

Glioblastoma cell differentiation trajectory predicts the immunotherapy response and overall survival of patients

Zihao Wang^{1,*}, Xiaopeng Guo^{1,*}, Lu Gao¹, Yu Wang¹, Wenbin Ma¹, Bing Xing¹

¹Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, P. R. China

*Equal contribution

Correspondence to: Bing Xing; email: xingbingemail@aliyun.com, <https://orcid.org/0000-0002-3864-5168>

Keywords: glioblastoma, cancer cell differentiation, trajectory analysis, overall survival, immunotherapy response

Received: May 9, 2020

Accepted: June 25, 2020

Published: September 21, 2020

Copyright: © 2020 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Glioblastoma (GBM) is the most common and lethal primary brain tumor. In this study, we aimed to investigate the differentiation states of GBM cells and their clinical relevance. Integrated single-cell RNA-seq (scRNA-seq) data and bulk RNA-seq data from GBM samples were used for analysis. Two subsets of GBM cells in distinct differentiation states were characterized, and 498 GBM cell differentiation-related genes (GDRGs) were identified. GDRGs were significantly correlated with immune regulation and metabolic pathways. We classified the GBM patients into two groups based on the expression of GDRGs in tumors and found that the cell differentiation-based classification successfully predicted patient overall survival (OS), immune checkpoint expression and likelihood of immunotherapy response in GBMs. *FN1*, *APOE*, *RPL7A* and *GSTM2* were the 4 most significant survival-predicting GDRGs, and patients with different expression levels of each of these genes had distinct survival outcomes. Finally, a nomogram composed of the GDRG signature, age, pharmacotherapy, radiotherapy, IDH mutations and MGMT promoter methylation was generated and validated in two large GBM cohorts to predict GBM prognosis. This study highlights the significant roles of cell differentiation in predicting the clinical outcomes of GBM patients and their potential response to immunotherapy, suggesting promising therapeutic targets for GBM.

INTRODUCTION

Glioblastoma (GBM) is the most common primary malignant brain tumor in adults, comprising 14.6% of all tumors and 48.3% of malignant tumors of the central nervous system (CNS) [1]. Due to the highly aggressive nature of GBM, neurological deficits, seizures and symptoms of intracranial hypertension occur rapidly in GBM patients within days or months [2]. Despite considerable advances in the development of treatments, including surgical resection, radiotherapy, and chemotherapy, little progress toward prolonged survival and better prognosis has been achieved over the last few decades [3]. The modest median overall survival (OS) time is approximately 14 months, and only 5% to 6.8% of GBM patients survive 5 years after

diagnosis [1, 4]. Multiple clinical trials, including those on immunotherapy, have been conducted for GBM patients, but the results have been largely disappointing [5]. Prognosis predictors of GBM patients have been studied, and patient age, the extent of tumor resection, and several molecular alterations were reported as promising predictors [6–8]. Since autophagy plays a vital role in GBM development and progression, in our previous study, we generated a risk score nomogram based on 3 autophagy-related genes, *NRG1*, *ITGA3*, and *MAP1LC3A*, to predict the survival of GBM patients [9]. However, biomarkers and predictors for patient outcome and the immunotherapy response of GBMs have not been fully elucidated, and existing predictive models are far from satisfactory.

Multiple factors in the tumor microenvironment influence cancer cells during their differentiation from cancer stem cells (CSCs), leading to heterogeneous cell differentiation states and cell fates [10, 11]. Single-cell transcriptomics analysis has recently emerged as a powerful method to provide opportunities to characterize cell states and their transitions by simultaneously investigating the comprehensive nature of the genomes of an entire tumor sample at microscopic resolution [12]. Ordering such comprehensive tumor-constituting cells into trajectories helps us understand tumor cell subsets based on differentiation states and unveils the genetic cascades and related tumorigenic pathways accompanying cell fate specification [13]. Monocle 2, a recently generated algorithm that uses a reversed graph embedding strategy, is capable of accurately reconstructing single-cell trajectories according to the features of cell differentiation [14]. Therefore, by combining single-cell genomics and trajectory analyses, the cell subsets in differentiation states with distinct characteristics can be classified, and the new classifications have been reported to be correlated with diagnosis, progression and therapeutic outcomes in several diseases and tumors [15–19]. However, it remains unclear whether GBM cells are in different differentiation states and whether a new classification of GBM patients based on cell differentiation trajectories correlates with tumor biological behaviors and plays a role in predicting patient survival and the immunotherapy response.

Therefore, in this study, we included the transcriptomic data of human GBMs to verify our hypothesis that GBM cancer cells have diverse differentiation characteristics and that the classification of patients based on the differentiation features of GBM cells can predict the tumor immunotherapy response and patient survival. First, we identified two GBM cell subsets in distinct differentiation states by trajectory analysis using single-cell RNA-sequencing (scRNA-seq) data and identified significant GBM cell differentiation-related genes (GDRGs). Second, we explored the biological functions of the GDRGs and found that they are related to tumor immune regulation and metabolic pathways. Third, we included GBM patients from The Cancer Genome Atlas (TCGA) database and classified them based on the expression patterns of GDRGs and demonstrated that this GBM cell differentiation state-based classification method revealed significantly different OS outcomes and different likelihoods of an immunotherapy response. Then, *FN1*, *APOE*, *RPL7A* and *GSTM2* were identified as the 4 key OS-predicting GDRGs, and a clinically applicable prognostic nomogram using these 4 GDRGs and other clinicopathological variables was successfully developed for GBM patients. Finally, the above findings were validated using the GBM patient cohort from the Chinese

Glioma Genome Atlas (CGGA) database. We identified distinct intratumoral GBM cell differentiation states and highlighted their essential role in predicting the clinical outcomes of GBM patients and tumor responses to immunotherapy.

RESULTS

Identification of 13 cell clusters in human GBMs using scRNA-seq data reveals high cell heterogeneity

A schematic diagram of the study design and principal findings is shown in Figure 1. Following the quality control standard and the normalization of GBM scRNA-seq data, 194 low-quality cells were excluded, and 2,149 cells from GBM cores were included in the analysis (Figure 2A). The number of genes detected was significantly related to the sequencing depth (Figure 2B). A total of 19,752 corresponding genes were included, and the variance analysis revealed 1,500 highly variable genes (Figure 2C). Principal component analysis (PCA) was performed to identify available dimensions and screen correlated genes. The top 20 significantly correlated genes are displayed as dot plots and heatmaps in Supplementary Figure 1. However, the PCA results did not demonstrate clear separations among cells in human GBMs (Figure 2D). We selected 20 principal components (PCs) with an estimated P value < 0.05 for subsequent analysis (Figure 2E).

Afterwards, the t-distributed stochastic neighbor embedding (tSNE) algorithm was applied, and cells in human GBMs were successfully classified into 13 separate clusters (Figure 2F). Differential expression analysis was performed, and a total of 8,025 marker genes from all 13 clusters were identified (Figure 2G). According to the expression patterns of the marker genes, these clusters were annotated by singleR and CellMarker (Figure 3A). Cluster 0, containing 518 cells, was annotated as GBM CSCs; clusters 1, 2, 6 and 10, containing 878 cells, were annotated as GBM cancer cells or GBM cells; cluster 3, containing 196 cells, was annotated as astrocytes; cluster 11, containing 44 cells, was annotated as oligodendrocytes; clusters 4, 5 and 9, containing 319 cells, were annotated as tumor-associated macrophages; cluster 8, containing 77 cells, was annotated as typical M1 macrophages; cluster 7, containing 81 cells, was annotated as typical M2 macrophages; and cluster 12, containing 36 cells, was annotated as T cells.

GBM cells can be divided into two subsets with distinct differentiation patterns, with their GDRGs correlating with immune and metabolic pathways

Trajectory analysis was performed to project all cells from GBMs onto one root and two branches, termed

branches I and II (Figure 3B). The results showed that GBM CSCs were mainly located in the root, whereas GBM cells were mostly located in the branches. Interestingly, GBM cells in branch I, defined here as type I GBM cells, were all from clusters 2, 6, and 10 (434 cells), and GBM cancer cells in branch II, defined here as type II GBM cells, were all from cluster 1 (444 cells). The degree of differentiation of cells in the type I and type II GBM subsets varied significantly. The branch-dependent marker genes of type I and type II GBM cells were determined to be GDRGs. Differential analysis was performed, and 265 marker genes were identified as type I GDRGs, and 193 marker genes were identified as type II GDRGs. Therefore, a total of 498 GDRGs were ultimately identified in GBM (Supplementary Table 1).

Gene set enrichment analysis (GSEA) was performed to identify related molecular mechanisms and pathways of GBM cells with distinct differentiation patterns. The

results revealed that type I GDRGs were significantly negatively correlated with the regulation of immune processes, such as antigen processing and immune cell differentiation (Figure 3C), and that type II GDRGs were significantly positively correlated with metabolism-related pathways, such as carbon metabolism, amino acid biosynthesis, glycolysis, and gluconeogenesis (Figure 3D). These findings from GDRGs indicate that GBM cells in distinct differentiation states demonstrate distinct tumor biology characteristics, which might provide new evidence for the molecular signatures of GBMs, including both intrinsic properties and the regulation of related pathways.

Then, we determined whether the observed GBM cell subsets could be identified using bulk RNA-seq data. As shown in Figure 4A–4C, the correlation analyses demonstrated that the type I GDRGs were highly correlated between scRNA-seq and bulk RNA-seq data from both the TCGA and CGGA databases, as were the

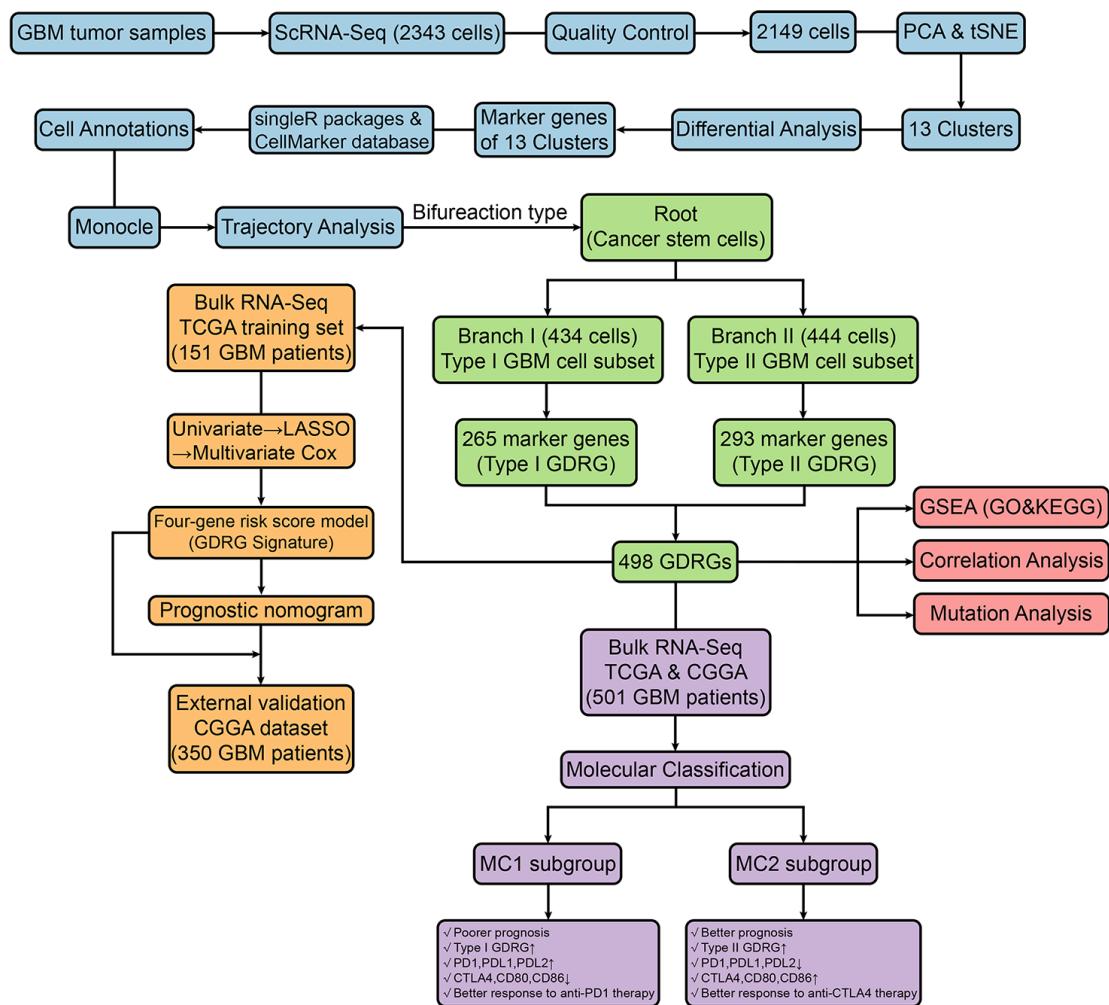


Figure 1. Schematic diagram showing the study design and principal findings.

