, ** means $P < 0.01$, and *** means $P < 0.001$.

T cells, resting dendritic cells, and resting mast cells were significantly more abundant in Stemness Subtype I (Figure 6B). Regarding the immune classification of GBM, Stemness Subtype I consisted of more proportions of high- and medium-immunity tumors, whereas Stemness Subtype II contained mainly low-immunity tumors ($P < 0.001$; Figure 6C). These interesting findings demonstrated that Stemness Subtype I tumors had relatively low immune infiltration levels and high tumor purity but also possessed relatively high immunity.

Stemness Subtype I was more sensitive to immunotherapy but resistant to TMZ

The expression levels of *PD1/PD-L1/PD-L2* and *CTLA/CD80/CD86* were exactly reversed in the two stemness subtypes (Figure 6D). The expression levels of *PD1* and its ligands (*PD-L1* and *PD-L2*) were significantly higher in Stemness Subtype I (all $P < 0.001$), and those of *CTLA* and its ligands (*CD80* and *CD86*) were significantly higher in Stemness Subtype II (all $P < 0.05$). Then, the TIDE algorithm was used to predict the likelihood of immunotherapy responses in GBM patients. As shown in Figure 6E, the proportion of responders to immunotherapy in the Stemness Subtype I group was more than two times that in the Stemness Subtype II group (44.6 versus 21.8%, $P < 0.001$). Regarding the TCGA molecular subtypes, the response rate of the proneural subtype was the highest (40.4%), followed by the neural (34.5%), classical (34.3%), and mesenchymal (21.1%) subtypes ($P = 0.004$). In addition, consistent with the findings of previous studies, the proportion of responders gradually decreased from high-immunity (81.0%) and medium-immunity (47.8%) to low-immunity (14.3%) tumors ($P < 0.001$). The stemness index was significantly higher in responders to immunotherapy than in nonresponders ($P < 0.001$) (Figure 6F). Then, subclass mapping analysis was used to predict the response to ICI therapy, including PD1 and CTLA4 inhibitors, of the two stemness subtypes. We found that patients in the Stemness Subtype I group were more sensitive to anti-PD1 therapy (FDR = 0.012), while those in the Stemness Subtype II group were more likely to respond to anti-CTLA4 therapy (FDR = 0.005) (Figure 6G).

Because TMZ treatment was the standard chemotherapy for GBM postoperatively, we evaluated the TMZ response of the two stemness subtypes by using the pRRophetic algorithm based on the GDSC database. By applying ridge regression analysis, we estimated the IC50 of TMZ in each GBM patient. The estimated IC50 values of TMZ were significantly lower in patients with Stemness Subtype II ($P = 4.34 \times 10^{-28}$; Figure 6H), indicating that GBM patients with Stemness Subtype I tended to be more resistant to TMZ therapy than those with Stemness Subtype II.

Identification of potential compounds targeting the two stemness subtypes

CMap analysis was carried out to explore the potential compounds targeting each stemness subtype of GBM. Forty-one highly expressed genes in Stemness Subtype I and 89 highly expressed genes in Stemness Subtype II were considered potential targets of compounds and were then employed in the CMap database. MoA analysis revealed 24 molecular pathways targeted by 34 compounds in Stemness Subtype I and 63 pathways targeted by 76 compounds in Stemness Subtype II (Supplementary

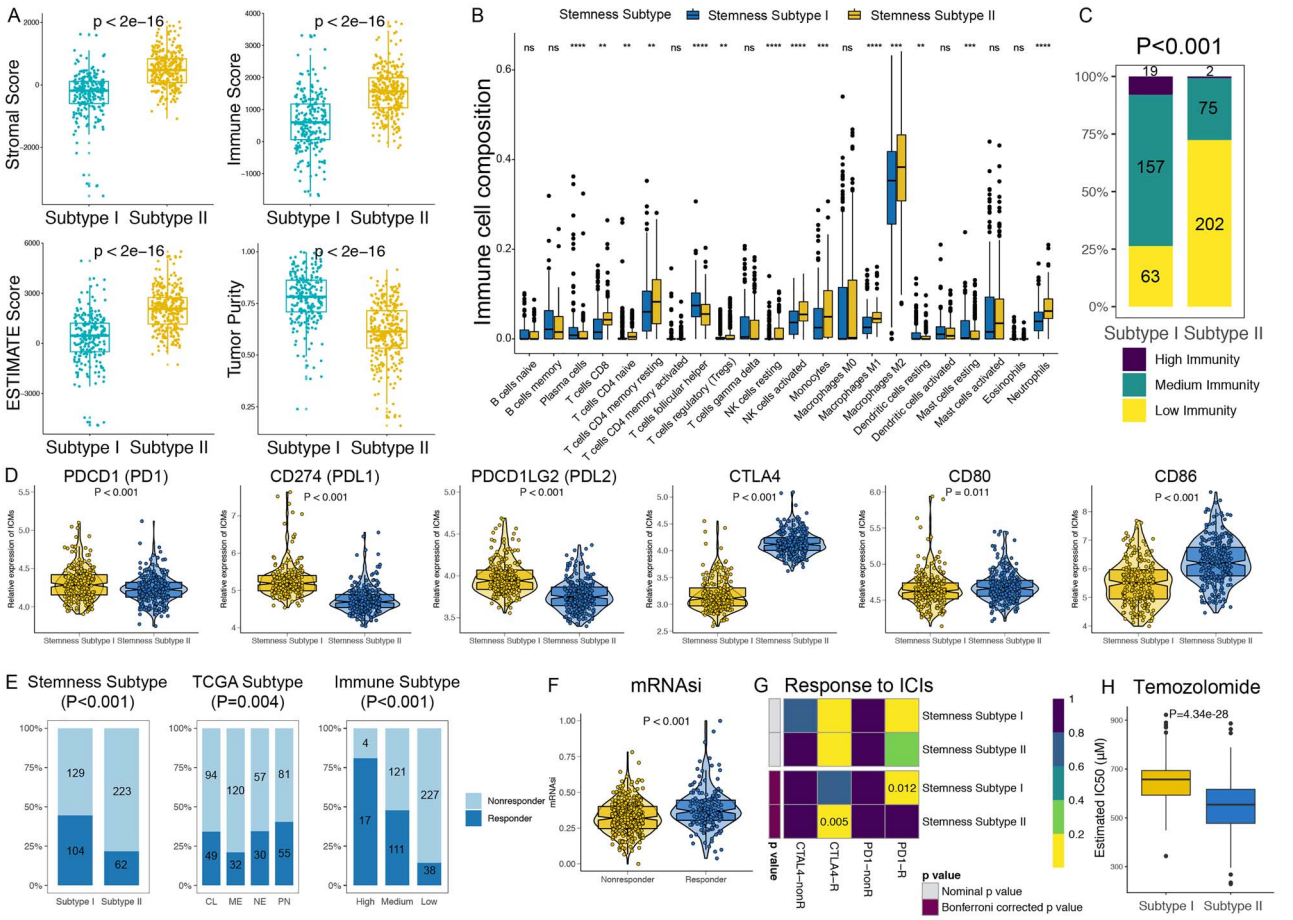

Figure 6. Distinct TIME and immunogenomic patterns of two stemness subtypes lead to different sensitivity to immunotherapy and TMZ. (**A**) Comparisons of stromal score, immune score, ESTIMATE score and tumor purity between Stemness Subtype I and II. (**B**) Comparisons of the abundances of 22 immune cells in two subtypes. (**C**) The different proportions of high, medium and low immunity tumors in two stemness subtypes. (**D**) The expression levels of *PD-1*, *PD-L1*, *PD-L2*, *CTLA-4*, *CD80* and *CD86* in Stemness Subtype I and II. (**E**) Comparisons of the proportions of nonresponders and responders to immunotherapy among different classification methods, including stemness subtypes (left panel), TCGA subtypes (middle panel) and immune subtypes (right panel). (**F**) Violin plot showed significantly higher stemness index in responders. (**G**) Subclass mapping analysis for predicting the likelihood of response to ICI therapy of different stemness subtypes. (**H**) Boxplot demonstrated the difference in the IC50 value estimated by pRRophetic algorithm based on GDSC database between Stemness Subtype I and II.

Figure 3). Regarding the most critical MoAs for each subtype, there were 12 compounds sharing the same MoA as topoisomerase inhibitors in Stemness Subtype I and another 12 compounds sharing the same MoA as monoamine oxidase inhibitor in Stemness Subtype II. Further studies are needed to verify the therapeutic value of these compounds by inhibiting or reversing the corresponding MoAs.

Construction and validation of the Stemness Subtype Predictor

First, in the training set, four machine learning algorithms were employed to determine the most critical stemness subtype-relevant features via the expression levels of 130 stemness-based DEGs. A total of 43, 112, 30 and 71 genes were identified by LASSO, RFB, SVM, and XGBoost analyses, respectively (Supplementary Figure 4). ROC curve analysis demonstrated that the four machine learning algorithms had excellent performance in feature selection, with AUCs of 1.00 in the training set and AUCs >0.95 in the testing set (Figure 7A). The Venn diagram identified seven critical genes that were shared by the four feature selection algorithms (Figure 7B). Afterwards, multivariate logistic regression analysis was performed to construct the diagnostic predictive model (Table 4). The formula of the Stemness Subtype Predictor was as follows: Stemness Subtype Predictor = 0.523 + 1.794 × (expression of *LTF*) + 1.410 × (expression of *TAGLN*) + 1.027 × (expression of *C5AR1*) - 2.148 × (expression of *RAB33A*) + 1.581 × (expression of *CFI*) + 1.392 × (expression of *CH25H*) + 1.964 × (expression of *RNASE2*). The optimal cutoff value for discrimination was 0.0565, suggesting that patients with a score < 0.0565 were considered to have Stemness Subtype I and those with a score > 0.0565 were considered to have Stemness Subtype II. ROC analysis demonstrated an AUC of 0.9889 for distinguishing Stemness Subtypes I and II, with a sensitivity of 94.86%, a specificity of 97.51%, and an accuracy of 96.28% in the training set (Figure 7C). In addition, the Stemness Subtype Predictor also had an excellent performance in discriminating the stemness subtypes as evaluated in the test set, with an AUC of 0.9599, a sensitivity of 91.38%, a specificity of 94.05%, and an accuracy of 92.96% (Figure 7C).