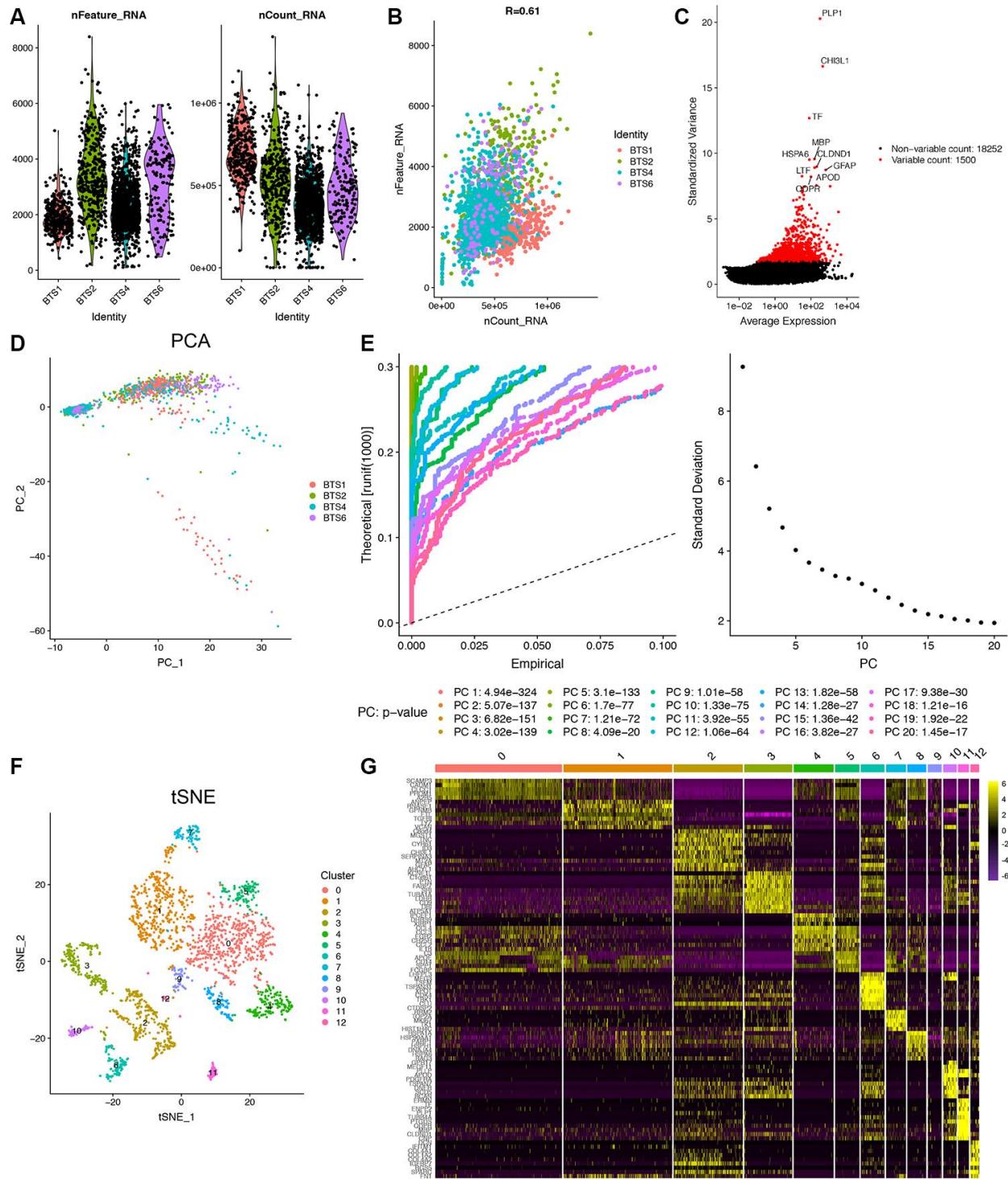


Figure 2. Identification of 13 cell clusters with diverse annotations revealing high cellular heterogeneity in GBM tumors based on single-cell RNA-seq data. (A) After quality control of the 2,343 cells from the tumor cores of 4 human GBM samples, 2,149 cells were included in the analysis. **(B)** The numbers of detected genes were significantly related to the sequencing depth, with a Pearson's correlation coefficient of 0.61. **(C)** The variance diagram shows 19,752 corresponding genes throughout all cells from GBMs. The red dots represent highly variable genes, and the black dots represent nonvariable genes. The top 10 most variable genes are marked in the plot. **(D)** PCA did not demonstrate clear separations of cells in GBMs. **(E)** PCA identified the 20 PCs with an estimated P value < 0.05. **(F)** The tSNE algorithm was applied for dimensionality reduction with the 20 PCs, and 13 cell clusters were successfully classified. **(G)** The differential analysis identified 8,025 marker genes. The top 20 marker genes of each cell cluster are displayed in the heatmap. A total of 96 genes are listed beside of the heatmap after omitting the same top marker genes among clusters. The colors from purple to yellow indicate the gene expression levels from low to high.

type II GDRGs. These findings suggest that type I and type II GBM cells can also be identified by GDRG expression using bulk RNA-seq data, since these highly correlated genes could indicate a common cellular origin.

GDRGs in the two GBM cell subsets are functionally correlated and mostly mutated

To determine whether the gene profiles originating from different GBM cell subsets were functionally correlated, we utilized metagenes to represent the overall expression patterns of the corresponding gene profile. The expression of type I and type II metagenes, consisting of type I and type II GDRGs, respectively, were derived from the weighted averages of the expression of the constituent genes. As shown in Figure 4D–4F, the results demonstrated significantly strong correlations between type I and type II metagene expression in both scRNA-seq and bulk RNA-seq data, indicating that type I and type II GDRGs are

functionally correlated even though they originate from distinct GBM subsets with different differentiation patterns.

We also analyzed the somatic mutation statuses of the GDRGs in the TCGA cohort. Most GDRGs (90.8%, 452/498) harbored mutations, and the top 9 mutated GDRGs with mutation frequencies $\geq 5\%$ in GBM patients are shown in Figure 4G. Epidermal growth factor receptor (*EGFR*) exhibited the highest mutation frequency (53%), followed by *CDK4* (16%) and *TSPAN31* (16%). There were 246 genes (92.8%) exhibiting mutations among type I GDRGs and 262 genes (89.4%) exhibiting mutations among type II GDRGs. The mutation frequencies were not significantly different between the two groups ($P=0.183$, Figure 4H). However, 8 of the top 9 mutated GDRGs were type II GDRGs, and only 1 was a type I GDRG. These findings demonstrate the high mutation status heterogeneity in GDRGs, suggesting the pivotal roles of GDRGs in the development and progression of GBMs.

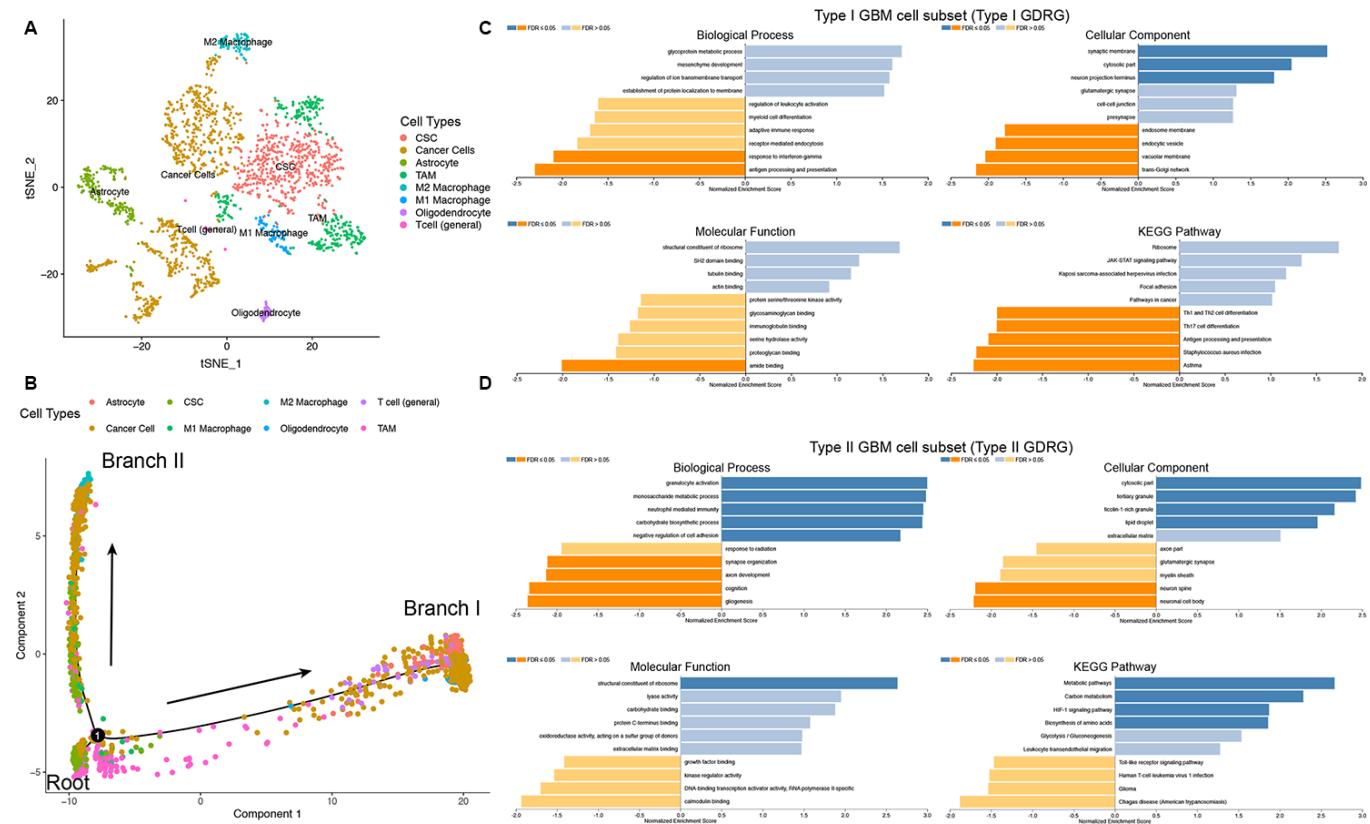


Figure 3. Cell annotation, trajectory analysis and GSEA of two GBM cell subsets with distinct differentiation patterns. (A) All 13 clusters of cells in GBMs were annotated by singleR and CellMarker according to the composition of the marker genes. (B) Trajectory analysis revealed two subsets of GBM cells with distinct differentiation patterns. GBM CSCs were mainly located in the root, whereas GBM cells were located in either branch. Branch I GBM cells were defined by the type I GBM cell subset (434 GBM cells). Branch II GBM cells were defined by the type II GBM cell subset (444 GBM cells). (C and D) GSEA of type I and II GBM cell subsets was performed to identify related molecular mechanisms and pathways. An FDR ≤ 0.05 was considered statistically significant.