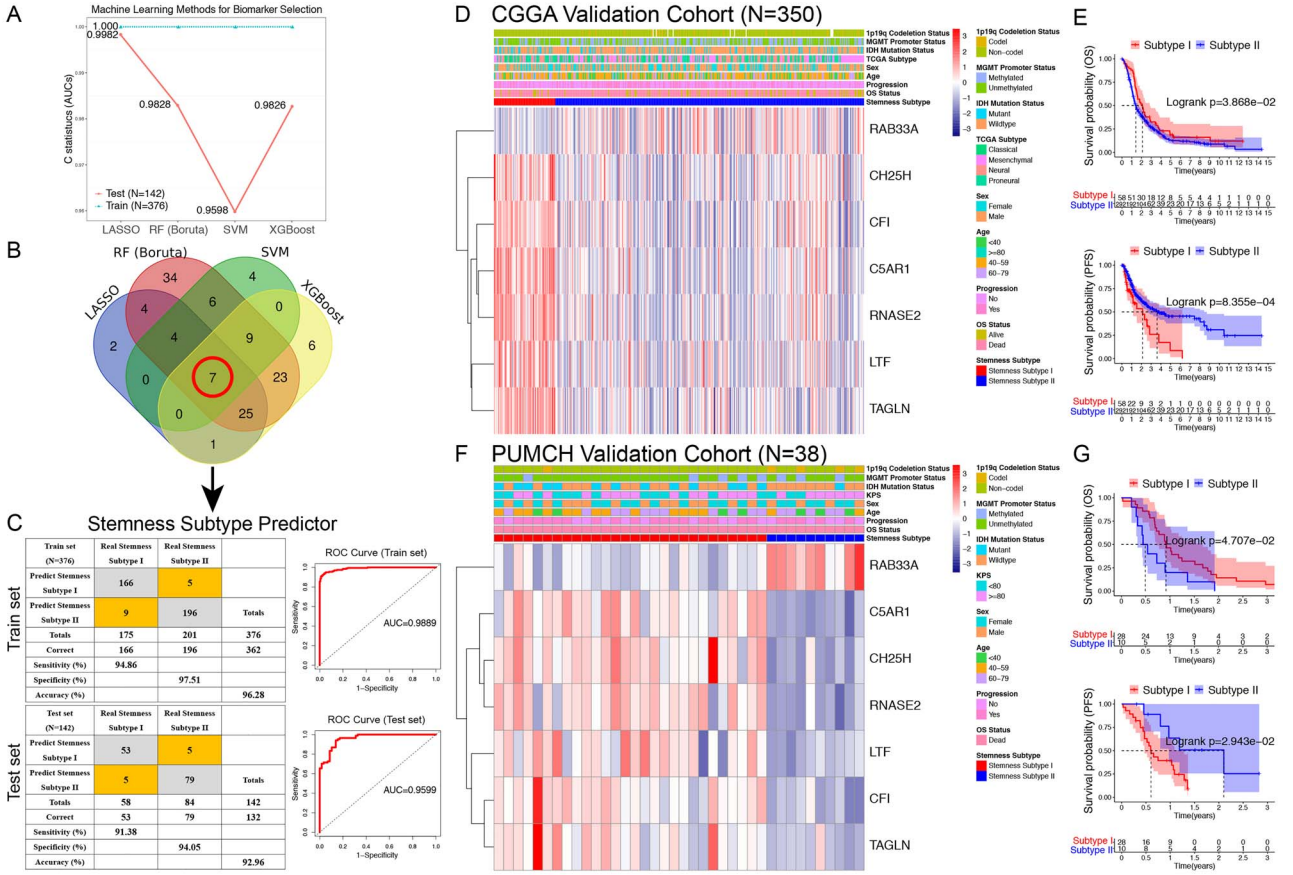

Figure 7. Construction and validation of the Stemness Subtype Predictor. (**A**) The performances of four machine-learning algorithms (LASSO, RFB, SVM and XGBoost) for feature selection were, respectively, evaluated in the train set and test set. AUCs were generated by ROC analysis. (**B**) Venn diagram identified seven most critical stemness subtype specific genes that were shared by four feature selection algorithms. (**C**) Left panel: confusion matrices of binary results of the Stemness Subtype Predictor for the train set (upper) and test set (lower). Right panel: ROC curves of the Stemness Subtype Predictor in distinguishing two subtypes in the train set (Upper) and test set (Lower). The CGGA RNA sequencing data (**D**) and the PUMCH qRT-PCR data (**F**) were enrolled to further validated the clinical application value of the stemness-based classification, which were visualized by heatmaps. Kaplan–Meier survival analysis also suggested that patients' OS and PFS were significantly different between Stemness Subtype I and II both in both CGGA (**E**) and PUMCH cohort (**G**).