GDRG-based classification of GBM patients correlates with different OS outcomes and clinicopathological characteristics

To establish a classification of GBMs based on the expression patterns of GDRGs, machine learning-based unsupervised consensus clustering was performed on 151 GBM patients from the TCGA database. According to the relative change in the area under the cumulative distribution function (CDF) curve and consensus

heatmap, the optimal number of clusters was determined to be two (k value = 2), and no appreciable increase was observed in the area under the CDF curve (Figure 5A–C). Thus, all GBMs were classified into two groups: 80 (53.0%) in molecular cluster 1 (MC1) and 71 (47.0%) in molecular cluster 2 (MC2). Kaplan-Meier survival analysis indicated that patients with MC1 GBMs experienced significantly poorer OS than patients with MC2 GBMs (log-rank $P=5.65\times 10^{-3}$; Figure 5D).

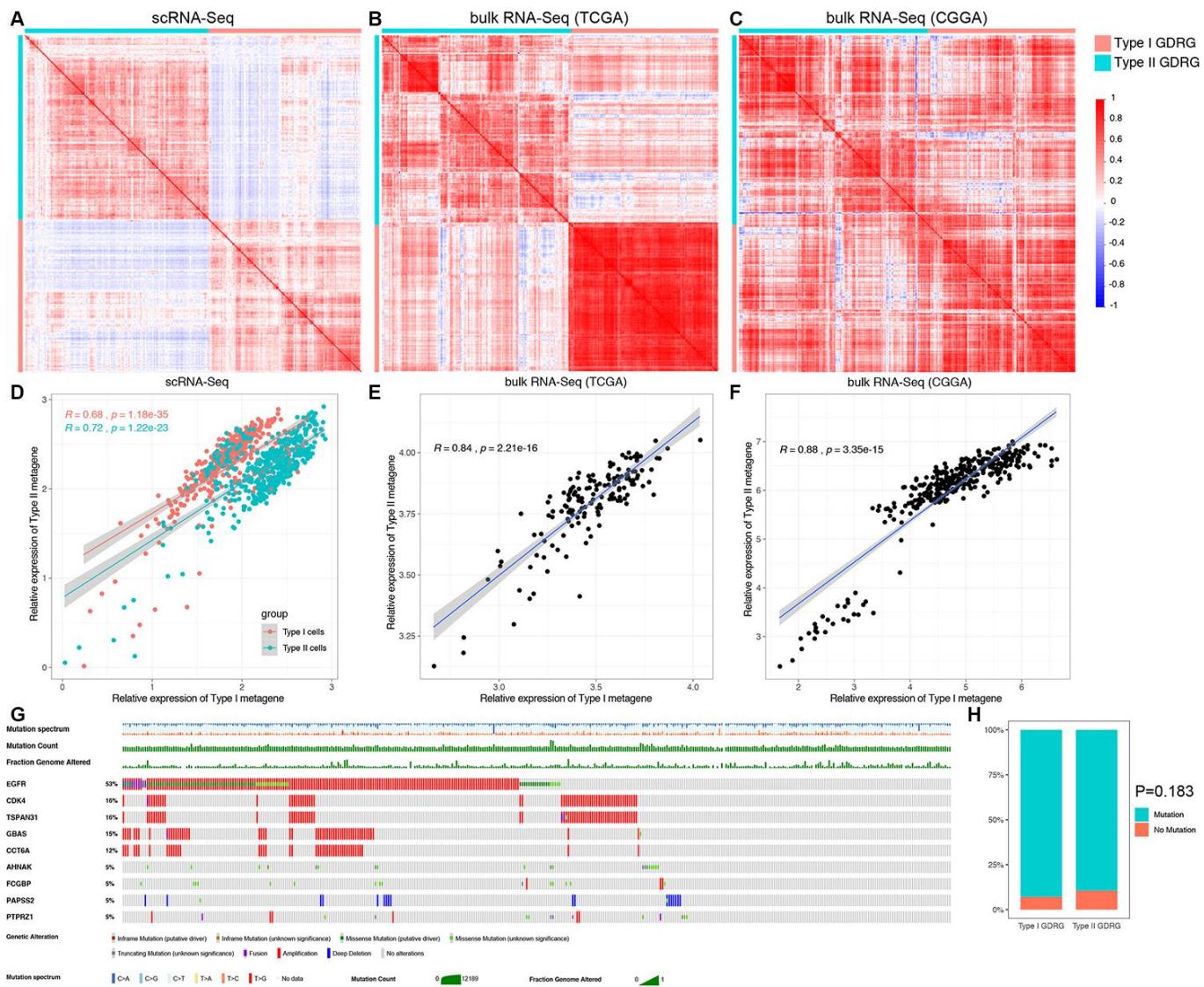


Figure 4. Correlation analysis and somatic mutation analysis of the two subtypes of GDRGs. The correlation heatmaps, which were generated to determine whether the observed GBM cell subsets could be identified using bulk RNA-seq data, demonstrated that the type I GDRGs were highly correlated in both scRNA-seq data (A) and bulk RNA-seq data, including the TCGA (B) and CGGA cohorts (C). The same result was observed for the type II GDRGs (A–C). The correlation analysis demonstrated that the expression of type I and type II metagenes was significantly correlated in both scRNA-seq data (D) and bulk RNA-seq data, including the TCGA (E) and CGGA cohorts (F). (G) OncoPlot analysis of the somatic mutation status of the GDRGs in the TCGA cohort revealed the top 9 mutated genes with mutation frequencies $\geq 5\%$. (H) Mutation frequencies of type I and type II GDRGs. A total of 246 genes (92.8%) were mutated in type I GDRGs, and 262 genes (89.4%) were mutated in type II GDRGs ($P=0.183$).

Afterwards, we validated the GDRG-based classification in the CGGA cohort. As shown in Figure 5G–5I, the optimal number of clusters was also determined to be two (k value = 2), and the patients were also classified as MC1 (265 patients, 75.7%) and MC2 (85 patients, 24.3%). Kaplan-Meier survival analysis also demonstrated that patients with MC1 GBMs had poorer OS than patients with MC2 GBMs ($\log\text{-rank } P=1.26\times 10^{-2}$; Figure 5J).

Additionally, we compared the expression patterns of the GDRGs and clinicopathological characteristics between two MCs of patients in the TCGA cohort. As shown in Figure 5E and 5F, the expression levels of the type I GDRG metagene were significantly higher ($P<0.001$) and the expression levels of the type II GDRG metagene were significantly lower ($P<0.001$) in patients with MC1 GBMs than in patients with MC2

GBMs. The same findings were also observed in the CGGA validation cohort (Figure 5K and 5L). Hence, we hypothesize that MC1 patients correspond mostly to the functional properties of the type I GBM subset and MC2 patients correspond mostly to the functional properties of the type II GBM subset.

The demographics and clinicopathological characteristics of patients with MC1 and MC2 GBMs are shown in Table 1. Compared with MC2 patients, MC1 patients were significantly younger in both the TCGA cohort ($P=0.007$) and CGGA cohort ($P=0.002$). However, no significant differences were observed in other variables between the two MCs of patients (all $P > 0.05$). Overall, the above findings indicate that this GDRG-based classification of GBM patients is robust and reliable across different populations, and different survival outcomes can be clearly discriminated according to this classification.

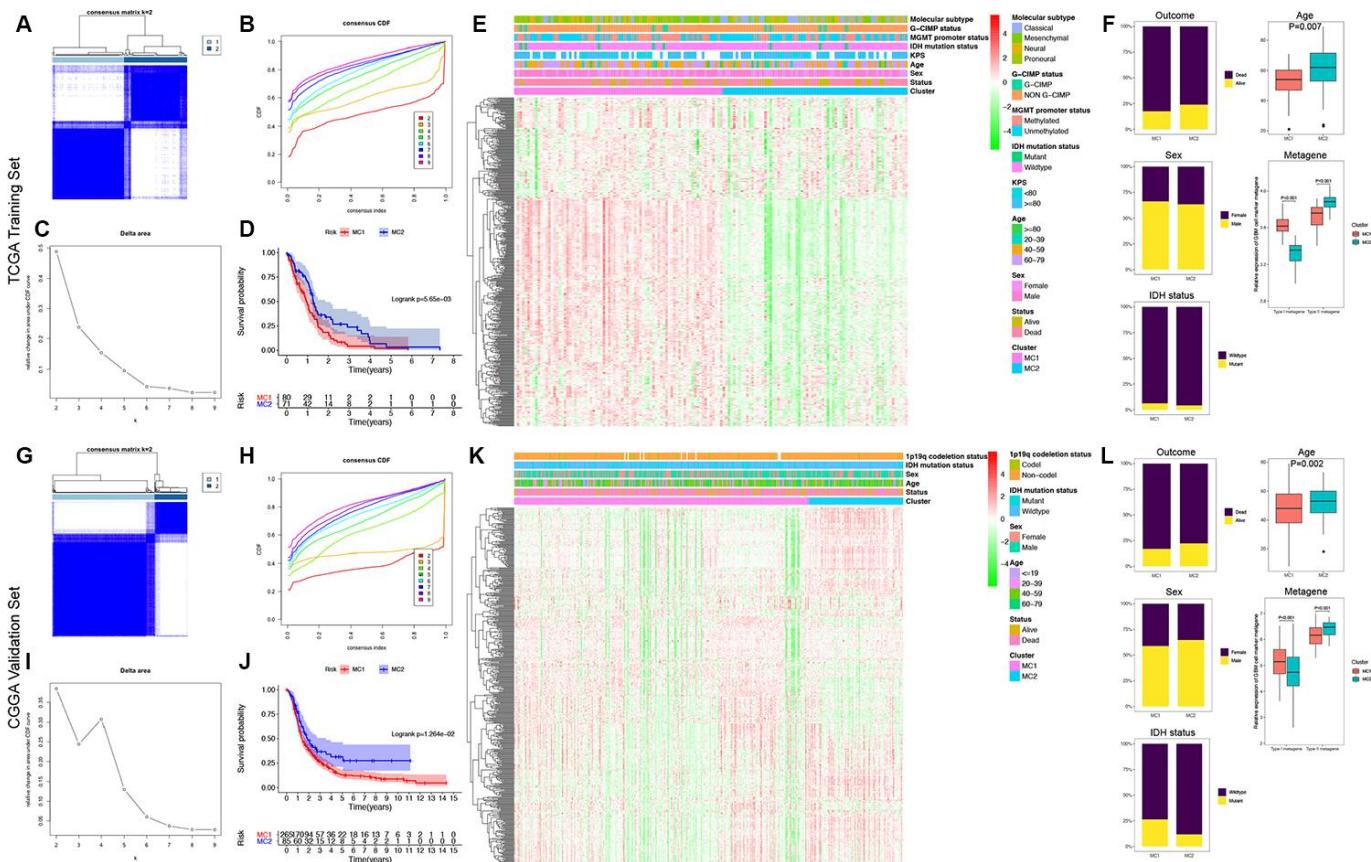


Figure 5. Identification and validation of the GDRG-based classification of GBM patients. Consensus clustering matrix for $k = 2$, which was the optimal cluster number in the TCGA training cohort (A) and CGGA validation cohort (G). (B and H) CDF curves of the consensus score ($k = 2$ -9) in the TCGA and CGGA cohorts. (C and I) Relative change in the area under the CDF curve ($k = 2$ -9) in the TCGA and CGGA cohorts. Kaplan-Meier survival analyses of the patients with MC1 and MC2 GBMs in the TCGA (D) and CGGA (J) cohorts, indicating that the patients with MC1 GBMs had poorer OS than those with MC2 GBMs. Heatmap and clinicopathological features of the two MCs in the TCGA (E) and CGGA (K) cohorts showing that the expression levels of the type I GDRG metagene were significantly higher and the levels of the type II GDRG metagene were significantly lower in patients with MC1 GBMs than in patients with MC2 GBMs. (F and L) Comparisons of the clinicopathological variables and type I and II metagenes between the two MCs of GBM patients in the TCGA and CGGA cohorts.

Table 1. Demographics and clinicopathological characteristics of GBM patients in the TCGA training cohort and CGGA validation cohort in the MC1 and MC2 subgroups based on the molecular classification.

Variable	TCGA training cohort			CGGA validation cohort		
	Total (n=151)	MC1 (n=80)	MC2 (n=71)	Total (n=350)	MC1 (n=265)	MC2 (n=85)
Age (years)	59.6±13.7	58.7±13.3	60.6±14.1	48.1±13.3	47.0±13.8	51.5±11.1
Sex						
Female	53	27	26	139	109	30
Male	98	53	45	211	156	55
KPS						
< 80	32	16	16	NA	-	-
≥ 80	81	40	41	NA	-	-
NA	38	24	14	NA	-	-
Pharmacotherapy						
TMZ	64	33	31	61 (No)	51	10
TMZ+BEV	26	17	9	269 (Yes)	201	68
Others (No TMZ)	19	11	8	-	-	-
No or NA	42	19	23	20 (NA)	13	7
Radiotherapy						
No	22	15	7	48	39	9
Yes	122	64	58	283	214	69
NA	7	1	6	19	12	7
Surgery						
Biopsy only	16	6	10	NA	-	-
Tumor resection	135	74	61	NA	-	-
IDH mutation status						
Wild-type	147	75	68	270	195	75
Mutant	8	5	3	80	70	10
MGMT promoter status						
Methylated	66	40	26	NA	-	-
Unmethylated	85	40	45	NA	-	-
TERT status						
Wild-type	146	78	68	NA	-	-
Mutant	5	2	3	NA	-	-
BRAF status						
Wild-type	146	76	70	NA	-	-
Mutant	5	4	1	NA	-	-
ATRX status						
Wild-type	140	72	68	NA	-	-
Mutant	11	8	3	NA	-	-
EGFR status						
Wild-type	97	56	41	NA	-	-
Mutant	54	24	30	NA	-	-
G-CIMP status						
Non G-CIMP	140	74	66	NA	-	-
G-CIMP	11	6	5	NA	-	-
Molecular subtype						
Classical	38	11	27	NA	-	-
Neural	26	15	11	NA	-	-
Mesenchymal	49	40	9	NA	-	-
Proneural	38	14	24	NA	-	-

1p/19q status					
Noncodeletion	NA		323	240	83
Codeletion	NA		17	15	2
NA	NA		10	10	0

"Others (No TMZ)" in pharmacotherapy included PCV, PCV+BEV, and other drugs, including avastin, carmustine, and irinotecan.

MC, molecular cluster; GBM, glioblastoma; NA, not available; KPS, Karnofsky performance status; TMZ, temozolomide; BEV, bevacizumab; PCV, procarbazine lomustine vincristine.

GDRG-based classification of GBM patients correlates with the expression patterns of immune checkpoints and different likelihoods of an immunotherapy response

The expression of 6 main immune checkpoints, namely, PDCD1 (PD1), CD274 (PDL1), PDCD1LG2 (PDL2), CTLA4, CD80 and CD86, was determined and compared between two GBM cell subsets and two MCs of GBM patients (Figure 6). In terms of scRNA-seq data, PD1, PDL1 and PDL2 were highly expressed in type I GBM cells, whereas CTLA4, CD80 and CD86 were highly expressed in type II GBM cells (Figure 6A). In terms of bulk RNA-seq data, PD1, PDL1 and PDL2 were highly expressed in MC1 GBM patients whereas CTLA4, CD80 and CD86 were highly expressed in MC2 GBM patients in both the TCGA database (Figure 6B) and CGGA database (Figure 6C).

Afterwards, the tumor immune dysfunction and exclusion (TIDE) algorithm was used to predict the likelihood of an immunotherapy response. According to the results from the TCGA training cohort, MC2 GBM patients (43.7%, 31/71) were more likely to respond to immunotherapy than MC1 patients (20.0%, 16/80, $P = 0.003$). Similarly, in the CGGA validation cohort, MC2 GBM patients (45.9%, 39/85) were more likely to respond to immunotherapy than MC1 patients (37.6%, 62/165, $P < 0.001$).

Then, SubMap analysis was performed to predict the likelihood of a clinical response to PD1 and CTLA4 inhibitors in the two MCs of GBM patients. MC1 GBM patients in both the TCGA and CGGA cohorts were more sensitive to anti-PD1 therapies, with Bonferroni-corrected P values of 0.001 and 0.033, respectively, while MC2 GBM patients were more sensitive to anti-CTLA4 therapies, with Bonferroni-corrected P values of 0.019 and 0.047, respectively (Figure 6D and 6E).

FN1, APOE, RPL7A and GSTM2 are the 4 most significant survival-predicting GDRGs in human GBMs

Univariate Cox analysis was performed, and 45 prognosis-associated GDRGs were identified in the

TCGA training set. Least absolute shrinkage and selection operator (LASSO) followed by multivariate Cox analysis was then performed, and 4 significant survival-predicting GDRGs were identified (Supplementary Figure 2): fibronectin 1 (*FN1*, HR=5.28, $P < 0.001$), apolipoprotein E (*APOE*, HR=0.39, $P=0.001$), ribosomal protein L7a (*RPL7A*, HR=0.27, $P=0.021$), and glutathione S-transferase mu 2 (*GSTM2*, HR=0.41, $P=0.042$). The expression levels of the 4 most significant prognostic GDRGs in both the scRNA-seq and bulk RNA-seq profiles are shown in Supplementary Figure 3. *FN1* was significantly upregulated in GBM cells and T cells (general); *APOE* was significantly upregulated in GBM cells, CSCs and macrophages; *RPL7A* was significantly upregulated in GBM cells, CSCs and macrophages; and *GSTM2* was significantly upregulated in GBM cells and astrocytes (Supplementary Figure 3A). Furthermore, the expression of these 4 GDRGs was validated using the Gene Expression Profiling Interactive Analysis (GEPIA) database, which includes 163 GBM (TCGA) samples and 207 normal (GTEx) samples. We found that all 4 survival-predicting GDRGs were upregulated in GBMs compared to normal tissues (Supplementary Figure 3B).

Kaplan-Meier survival analysis demonstrated that high expression of *FN1* and low expression of *APOE*, *GSTM2* and *RPL7A* were associated with poor OS in GBM patients (Supplementary Figure 3C).

Generation and validation of the GDRG-based prognostic risk score model to predict GBM patient survival

The prognostic risk score model was developed based on the above 4 GDRGs with the following formula: Risk score = $\text{Exp}_{FN1} \times 1.66 + \text{Exp}_{APOE} \times (-0.93) + \text{Exp}_{RPL7A} \times (-1.30) + \text{Exp}_{GSTM2} \times (-0.90)$. Risk scores were calculated for all patients in the TCGA training set, and the patients were divided into either a high-risk (high score) group or a low-risk (low score) group using the median value of the risk score as the cutoff value (Figure 7C). Kaplan-Meier survival analysis demonstrated that patients in the high-risk group had significantly poorer OS than those in the low-risk group (log-rank, $P = 2.778 \times 10^{-4}$; Figure 7A). The

C-index of the GDRG signature for OS prediction was 0.781 (95% CI=0.742 to 0.820). Time-dependent receiver operating characteristic (ROC) analysis also indicated that the GDRG signature showed excellent performance in predicting the 0.5-, 1-, 2- and 3-year OS rates, with respective area under the curve (AUC) values of 0.767, 0.712, 0.752 and 0.776 (Figure 7B).

Then, the predictive formula was validated using the CGGA cohort in a similar manner. As shown in Figure 7F, all GBM patients were classified into high-risk or low-risk groups. Consistent with the results from the TCGA training set, the survival analysis also demonstrated that patients in the high-risk group had significantly poorer OS than patients in the low-risk group (log-rank, $P = 9.783 \times 10^{-6}$; Figure 7D). The C-index of the GDRG signature was 0.715 (95% CI=0.676

to 0.754). Time-dependent ROC analysis also suggested favorable values in predicting OS in the CGGA validation set (Figure 7E). These results indicate that the GDRG-based prognostic risk score model can serve as a reliable prognostic predictor for different populations of GBM patients.