Table 4. Multivariate logistic regression analysis of the seven genes selected by multiple machine learning algorithms

Variables	β	OR (95%CI)	*P*-value
Stemness subtype predictor			
Intercept	0.523	–	3.64e-02
LTF	1.794	6.013 (3.173–12.872)	3.94e-07
TAGLN	1.410	4.095 (2.235–8.25)	1.89e-05
C5AR1	1.027	2.794 (1.263–6.481)	1.26e-02
RAB33A	−2.148	0.117 (0.049–0.236)	6.63e-08
CFI	1.581	4.860 (2.488–10.509)	1.41e-05
CH25H	1.392	4.022 (2.159–8.167)	3.39e-05
RNASE2	1.964	7.126 (2.859–21.489)	1.14e-04

Abbreviations: β, regression coefficient; OR, odds ratio; 95%CI, 95% confidence interval.
Bold type of *P*-value means $P < 0.05$.

Clinical applications of the stemness-based classification in two independent cohorts

We enrolled another two independent GBM cohorts, bulk RNA sequencing data from patients in the CGGA database and qRT-PCR data from patients at PUMCH, to explore the clinical application value of the novel stemness-based classification in different patient populations. By applying the Stemness Subtype Predictor, we calculated the score for each patient and then categorized the patients into two groups based on the cutoff value of the diagnostic predictor and the expression profiles of seven critical genes. The CGGA cohort consisted of 58 patients with Stemness Subtype I (16.6%) and 292 patients with Stemness Subtype II (83.4%) (Figure 7D), whereas the PUMCH cohort contained 28 patients with Stemness Subtype I (73.7%) and 10 patients with Stemness Subtype II (26.3%) (Figure 7F). K-M survival analysis indicated that patients with Stemness Subtype I presented significantly better OS and poorer PFS in both the CGGA and PUMCH GBM cohorts (Figure 7E and G), which was consistent with the previous results in the TCGA cohort. In addition, univariate and subsequent multivariate Cox regression analyses revealed significant correlations between the stemness subtype and OS/PFS in both the CGGA and PUMCH cohorts (Tables 2 and 3), indicating that the stemness subtype was also an independent predictor for the prognosis of GBM patients.

In addition, the demographic and clinicopathological features of GBM patients with Stemness Subtypes I and II in the two independent GBM cohorts were compared (Table 1). As shown in Supplementary Figures 5A and 6A, patients in the Stemness Subtype I group were significantly younger than those in the Stemness Subtype II group ($P < 0.05$), and the Stemness Subtype I group had more patients with 1p19q codeletion ($P < 0.05$) in both cohorts. The proportions of TCGA molecular subtypes also differed between Stemness Subtypes I and II in the CGGA cohort ($P = 0.013$). Regarding the most common biomarkers of GBM, the Stemness Subtype I group had a higher proportion of patients with IDH mutations ($P < 0.05$) in the two cohorts, and more patients with TP53 mutations were also observed in the Stemness Subtype I group in the PUMCH cohort (57.1% versus 20.0%, $P = 0.043$). However, the mutation frequencies of *BRAF*, *PTEN*, *EGFR*, *TERT*, and *ATRX* did not differ significantly between the two subtypes in the PUMCH cohort. TMB was not compared due to a lack of whole genome sequencing or whole exome sequencing data.

TIME patterns were also evaluated in the CGGA cohort. Similar to the findings from TCGA, the immune, stromal, and ESTIMATE scores were significantly lower and the tumor purity score was significantly higher in the Stemness Subtype I group (all $P < 0.001$), demonstrating low abundances of immune and stromal cells and high tumor purity in Stemness Subtype I tumors (Supplementary Figure 5B). CIBERSORT further revealed that follicular helper T cells, resting dendritic cells, activated dendritic cells, and activated mast cells were significantly more abundant in Stemness Subtype I, whereas most CD4+ and CD8+ T cell subsets, NK cells, monocytes, macrophages, and neutrophils were more abundant in Stemness Subtype II (Supplementary Figure 5C).

Subsequently, the expression levels of immune checkpoint molecules were investigated in the two cohorts, which also demonstrated significant reverse expression patterns of *PD1/PD-L1/PD-L2* and *CTLA/CD80/CD86* in Stemness Subtypes I and II (Supplementary Figures 5D and 6C). Then, the TIDE algorithm was utilized to predict the likelihood of immunotherapy responses in CGGA GBM patients, which indicated that the proportion of responders to immunotherapy in the Stemness Subtype I group (36.2%) was significantly higher than that in the Stemness Subtype II group (18.2%) in the CGGA cohort ($P < 0.001$) (Supplementary Figure 5E). Subclass mapping analysis also revealed that patients with Stemness Subtype I were more sensitive to anti-PD1 therapy (FDR = 0.003), and those with Stemness Subtype II were more likely to respond to anti-CTLA4 therapy (FDR = 0.033) (Supplementary Figure 5F). Finally, we also found that CGGA patients with Stemness Subtype I were more sensitive to TMZ therapy due to their relatively lower estimated IC50 values of TMZ ($P = 8.15 \times 10^{-47}$; Supplementary Figure 5G). The response to immunotherapy and TMZ therapy in the PUMCH cohort could not be predicted due to the lack of transcriptomic data of PUMCH patients. Generally, all the above findings indicated that this novel stemness-based molecular classification of GBM patients was robust and reliable and can be clinically applied in different patient populations.