Development and validation of the clinically applicable prognostic nomogram with the GDRG signature and clinicopathological parameters

To investigate whether the prognostic significance of the GDRG signature is independent of the other clinicopathological variables in predicting the survival of GBM patients, univariate and multivariate Cox regression analyses were performed, and the results demonstrated that the GDRG signature was independently

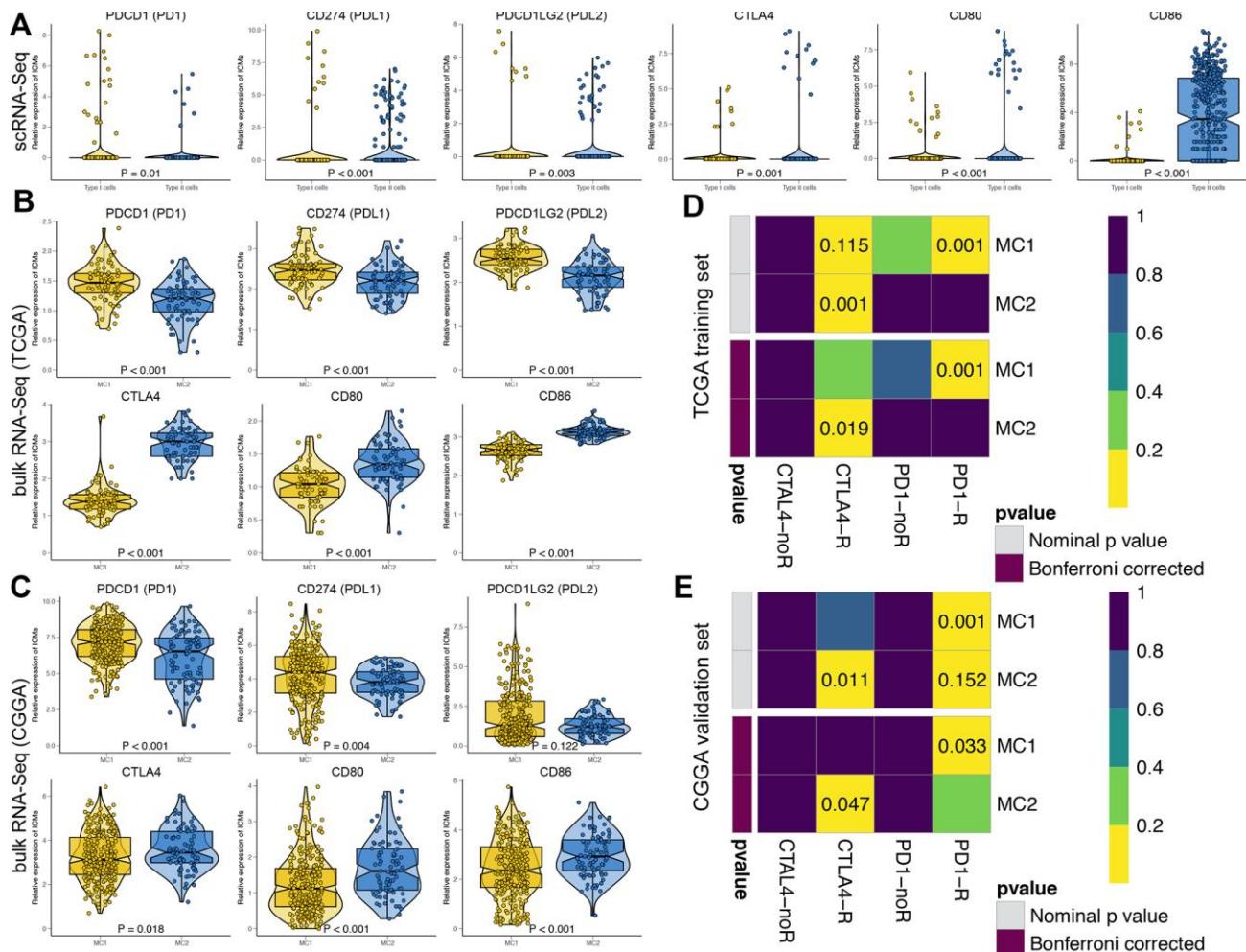


Figure 6. Predictions of the immunotherapy response in GBM patients. The violin plots present the expression of 6 principal immune checkpoint molecules, namely, PDCD1 (PD1), CD274 (PDL1), PDCD1LG2 (PDL2), CTLA4, CD80, and CD86, in scRNA-seq data (A) and bulk RNA-seq data, including the TCGA (B) and CGGA cohorts (C). Subclass mapping analysis was used to predict the likelihood of the clinical response to anti-PD1 and anti-CTLA4 therapy for MC1 and MC2 GBM patients from the TCGA (D) and CGGA (E) cohorts. R represents immunotherapy responders, while noR represents immunotherapy nonresponders. A Bonferroni-corrected P value < 0.05 was considered statistically significant.

associated with OS in both the TCGA and CGGA cohorts (Table 2). Finally, a prognostic nomogram was successfully developed to provide a clinically applicable quantitative approach for individual OS prediction. Age, pharmacotherapy, radiotherapy, IDH mutation status, MGMT promoter methylation status, and the GDRG signature were included in the final OS prediction model (Figure 8A). The C-index of the prognostic nomogram was 0.896 (95% CI=0.857 to 0.935). Time-dependent ROC analysis revealed excellent predictive abilities for the 0.5-, 1-, 2- and 3-year OS rates, with AUC values of 0.734, 0.771, 0.864 and 0.919, respectively (Figure 8B). The calibration plots showed excellent agreement between the predicted 0.5-, 1- and 3-year OS rates and the actual observations in the TCGA cohort (Figure 8D–8F). Then, the prognostic model was validated in the CGGA cohort, with a C-index of 0.729 (95% CI=0.690 to 0.768). The time-dependent AUCs for the 0.5-, 1-, 2-, and 3-year OS rates with the prognostic nomogram were 0.725, 0.696, 0.694, and 0.701, respectively (Figure 8C). The calibration plots also showed excellent agreement between the OS predictions and the actual observations for the probabilities of the 0.5-, 1- and 3-year survival rates in the validation set (Figure 8G–8I). The above findings suggest the appreciable reliability of the prognostic nomogram for OS prediction that can be applied in different populations of GBM patients.

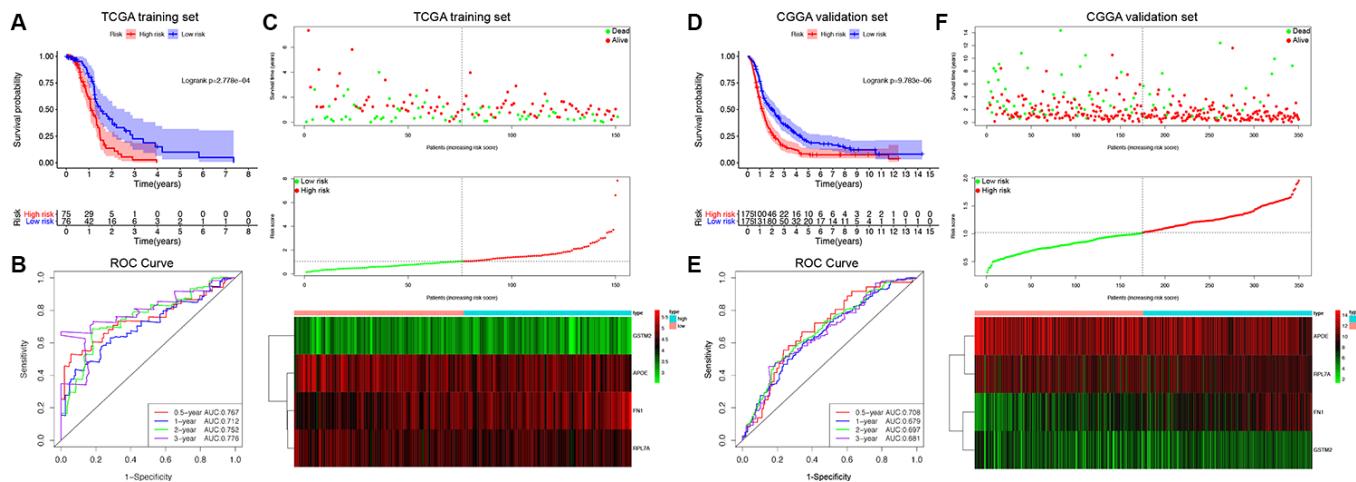


Figure 7. Survival analysis, prognostic performance and risk score analysis of the GDRG-based risk score model in GBM patients. Kaplan-Meier survival analysis was performed to estimate the OS of high-risk and low-risk patients in the TCGA training cohort (A) and CGGA validation cohort (D). The high-risk groups had significantly poorer OS than the low-risk groups. Time-dependent ROC curve analysis was performed to evaluate the prognostic performance of the GDRG signature for predicting the 0.5-, 1-, 2-, and 3-year OS rates in the TCGA (B) and CGGA cohorts (E). Risk score analysis of the GDRG signatures in the TCGA (C) and CGGA (F) cohorts were calculated, and the patients were divided into either a high-risk group or a low-risk group using the median value of the risk score as the cutoff value. Upper panel: Patient survival status and time distributed by the risk score. Middle panel: Risk score curves of the GDRG signatures. Bottom panel: Heatmaps of the expression levels of the 4 GDRGs in the GBM samples. The colors from green to red indicate the gene expression levels from low to high.

DISCUSSION

Cancer cells are derived from CSCs, initiate tumor mass growth, and drive tumor progression and invasion forward. They differentiate into diverse subpopulations during differentiation due to hyperproliferation and increased genetic instability [20]. As a result, one group of cells can express heterogeneous phenotypes within a tumor and stay in distinct differentiation states [19, 21]. GBM, the most common and lethal CNS neoplasm, has a highly heterogeneous intratumoral cell composition [22]. Since intratumoral heterogeneity is increasingly considered to be one of the main causes of treatment resistance, there is an urgent need to harness new technologies, including single-cell analysis, to explore cell heterogeneity in GBMs [23]. Several studies have researched the differentiation of GBM CSCs [24–26]. However, to date, studies on the differentiation states of GBM cancer cells are limited. Whether GBM cell differentiation states are correlated with clinical outcomes and the treatment response remains unknown. In this study, we demonstrated that GBM cells could be divided into two subsets with distinct differentiation characteristics, and the classification of patients based on GBM cell differentiation patterns was correlated with patient OS after treatment and the tumor response to immunotherapy. These findings were initially obtained based on scRNA-seq data and then validated using the bulk RNA-seq data of GBM patients from two large databases.

Table 2. Univariate and multivariate Cox proportional hazards analyses of clinicopathological variables and GDRG signatures in the TCGA GBM training cohort and CGGA GBM validation cohort.

Variable	TCGA training cohort (n=151)				CGGA validation cohort (n=350)			
	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
Age	1.028 (1.013-1.044)	1.98e-04	1.017 (1.000-1.035)	4.59e-02	1.078 (1.048-1.108)	8.35e-05	1.006 (1.001-1.011)	1.90e-02
Sex (Female/Male)	0.916 (0.626-1.341)	0.65	-	-	1.063 (0.837-1.350)	0.61	-	-
KPS (<80/≥80/NA)	0.926 (0.696-1.233)	0.59	-	-	NA		NA	
Pharmacotherapy	0.883		0.918		0.573		0.663	
(TMZ/TMZ+BEV/Others (No TMZ)/No or NA)	(0.852-0.913)	1.06e-04	(0.879-0.958)	2.87e-02	(0.432-0.759)	1.04e-04	(0.487-0.901)	8.80e-03
Radiotherapy (No/Yes/NA)	0.433 (0.262-0.714)	1.04e-03	0.273 (0.156-0.476)	4.69e-06	0.668 (0.492-0.908)	9.96e-03	0.722 (0.682-0.762)	4.11e-02
Surgery (Biopsy only/Tumor resection)	0.934 (0.523-1.667)	0.82	-	-	NA		NA	
IDH mutation status (Wild- type/Mutant)	0.262 (0.096-0.715)	8.91e-03	0.279 (0.240-0.318)	3.24e-02	0.752 (0.566-0.988)	3.89e-02	0.807 (0.767-0.847)	5.68e-03
MGMT promoter status (Methylated/Unmethylated)	1.434 (1.133-1.733)	6.84e-03	1.365 (1.325-1.404)	1.31e-02	NA		NA	
TERT promoter status (Wild-type/Mutant)	0.906 (0.287-2.861)	0.87	-	-	NA		NA	
BRAF status (Wild-type/Mutant)	1.973 (0.720-5.410)	0.19	-	-	NA		NA	
ATRX status (Wild-type/Mutant)	0.426 (0.187-0.973)	4.28e-02	0.899 (0.703-2.095)	0.32	NA		NA	
EGFR status (Wild-type/Mutant)	1.273 (0.873-1.857)	0.21	-	-	NA		NA	
G-CIMP status (Non- or G-CIMP)	0.241 (0.088-0.655)	5.32e-03	1.474 (0.912-2.043)	0.29				
Molecular subtype (Classical/Neural/Mesenchy- mal/Proneural)	0.971 (0.831-1.133)	0.71	-	-				
1p/19q status (Noncodeletion/Codeletion/N A)	NA		NA		0.913 (0.662-1.259)	0.58	-	-
GDRG signature	1.332 (1.153-1.540)	1.05e-04	1.297 (1.109-1.517)	1.17e-03	2.558 (1.815-3.606)	8.07e-08	2.609 (1.744-3.902)	3.03e-06

GDRG, GBM cell differentiation-related gene; GBM, glioblastoma; NA, not available; HR, hazard ratio; CI, confidence interval; KPS, Karnofsky performance status; 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.

All statistical tests were two-sided. Bold type indicates P<0.05.

Intratumoral heterogeneity refers to the different features of cells in a single tumor, and these cells manifest as a diverse collection of cells with distinct molecular signatures or differentiation states [27]. In this study, we identified 13 cell clusters in GBMs. Four of them were GBM cells, and 1 cluster was GBM CSCs. One cluster was annotated as astrocytes, one was annotated as oligodendrocytes, and the other 6 clusters were annotated as immune cells, mostly macrophages, which is consistent with the previous literature [28]. According to the trajectory analysis, GBM cells were projected into two subsets with remarkably distinct differentiation features, and subset-dependent GDRGs were identified. Using GSEA and correlation analysis,

we found that this differentiation model was associated with immune regulation and metabolic pathways in GBMs, implying that intrinsic correlations between GBM cell differentiation and intratumoral immune and metabolic biology exist.

EGFR, observed in approximately 57% of GBMs, acts as a major driver of tumor invasion, progression, and angiogenesis [29–31]. As shown in this study, *EGFR*, identified as a type II GDRG, was the top mutated GDRG, with a mutation frequency of 53%. This finding unveils the role of *EGFR* in regulating GBM cell differentiation and its subsequent role in interacting with tumor immunity and cell fate.

One hallmark of uncontrolled cancer progression and invasion is immune escape [20]. Immunotherapy can relieve the intratumoral immunosuppression status and acts as a promising treatment strategy for GBM patients; it has been previously shown to remarkably improve the survival of patients with several other tumors [5]. Unfortunately, the outcomes of almost all trials for GBMs, including those on immune checkpoint inhibitors, vaccinations, and adoptive T cell therapy, have not been as effective as expected. The low immunogenicity of GBM, the immune privilege of the CNS and the immunosuppressive microenvironment are considered key pathophysiologicals underlying the immunotherapy resistance of GBMs [32]. We revealed in this study that type I GDRGs are correlated with intratumoral immunosuppression, disturbing the processes of antigen processing and presentation and T cell differentiation in GBMs. TIDE analysis showed that patients with MC1 GBMs, namely, those harboring

more type I GDRGs, were less likely to respond to immunotherapy than patients with MC2 GBMs. The expression of immunotherapy-targeted molecules differed in GBMs with different cell differentiation patterns. Our results showed that MC1 GBMs expressed more PD1/PDL1/PDL2 molecules, while MC2 GBMs expressed more CTLA4/CD80/CD86 molecules. Moreover, the SubMap analysis confirmed that the immunoreaction was related to cell differentiation, predicting that patients with MC1 GBMs were more sensitive to anti-PD1 therapies, while patients with MC2 GBMs were more sensitive to anti-CTLA4 therapies. Therefore, based on these findings, we propose that the features of GBM cell differentiation states can be referred to as a good predictor for the GBM immunotherapy response.

We explored and validated the prognostic predictive value of the GDRGs and their correlations with patient

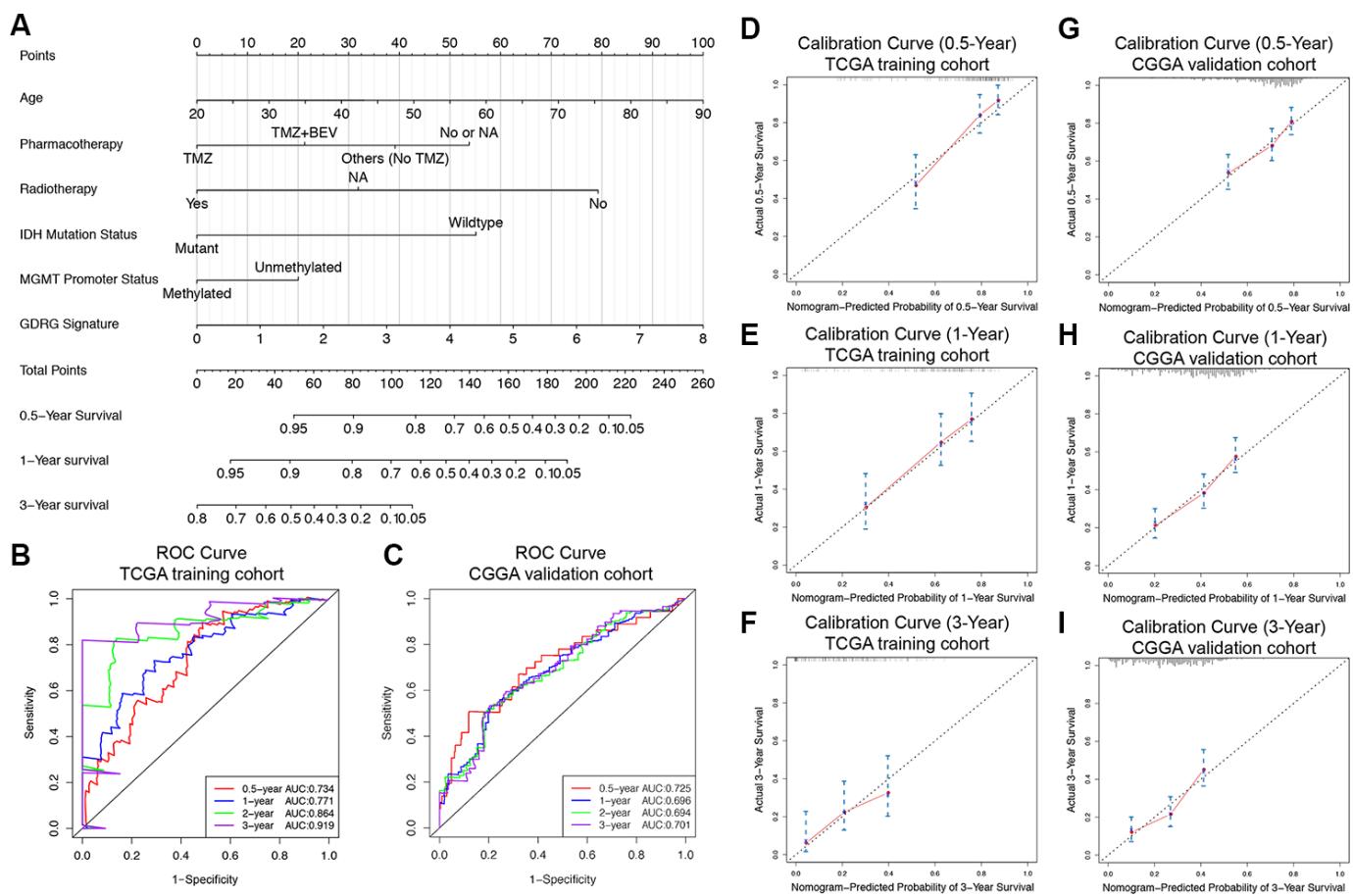


Figure 8. Prognostic nomogram to predict the 0.5-, 1-, and 3-year OS of GBM patients. (A) Nomogram model to predict the prognosis of GBM patients based on the TCGA training cohort. Age, pharmacotherapy, radiotherapy, IDH mutation status, MGMT promoter methylation status, and the GDRG signature were included in the prediction model. The prognostic performance of the nomogram demonstrated by the time-dependent ROC curve for predicting the 0.5-, 1-, and 3-year OS rates in the TCGA training cohort (B) and CGGA validation cohort (C). Calibration plots of the prognostic nomogram for predicting OS at 0.5, 1, and 3 years in the TCGA (D–F) and CGGA (G–I) cohorts. Actual survival is plotted on the y-axis, and nomogram-predicted probability is plotted on the x-axis.

clinical outcomes. The machine learning-based unsupervised clustering method was used in this study to classify GBM patients into two groups based on the cell differentiation states, namely, the GDRG features, of GBM cells. Two large GBM databases of different origins, the TCGA and CGGA, were used, and the survival analysis showed that patients with MC1 GBMs had poorer OS than patients with MC2 GBMs, indicating that the GDRG-based patient classification can be used to predict patient survival.

A nomogram is a multivariable regression model that is widely used in studies to predict clinical outcomes with intuitive visual presentations [33]. In this study, *FNI*, *APOE*, *RPL7A* and *GSTM2* were identified as the 4 most significant survival-predicting GDRGs in human GBMs. We successfully established a risk score formula based on these GDRGs and generated a clinically applicable nomogram with GDRG signatures and clinicopathological parameters to predict GBM patient outcome. We then validated this nomogram in two large GBM cohorts with long-term follow-up examinations, showing the high reliability of this nomogram. To our knowledge, this nomogram is the first to incorporate a cell differentiation-related signature for predicting GBM patient survival. This visualized scoring system may assist neurosurgeons and oncologists in performing survival predictions according to clinicopathological and cell differentiation information and in further proposing better treatment options.

The current study has some limitations. We conducted this analysis and made conclusions using data from published databases, and the prediction model was validated using TCGA and CGGA cohorts but not our own cohort. Additionally, detailed patient information was incomplete, and some clinical parameters, e.g., tumor imaging results, medical records and history, and operation note details, were not available for download and thus were not input into the nomogram. The predictive model needs to be further validated in prospective large-scale cohorts.

CONCLUSIONS

We used scRNA-seq and bulk RNA-seq data and found that GBM cells follow a two-dimensional differentiation trajectory and that their differentiation states correlate with several immune regulation and metabolic pathways. The classification of GBM patients based on GBM cell differentiation patterns can reliably predict patient survival, immune checkpoint expression, and the tumor immunotherapy response. We identified the key prognosis-predicting GDRGs and established a nomogram composed of patient clinicopathological variables and these GDRGs to predict GBM prognosis.

This study highlights the distinct cell differentiation trajectories of GBM cells and their essential roles in predicting the clinical outcome of GBM patients and the tumor immunotherapy response.

MATERIALS AND METHODS

Data acquisition and preprocessing

The scRNA-seq data and bulk RNA-seq data of human GBM samples were included in this study for analysis. The scRNA-seq data of a total of 3,589 cells of 4 human primary GBM samples, accession number GSE84465 [34], were obtained from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) database, containing 2,343 cells from tumor cores and 1,246 cells from peripheral regions, with a reading depth of 10 \times genomics based on Illumina NextSeq 500. The bulk RNA-seq profiles of GBM samples were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>) and the CGGA database (<http://www.cgga.org.cn>). We excluded samples with unavailable clinical information and ultimately included 151 GBMs from the TCGA cohort as the training set and 350 GBMs from the CGGA cohort as the validation set.