Discussion

This study conducted an in-depth analysis of the relationship between the stemness of GBM and the effects of immunotherapy and chemotherapy, proposed an approach to distinguish subtypes according to stemness, and verified this approach using patient samples and clinical data from two different cohorts. First, we used the OCLR algorithm to calculate the stemness index, mRNAsi, in 518 GBM patients from the TCGA database. Then, ESTIMATE was used to evaluate the tumor microenvironment, tumor purity and the abundance of stromal cells and immune cells in GBM patients. After analyzing the interaction between stemness and TIME, CIBERSORT was used to analyze the composition of TIICs. We divided the GBM patients into two subtypes based on their stemness index and compared their clinicopathological parameters to clarify the correlation between the stemness subtypes and clinical features. In Stemness Subtype I patients with high mRNAsi scores, immunotherapy, especially anti-PD-1/PD-L1 treatment, can achieve better therapeutic effects, as predicted by the TIDE algorithm, while TMZ will have worse therapeutic effects, as predicted by the pRRophetic package. Moreover, this group of patients have better OS and poorer PFS. To make the results more practical, we found potential compounds targeting genes related to the stemness subtype using CMap analysis, thus laying the foundation for research on treatment. Moreover, to facilitate the distinction between these two stemness subtypes in clinical practice, the seven most critical stemness subtype-related genes were identified and defined as stemness subtype predictors by using the LASSO, SVM, RFB and XGBoost machine learning methods. These 7 predictors were verified in the CGGA and PUMCH cohort.

Although the CNS is a relatively immunologically privileged site, immunotherapy has been widely studied in GBM patients in recent years, mainly including immune checkpoint inhibitors, antitumor vaccines and cellular immunotherapy. However, compared to other tumors, the results of these phase III clinical trials in GBM were not satisfactory [15, 57]. There are many factors affecting the effect of immunotherapy in GBM. Taking the PD-1/PD-L1 blockade treatment as an example, the expression level of *PD-L1*, TMB, mismatch repair deficiency and tumor-infiltrating lymphocytes were all able to affect the anti-PD-1/PD-L1 treatment [17, 58]. However, in the current phase III clinical trials, GBM patients have not been screened for the above factors, and such indiscriminate treatment may also be the reason for

the failure of these trials, which is also an issue that needs to be paid attention to in future researches.

Based on the above problems in immunotherapy, our study proposed a new GBM classification based on tumor stemness. We found that patients with Stemness Subtype I with a high mRNAsi score would have a better response to immunotherapy, which provided a new idea for screening patients. At present, the interaction between CSCs and the immune system has not been well studied because most of the animal models used in studies on CSCs lack major immune cell components, but there is still some evidence that CSCs may have a regulatory effect on the immune system in GBM patients [59]. In coculture studies, CSCs can induce the production of regulatory T (Treg) cells through PD-1 and induce the apoptosis of cytotoxic T cells to inhibit their tumor killing effect [60]. CSCs may also induce the production of a more immunosuppressive phenotype (M2) of tumor-associated macrophages by secreting interleukin 10 (IL-10) and tumor necrosis factor β (TNF-β) [61]. In addition, antitumor vaccines are related to CSCs, as the lysed products of CSCs are more effective as antigens of antitumor vaccines than differentiated tumor cells [62, 63]. All the above studies have indicated the correlation between CSCs and immunotherapy. Therefore, anti-CSC therapy combined with immunotherapy may be able to achieve a better therapeutic effect in GBM patients.

Potential anti-CSC compounds were analyzed from CMap in this study. In the case of Stemness Subtype I, amitriptyline was identified as one of the potential targeted drugs. Amitriptyline is a tricyclic antidepressant approved by FDA for the treatment of depression by blocking serotonin and norepinephrine and the reuptake of them. Recently, amitriptyline is also proposed to be used as an anti-tumor drug by affecting the neurotrophin receptor, Fas death receptor and c-Jun pathway to promote cell apoptosis [64–66]. Besides, it can affect the production of reactive oxygen species by affecting the synthesis of coenzyme Q10. This result has been verified in lung cancer cell lines [67]. Other tricyclic antidepressants were confirmed to induce autophagic cell death by affecting PI3K/AKT/mTOR signaling pathway in GBM cell line U-87MG [68]. In another GBM cell line T98G, the tricyclic antidepressant was proved to transform GBM stem cells into non-stem cells, thereby reducing the malignancy of GBM [69]. Although there is still a lack of research on this drug and immunotherapy at present, the above researches demonstrate the effectiveness of drug screening in our study and the feasibility of drug application in the treatment of GBM. In addition to amitriptyline, this study screened a variety of targeted drugs for the two stemness subtypes, which may provide a basis for future studies on combination therapy.

Our study also found that although immunotherapy had a significantly better effect on patients with Stemness Subtype I, the therapeutic effect of TMZ on this group of patients was not satisfactory, suggesting that CSCs in GBM were also related to TMZ and that TMZ may have some regulatory relationship with the immune system. An important reason for the current failure of TMZ treatment for GBM lies in its high heterogeneity, while the strong differentiation potential of CSCs is undoubtedly the greatest factor causing tumor heterogeneity [70]. In addition, studies on the causes of CSCs and radiotherapy resistance suggest that CSCs may enhance the ability of DNA damage repair to cause radiotherapy resistance [71], and this effect is also likely to play an important role in TMZ resistance. Many molecules, such as *Notch*, nuclear factor kappa B (*NF-κB*), enhancer of zeste homolog 2 (*EZH2*), and poly (ADP-ribose) polymerase (*PARP*) [72–75], may also be involved in the effect of CSCs on cytotoxic agents, suggesting that the combination of targeted therapy, immunotherapy, and anti-CSC therapy may provide survival benefits for GBM patients in the future. Aida Karachi et al. summarized the immunomodulatory effect of TMZ in GBM, clarifying its effect on cytotoxic T cells and antigen-presenting cells [76]. Meanwhile, the existence of lymphocytosis and increasing proportion of Treg cells caused by TMZ also promoted the formation of an immunosuppressant environment in GBM.

There are also some limitations of this research. First, when validating the results of this study, we only included 38 patients with complete clinical and pathological data from PUMCH, which was a relatively small sample size. Although there were enough CGGA samples as the validation set to support the conclusions of this research, we need to further expand the sample size from our own center in the future. Second, as patients receiving immunotherapy are currently very limited and the conclusions of this study were based on the analysis of transcriptomic data from public databases, the relationship between stemness subtypes and immunotherapy responsiveness needs to be validated in an immunotherapy cohort in the future.

In conclusion, immunotherapy still has great potential in GBM, and screening patients who are likely to benefit from immunotherapy is one of the important tasks at present. In this study, the patients were divided into two different subtype groups based on the stemness of GBM, and the response of patients with different subtypes to immunotherapy was predicted, providing a potential approach for the screening of patients for immunotherapy in the future. The stemness subtype predictors were also identified in this study, which made subgrouping based on stemness clinically feasible.

Key Points

- GBM is the most malignant and lethal primary intracranial tumor with extremely limited treatment options.
- Stemness plays a critical role in the initiation and treatment resistance of GBM and also exhibits a substantial impact on patients' response to immunotherapy.
- Integrated multiomic analysis revealed that based on distinct stemness features, GBM patients could be divided into two subtypes with reverse survival outcomes and tumor microenvironment, as well as different response to immunotherapy and temozolomide therapy.
- The novel stemness-based classification could provide promising prognostic predictors for GBM and may guide physicians in selecting potential immunotherapy responders for preferential use of immune checkpoint inhibitors.

Supplementary data

Supplementary data are available online in *Briefings in Bioinformatics*.

Authors' contributions

All authors designed and conducted this article. All authors read and approved the final manuscript. Zihao Wang and Yaning Wang analyzed and interpreted the results and drafted the manuscript. Yuekun Wang, Lu Gao and Xiaopeng

Guo performed the data curation and analysis. Tianrui Yang and Hao Xing compiled the clinical data from PUMCH. Yu Wang, Wenbin Ma and Bing Xing reviewed the manuscript. Notably, Zihao Wang and Yaning Wang equally share the first authorship; Yu Wang and Wenbin Ma equally share the corresponding authorship.

Data availability statement

The data underlying this article are available in the article and in its online supplementary material.

Funding

The Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grant number: 2016-I2M-2-001); the Beijing Municipal Natural Science Foundation (grant number: 7202150 and 19JCZDJC64200(Z)); Tsinghua University-Peking Union Medical College Hospital Initiative Scientific Research Program (grant number: 2019ZLH101); the Graduate Innovation Fund of the Chinese Academy of Medical Sciences and Peking Union Medical College (grant number: 2019-1002-73).