Processing of the GBM scRNA-seq data

In total, 2,343 cells from tumor cores were included in this analysis. The Seurat package in R 3.5.1 was used for quality control, statistical analysis, and exploration of the scRNA-seq data [35]. First, 194 low-quality cells were excluded based on the following quality control standards: 1) genes detected in < 3 cells were excluded; 2) cells with < 50 total detected genes were excluded; and 3) cells with $\geq 5\%$ of mitochondria-expressed genes were excluded. Then, the gene expression of the remaining 2,149 cells was normalized using a linear regression model. PCA was performed to identify significantly available dimensions with a P value < 0.05 [36]. Then, the t-distributed stochastic neighbor embedding (tSNE) algorithm was applied for dimensionality reduction with 20 initial PCs and for performing cluster classification analysis across all cells [37]. The differential expression analysis among all genes within cell clusters was performed using the limma package in R to identify the marker genes of each cluster. An adjusted P value < 0.05 and $|\log_2[\text{fold change (FC)}]| > 0.5$ were considered the cutoff criteria for identifying marker genes. Afterwards, different cell clusters were determined and annotated by the singleR package according to the composition patterns of the marker genes and were then manually verified and corrected with the CellMarker database [38, 39]. The corresponding genes of cell surface markers for the annotation of cell clusters are listed in Supplementary Table 2.

Trajectory analysis and GDRG identification

Single-cell pseudotime trajectories of the GBM scRNA-seq data were constructed using the Monocle 2 algorithm [14]. This algorithm adopts a machine learning technique, learning a parsimonious principal graph to reduce the given high-dimensional expression profiles to a low-dimensional space. Single cells were projected onto this space and ordered into a trajectory with branch points. For data interpretation, the cells in the same branch were generally considered to be in the same differentiation state, while cells located in different branches were considered to have different cell differentiation characteristics. In addition, differential expression analysis was performed between branches, and genes that showed differential expression levels were defined as branch-dependent or state-specific genes or marker genes. These marker genes of GBM cells located in different branches were defined as GDRGs.

GSEA, correlation analysis, and somatic mutation analysis of branch-dependent GDRGs

GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was performed to identify the related molecular mechanisms and pathways of GBM cells in different differentiation states [40]. The adjusted P value was used to correct the false positive results by using the default Benjamini-Hochberg false discovery rate (FDR) method. An FDR value ≤ 0.05 of the enrichment gene sets was considered statistically significant. Then, we determined whether the observed cell subtypes generated from the scRNA-seq GBM data could also be identified in the bulk RNA-seq data. Because the coregulation of transcriptional programs in bulk RNA-seq data would indicate a common cellular origin, we believed that cell populations would be best identified by distinguishing gene expression profiles consisting of highly correlated genes [15]. Thus, correlation analyses of the expression of GDRGs were performed using Pearson's correlation test, and the results were visualized as a heatmap. A P value < 0.05 and $|\text{correlation coefficient}| > 0.3$ were considered significantly correlative. To determine whether the gene profiles originating from different cell populations were functionally correlated, we utilized metagenes to represent the overall expression patterns of the corresponding gene profiles. The relative expression values of metagenes, an assembly of multiple genes, were derived as the weighted averages of the expression levels of the constituent genes using a logistic regression model [15]. Correlation analysis between branch-dependent GBM differentiation-related metagenes was performed by Pearson's correlation test in both the scRNA-seq data and bulk RNA-seq data. Moreover, we performed somatic mutation analysis to identify the

mutation statuses of the GDRGs. The cBioPortal database (<https://www.cbioperl.org/>) was used to download the somatic mutation data of GBM patients. The numbers of variant types and classifications were visualized with OncoPlot.

GDRG-based classifications of GBM patients in the TCGA and CGGA cohorts

Unsupervised consensus clustering, an algorithm based on k-means machine learning, was utilized to explore a molecular classification of both the TCGA and CGGA GBM cohorts based on the expression patterns of GDRGs using the 'ConsensusClusterPlus' package in R [41]. The clustering procedure, with 1,000 iterations, was performed by sampling 80% of the data in each iteration. The optimal number of clusters was determined by the relative change in the area under the CDF curves of the consensus score and consensus heatmap. Then, Kaplan-Meier survival analysis was performed to evaluate the prognosis of patients in different MCs. We also performed comparisons of the clinicopathological variables between different clusters of patients to further explore the associations between the GDRG-based MCs and the clinical features of GBM patients.

Immunotherapy response predictions

TIDE (<http://tide.dfci.harvard.edu/>) is a computational method that integrates the expression signatures of T cell dysfunction and exclusion to model tumor immune evasion [42]. We used the TIDE algorithm to predict the clinical response to immune checkpoint blockade (ICB) in GBM patients based on pretreatment genomics. The unsupervised subclass mapping method (SubMap; <https://cloud.genepattern.org/gp/>) was further applied to predict the response of GBM patients in different MCs to immunotherapy, including anti-PD1 and anti-CTLA4 therapy [43]. A Bonferroni-corrected P value < 0.05 was considered statistically significant.

Generation and validation of the GDRG-based prognostic risk score model

The associations between the expression levels of GDRGs and patient survival were first evaluated by univariate Cox regression analysis in the TCGA training cohort. The prognosis-related genes with a P value < 0.05 identified by the analysis were further screened by LASSO and multivariate Cox regression analyses. Consequently, a risk score model based on the key prognosis-related GDRGs was constructed to predict the prognosis of GBM patients [44]. We calculated the risk score of each patient by referring to our previously

constructed formula, Risk score = $\text{Exp}_{\text{GENE}1} \times \beta_1 + \text{Exp}_{\text{GENE}2} \times \beta_2 + \dots + \text{Exp}_{\text{GENE}n} \times \beta_n$, in which “Exp” represents the expression level of the corresponding gene, and “ β ” represents the regression coefficient calculated by the multivariate Cox analysis [9]. Accordingly, all GBM patients in the TCGA database were stratified into either the low-risk (low score) group or the high-risk (high score) group. Kaplan-Meier survival analysis was performed to estimate the OS of these two groups, and survival differences were evaluated by a two-sided log-rank test. The predictive accuracy of the GDRG-based prognostic model was evaluated by Harrell's concordance index (C-index) and time-dependent ROC curve analysis within 0.5, 1 and 3 years by utilizing the survcomp and survivalROC packages in R [33, 45]. Both the C-index and AUC ranged from 0.5 to 1, with 1 indicating perfect discrimination and 0.5 indicating no discrimination. Finally, the prognostic model generated by the TCGA training cohort was verified in the CGGA validation cohort.

Afterwards, univariate and multivariate Cox regression analyses were performed in both GBM cohorts to determine whether the predictive performance of the GDRG signatures could be independent of the clinicopathological variables. These variables included age, sex, Karnofsky performance status score, pharmacotherapy, radiotherapy, surgery, isocitrate dehydrogenase (IDH) mutation status, O6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation status, telomerase reverse transcriptase (TERT) promoter mutation status, B-Raf proto-oncogene (BRAF) mutation status, X-linked alpha thalassemia mental retardation syndrome gene (ATRX) mutation status, EGFR mutation status, glioma CpG island methylator phenotype (G-CIMP) status, molecular subtype (classical, neural, mesenchymal, and proneural), and 1p/19q status.

Development and validation of the prognostic nomogram with the GDRG signature and clinicopathological parameters

Following univariate and multivariate Cox regression analyses, all the identified independent prognostic parameters were utilized to develop a prognostic nomogram for predicting the 0.5-, 1-, and 3-year survival outcomes of GBM patients using the rms package in R. Calibration plots at 0.5, 1, and 3 years were constructed to graphically evaluate the discriminative ability of the nomogram [45]. Then, the discrimination performance of the nomogram was quantitatively assessed by the C-index and ROC curve [33]. Finally, the nomogram constructed using the data from the TCGA cohort was validated in the CGGA cohort.

Data accessibility

The scRNA-seq data of GBM samples (accession number GSE84465) were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). The bulk RNA-seq profiles of GBM samples were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>) and the CGGA database (<http://www.cgga.org.cn>).

Ethics approval

Ethics committee approval for our study was not required because the data were obtained from publicly available databases.

AUTHOR CONTRIBUTIONS

ZW, XG and LG performed the data curation and analysis. YW and WM analyzed and interpreted the results. ZW, XG and BX drafted and reviewed the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Dr. Wang and Dr. Guo are grateful for the invaluable support received from Professor Bing Xing over the years.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

FUNDING

This study was supported by the Graduate Innovation Fund of Peking Union Medical College (NO. 2019-1002-73 from ZW and NO. 2017-1002-2-18 from XG). The sponsors had no role in the design, execution, interpretation, or writing of the study.

REFERENCES

1. Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012-2016. Neuro Oncol. 2019 (Suppl 5); 21:v1–100.
<https://doi.org/10.1093/neuonc/noz150>
PMID:[31675094](#)
2. Omuro A, DeAngelis LM. Glioblastoma and other Malignant gliomas: a clinical review. JAMA. 2013; 310:1842–50.
<https://doi.org/10.1001/jama.2013.280319>
PMID:[24193082](#)