References

1. Wen PY, Huse JT. 2016 World Health Organization classification of central nervous system Tumors. *Continuum (Minneap Minn)* 2017;**23**:1531–47.
2. Jiang T, Mao Y, Ma W, *et al*. CGCG clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett* 2016;**375**:263–73.
3. Stupp R, Taillibert S, Kanner A, *et al*. Effect of tumor-treating fields plus maintenance Temozolomide vs maintenance Temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA* 2017;**318**:2306–16.
4. Gilbert MR, Dignam JJ, Armstrong TS, *et al*. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* 2014;**370**:699–708.
5. Sandmann T, Bourgon R, Garcia J, *et al*. Patients with proneural glioblastoma may derive overall survival benefit from the addition of bevacizumab to first-line radiotherapy and Temozolomide: retrospective analysis of the AVAglio trial. *J Clin Oncol* 2015;**33**:2735–44.
6. Malta TM, Sokolov A, Gentles AJ, *et al*. Machine learning identifies Stemness features associated with oncogenic dedifferentiation. *Cell* 2018;**173**:338–354.e315.
7. Bjerkvig R, Tysnes BB, Aboody KS, *et al*. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat Rev Cancer* 2005;**5**:899–904.
8. Shibue T, Weinberg RAEMT. CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol* 2017;**14**:611–29.
9. Ge Y, Gomez NC, Adam RC, *et al*. Stem cell lineage infidelity drives wound repair and cancer. *Cell* 2017;**169**:636–650.e614.
10. Esparza R, Azad TD, Feroze AH, *et al*. Glioblastoma stem cells and stem cell-targeting immunotherapies. *J Neurooncol* 2015;**123**:449–57.
11. Rizvi NA, Hellmann MD, Snyder A, *et al*. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015;**348**:124–8.
12. Snyder A, Makarov V, Merghoub T, *et al*. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 2014;**371**:2189–99.
13. Watkins S, Robel S, Kimbrough IF, *et al*. Disruption of astrocyte-vascular coupling and the blood-brain barrier by invading glioma cells. *Nat Commun* 2014;**5**:4196.
14. Louveau A, Smirnov I, Keyes TJ, *et al*. Structural and functional features of central nervous system lymphatic vessels. *Nature* 2015;**523**:337–41.
15. Reardon DA, Brandes AA, Omuro A, *et al*. Effect of Nivolumab vs bevacizumab in patients with recurrent glioblastoma: the CheckMate 143 phase 3 randomized clinical trial. *JAMA Oncol* 2020;**6**:1003–10.
16. Yang T, Kong Z, Ma WPD. 1/PD-L1 immune checkpoint inhibitors in glioblastoma: clinical studies, challenges and potential. *Hum Vaccin Immunother* 2021;**17**:546–553.
17. Touat M, Li YY, Boynton AN, *et al*. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* 2020;**580**:517–23.
18. Bouffet E, Larouche V, Campbell BB, *et al*. Immune checkpoint inhibition for Hypermutant glioblastoma Multiforme resulting from germline Biallelic mismatch repair deficiency. *J Clin Oncol* 2016;**34**:2206–11.
19. Zhao S, Fung-Leung WP, Bittner A, *et al*. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 2014;**9**:e78644.
20. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 2005;**21**:2067–75.
21. Mayakonda A, Lin DC, Assenov Y, *et al*. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018;**28**:1747–56.
22. Skidmore ZL, Wagner AH, Lesurf R, *et al*. GenVisR: genomic visualizations in R. *Bioinformatics* 2016;**32**:3012–4.
23. Budczies J, Seidel A, Christopoulos P, *et al*. Integrated analysis of the immunological and genetic status in and across cancer types: impact of mutational signatures beyond tumor mutational burden. *Onco Targets Ther* 2018;**7**:e1526613.
24. Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;**12**:R41.
25. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* 2013;**14**:244.
26. Sokolov A, Carlin DE, Paull EO, *et al*. Pathway-based genomics prediction using generalized elastic net. *PLoS Comput Biol* 2016;**12**:e1004790.
27. Yoshihara K, Shahmoradgoli M, Martínez E, *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;**4**:2612.
28. Newman AM, Liu CL, Green MR, *et al*. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;**12**:453–7.
29. He Y, Jiang Z, Chen C, *et al*. Classification of triple-negative breast cancers based on Immunogenomic profiling. *J Exp Clin Cancer Res* 2018;**37**:327.
30. Hänzelmann S, Castelo R, Guinney JGSVA. Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;**14**:7.
31. Zhang C, Chen T, Li Z, *et al*. Depiction of tumor stemlike features and underlying relationships with hazard immune infiltrations based on large prostate cancer cohorts. *Brief Bioinform* 2020.

32. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 2000;**25**:25–9.
33. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–d361.
34. Wang J, Vasaikar S, Shi Z, *et al*. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 2017;**45**:W130–w137.
35. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;**26**:1572–3.
36. Şenbabaoğlu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. *Sci Rep* 2014;**4**:6207.
37. Wang Z, Guo X, Gao L, *et al*. Glioblastoma cell differentiation trajectory predicts the immunotherapy response and overall survival of patients. *Aging (Albany NY)* 2020;**12**:18297–321.
38. Yang C, Huang X, Liu Z, *et al*. Metabolism-associated molecular classification of hepatocellular carcinoma. *Mol Oncol* 2020;**14**:896–913.
39. Jiang P, Gu S, Pan D, *et al*. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;**24**:1550–8.
40. Hoshida Y, Brunet JP, Tamayo P, *et al*. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One* 2007;**2**:e1195.
41. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014;**9**:e107468.
42. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol* 2014;**15**:R47.
43. Subramanian A, Narayan R, Corsello SM, *et al*. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e1417.
44. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 2007;**26**:5512–28.
45. Kim S. Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms. *Int J Data Min Bioinform* 2014;**10**:374–90.
46. Yperman J, Becker T, Valkenborg D, *et al*. Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. *BMC Neurol* 2020;**20**:105.
47. Li W, Yin Y, Quan X, *et al*. Gene expression value prediction based on XGBoost algorithm. *Front Genet* 2019;**10**:1077.
48. Xu RH, Wei W, Krawczyk M, *et al*. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater* 2017;**16**:1155–61.
49. Li G, Xu W, Zhang L, *et al*. Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma. *EBioMedicine* 2019;**47**:128–41.
50. Horbinski C, Kofler J, Kelly LM, *et al*. Diagnostic use of IDH1/2 mutation analysis in routine clinical testing of formalin-fixed, paraffin-embedded glioma tissues. *J Neuropathol Exp Neurol* 2009;**68**:1319–25.
51. Chan AK, Yao Y, Zhang Z, *et al*. TERT promoter mutations contribute to subset prognostication of lower-grade gliomas. *Mod Pathol* 2015;**28**:177–86.
52. Zacher A, Kaulich K, Stepanow S, *et al*. Molecular diagnostics of gliomas using next generation sequencing of a glioma-tailored gene panel. *Brain Pathol* 2017;**27**:146–59.
53. Snuderl M, Eichler AF, Ligon KL, *et al*. Polysomy for chromosomes 1 and 19 predicts earlier recurrence in anaplastic oligodendrogliomas with concurrent 1p/19q loss. *Clin Cancer Res* 2009;**15**:6430–7.
54. Reifenberger G, Hentschel B, Felsberg J, *et al*. Predictive impact of MGMT promoter methylation in glioblastoma of the elderly. *Int J Cancer* 2012;**131**:1342–50.
55. Thorsson V, Gibbs DL, Brown SD, *et al*. The immune landscape of cancer. *Immunity* 2018;**48**:812–830.e814.
56. Rooney MS, Shukla SA, Wu CJ, *et al*. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015;**160**:48–61.
57. Weller M, Butowski N, Tran DD, *et al*. Rindopepimut with temozolomide for patients with newly diagnosed, EGFRvIII-expressing glioblastoma (ACT IV): a randomised, double-blind. *international phase 3 trial, Lancet Oncol* 2017;**18**:1373–85.
58. Wang X, Guo G, Guan H, *et al*. Challenges and potential of PD-1/PD-L1 checkpoint blockade immunotherapy for glioblastoma. *J Exp Clin Cancer Res* 2019;**38**:87.
59. Lathia JD, Mack SC, Mulkearns-Hubert EE, *et al*. Cancer stem cells in glioblastoma. *Genes Dev* 2015;**29**:1203–17.
60. Di Tomaso T, Mazzoleni S, Wang E, *et al*. Immunobiological characterization of cancer stem cells isolated from glioblastoma patients. *Clin Cancer Res* 2010;**16**:800–13.
61. Wu A, Wei J, Kong LY, *et al*. Glioma cancer stem cells induce immunosuppressive macrophages/microglia. *Neuro Oncol* 2010;**12**:1113–25.
62. Pellegatta S, Poliani PL, Corno D, *et al*. Neurospheres enriched in cancer stem-like cells are highly effective in eliciting a dendritic cell-mediated immune response against malignant gliomas. *Cancer Res* 2006;**66**:10247–52.
63. Xu Q, Liu G, Yuan X, *et al*. Antigen-specific T-cell response from dendritic cell vaccination using cancer stem-like cell-associated antigens. *Stem Cells* 2009;**27**:1734–40.
64. Pula G, Pistilli A, Montagnoli C, *et al*. The tricyclic antidepressant amitriptyline is cytotoxic to HTB114 human leiomyosarcoma and induces p75(NTR)-dependent apoptosis. *Anticancer Drugs* 2013;**24**:899–910.
65. Jahchan NS, Dudley JT, Mazur PK, *et al*. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov* 2013;**3**:1364–77.
66. Yuan SY, Cheng CL, Ho HC, *et al*. Nortriptyline induces mitochondria and death receptor-mediated apoptosis in bladder cancer cells and inhibits bladder tumor growth in vivo. *Eur J Pharmacol* 2015;**761**:309–20.
67. Ortiz T, Villanueva-Paz M, Díaz-Parrado E, *et al*. Amitriptyline down-regulates coenzyme Q(10) biosynthesis in lung cancer cells. *Eur J Pharmacol* 2017;**797**:75–82.
68. Jeon SH, Kim SH, Kim Y, *et al*. The tricyclic antidepressant imipramine induces autophagic cell death in U-87MG glioma cells. *Biochem Biophys Res Commun* 2011;**413**:311–7.
69. Bielecka-Wajdman AM, Lesiak M, Ludyga T, *et al*. Reversing glioma malignancy: a new look at the role of antidepressant drugs as adjuvant therapy for glioblastoma multiforme. *Cancer Chemother Pharmacol* 2017;**79**:1249–56.
70. Chen J, Li Y, Yu TS, *et al*. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* 2012;**488**:522–6.
71. Bao S, Wu Q, McLendon RE, *et al*. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* 2006;**444**:756–60.

72. Wang J, Wakeman TP, Lathia JD, *et al*. Notch promotes radioresistance of glioma stem cells. *Stem Cells* 2010;**28**:17–28.
73. Bhat KPL, Balasubramaniyan V, Vaillant B, *et al*. Mesenchymal differentiation mediated by NF-κB promotes radiation resistance in glioblastoma. *Cancer Cell* 2013;**24**: 331–46.
74. Venere M, Hamerlik P, Wu Q, *et al*. Therapeutic targeting of constitutive PARP activation compromises stem cell phenotype and survival of glioblastoma-initiating cells. *Cell Death Differ* 2014;**21**:258–69.
75. Kim SH, Joshi K, Ezhilarasan R, *et al*. EZH2 protects glioma stem cells from radiation-induced cell death in a MELK/FOXM1-dependent manner. *Stem Cell Reports* 2015;**4**:226–38.
76. Karachi A, Dastmalchi F, Mitchell DA, *et al*. Temozolomide for immunomodulation in the treatment of glioblastoma. *Neuro Oncol* 2018;**20**:1566–72.