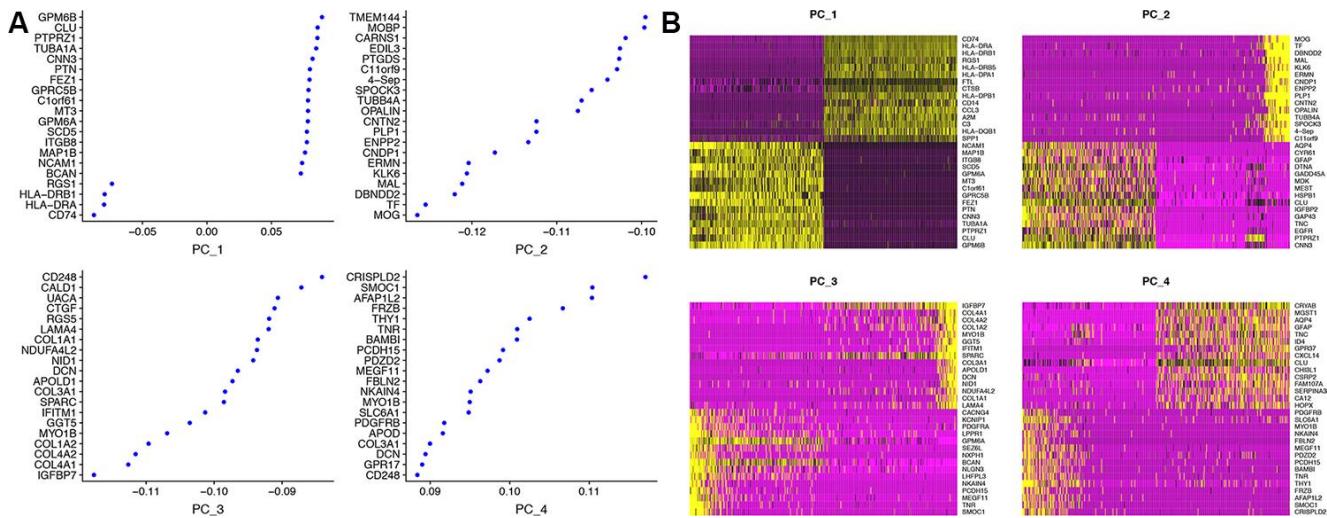
3. Alexander BM, Cloughesy TF. Adult glioblastoma. *J Clin Oncol.* 2017; 35:2402–09.
<https://doi.org/10.1200/JCO.2017.73.0119>
PMID:[28640706](#)
4. Delgado-López PD, Corrales-García EM. Survival in glioblastoma: a review on the impact of treatment modalities. *Clin Transl Oncol.* 2016; 18:1062–71.
<https://doi.org/10.1007/s12094-016-1497-x>
PMID:[26960561](#)
5. Medikonda R, Dunn G, Rahman M, Fecchi P, Lim M. A review of glioblastoma immunotherapy. *J Neurooncol.* 2020. [Epub ahead of print].
<https://doi.org/10.1007/s11060-020-03448-1>
PMID:[32253714](#)
6. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, Lang FF, McCutcheon IE, Hassenbusch SJ, Holland E, Hess K, Michael C, Miller D, Sawaya R. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg.* 2001; 95:190–98.
<https://doi.org/10.3171/jns.2001.95.2.0190>
PMID:[11780887](#)
7. Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatry D, Feng B, Sinha R, Louis DN, Quackenbush J, Black PM, Chin L, DePinho RA. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res.* 2006; 66: 11502–13.
<https://doi.org/10.1158/0008-5472.CAN-06-2072>
PMID:[17114236](#)
8. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science.* 2008; 321:1807–12.
<https://doi.org/10.1126/science.1164382>
PMID:[18772396](#)
9. Wang Z, Gao L, Guo X, Feng C, Lian W, Deng K, Xing B. Development and validation of a nomogram with an autophagy-related gene signature for predicting survival in patients with glioblastoma. *Aging (Albany NY).* 2019; 11:12246–69.
<https://doi.org/10.18632/aging.102566>
PMID:[31844032](#)
10. Lumertz da Rocha E, Rowe RG, Lundin V, Malleshaiah M, Jha DK, Rambo CR, Li H, North TE, Collins JJ, Daley GQ. Reconstruction of complex single-cell trajectories using CellRouter. *Nat Commun.* 2018; 9:892.
<https://doi.org/10.1038/s41467-018-03214-y>
PMID:[29497036](#)
11. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature.* 2017; 541:331–38.
<https://doi.org/10.1038/nature21350>
PMID:[28102262](#)
12. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol.* 2016; 34:1145–60.
<https://doi.org/10.1038/nbt.3711>
PMID:[27824854](#)
13. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science.* 2019; 365:eaav9314.
<https://doi.org/10.1126/science.aav9314>
PMID:[31346039](#)
14. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods.* 2017; 14:979–82.
<https://doi.org/10.1038/nmeth.4402> PMID:[28825705](#)
15. Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringnér M, Bergh J, Björklund Å, Pietras K. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun.* 2018; 9:5150.
<https://doi.org/10.1038/s41467-018-07582-3>
PMID:[30514914](#)
16. Cai Y, Dai Y, Wang Y, Yang Q, Guo J, Wei C, Chen W, Huang H, Zhu J, Zhang C, Zheng W, Wen Z, Liu H, et al. Single-cell transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis. *EBioMedicine.* 2020; 53:102686.
<https://doi.org/10.1016/j.ebiom.2020.102686>
PMID:[32114394](#)
17. Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, Cole MB, Flores Q, Choi YG, Yosef N, Purdom E, Dudoit S, Risso D, Ngai J. Deconstructing olfactory stem cell trajectories at single-cell resolution. *Cell Stem Cell.* 2017; 20:817–30.e8.
<https://doi.org/10.1016/j.stem.2017.04.003>
PMID:[28506465](#)
18. Praktikno SD, Obermayer B, Zhu Q, Fang L, Liu H, Quinn H, Stoeckius M, Kocks C, Birchmeier W, Rajewsky N. Tracing tumorigenesis in a solid tumor model at single-cell resolution. *Nat Commun.* 2020; 11:991.
<https://doi.org/10.1038/s41467-020-14777-0>
PMID:[32080185](#)
19. Tsoi J, Robert L, Paraiso K, Galvan C, Sheu KM, Lay J, Wong DJ, Atefi M, Shirazi R, Wang X, Braas D, Grasso

- CS, Palaskas N, et al. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell.* 2018; 33:890–904.e5.
<https://doi.org/10.1016/j.ccr.2018.03.017>
PMID:[29657129](#)
20. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74.
<https://doi.org/10.1016/j.cell.2011.02.013>
PMID:[21376230](#)
21. Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, Fronick CC, Fulton RS, Church DM, Ley TJ. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat Commun.* 2019; 10:3660.
<https://doi.org/10.1038/s41467-019-11591-1>
PMID:[31413257](#)
22. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suvà ML, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* 2014; 344:1396–401.
<https://doi.org/10.1126/science.1254257>
PMID:[24925914](#)
23. Qazi MA, Vora P, Venugopal C, Sidhu SS, Moffat J, Swanton C, Singh SK. Intratumoral heterogeneity: pathways to treatment resistance and relapse in human glioblastoma. *Ann Oncol.* 2017; 28:1448–56.
<https://doi.org/10.1093/annonc/mdx169>
PMID:[28407030](#)
24. Guichet PO, Guelfi S, Teigell M, Hoppe L, Bakalara N, Bauchet L, Duffau H, Lamszus K, Rothhut B, Hugnot JP. Notch1 stimulation induces a vascularization switch with pericyte-like cell differentiation of glioblastoma stem cells. *Stem Cells.* 2015; 33:21–34.
<https://doi.org/10.1002/stem.1767>
PMID:[24898819](#)
25. Hu B, Wang Q, Wang YA, Hua S, Sauvé CG, Ong D, Lan ZD, Chang Q, Ho YW, Monasterio MM, Lu X, Zhong Y, Zhang J, et al. Epigenetic activation of Wnt5A drives glioblastoma stem cell differentiation and invasive growth. *Cell.* 2016; 167:1281–95.e18.
<https://doi.org/10.1016/j.cell.2016.10.039>
PMID:[27863244](#)
26. Rajakulendran N, Rowland KJ, Selvadurai HJ, Ahmadi M, Park NI, Naumenko S, Dolma S, Ward RJ, So M, Lee L, MacLeod G, Pasiliao C, Brandon C, et al. Wnt and notch signaling govern self-renewal and differentiation in a subset of human glioblastoma stem cells. *Genes Dev.* 2019; 33:498–510.
<https://doi.org/10.1101/gad.321968.118>
PMID:[30842215](#)
27. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* 2018; 15:81–94.
<https://doi.org/10.1038/nrclinonc.2017.166>
PMID:[29115304](#)
28. Martinez-Lage M, Lynch TM, Bi Y, Cocito C, Way GP, Pal S, Haller J, Yan RE, Ziobor A, Nguyen A, Kandpal M, O'Rourke DM, Greenfield JP, et al. Immune landscapes associated with different glioblastoma molecular subtypes. *Acta Neuropathol Commun.* 2019; 7:203.
<https://doi.org/10.1186/s40478-019-0803-6>
PMID:[31815646](#)
29. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu CJ, et al, and TCGA Research Network. The somatic genomic landscape of glioblastoma. *Cell.* 2013; 155:462–77.
<https://doi.org/10.1016/j.cell.2013.09.034>
PMID:[24120142](#)
30. Eskilsson E, Røslund GV, Solecki G, Wang Q, Harter PN, Graziani G, Verhaak RG, Winkler F, Bjerkvig R, Miletic H. EGFR heterogeneity and implications for therapeutic intervention in glioblastoma. *Neuro Oncol.* 2018; 20:743–52.
<https://doi.org/10.1093/neuonc/nox191>
PMID:[29040782](#)
31. Furnari FB, Cloughesy TF, Cavenee WK, Mischel PS. Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. *Nat Rev Cancer.* 2015; 15:302–10.
<https://doi.org/10.1038/nrc3918> PMID:[25855404](#)
32. Weenink B, French PJ, Silleveld Smitt PA, Debets R, Geurts M. Immunotherapy in glioblastoma: current shortcomings and future perspectives. *Cancers (Basel).* 2020; 12:751.
<https://doi.org/10.3390/cancers12030751>
PMID:[32235752](#)
33. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996; 15:361–87.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
PMID:[8668867](#)
34. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, Li G, Chang SD, Connolly ID, et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* 2017; 21:1399–410.
<https://doi.org/10.1016/j.celrep.2017.10.030>
PMID:[29091775](#)

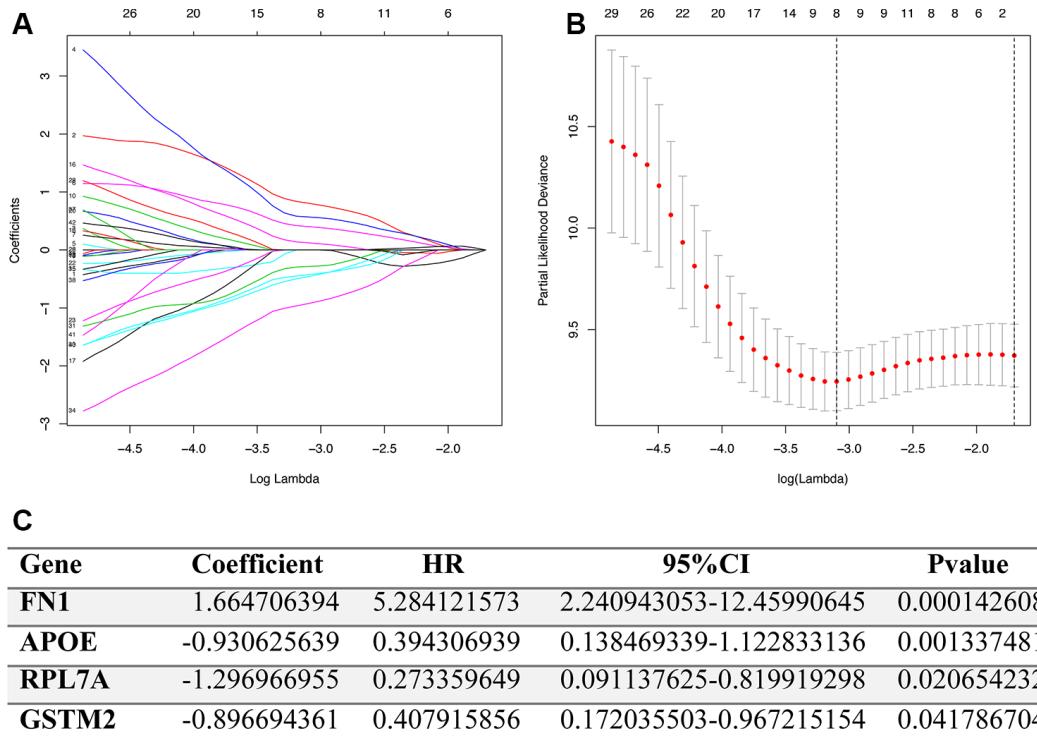
35. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018; 36:411–20.
<https://doi.org/10.1038/nbt.4096>
PMID:[29608179](#)
36. Lall S, Sinha D, Bandyopadhyay S, Sengupta D. Structure-aware principal component analysis for single-cell RNA-seq data. *J Comput Biol.* 2018. [Epub ahead of print].
<https://doi.org/10.1089/cmb.2018.0027>
PMID:[30133312](#)
37. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015; 33:495–502.
<https://doi.org/10.1038/nbt.3192>
PMID:[25867923](#)
38. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* 2019; 20:163–72.
<https://doi.org/10.1038/s41590-018-0276-y>
PMID:[30643263](#)
39. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 2019; 47:D721–28.
<https://doi.org/10.1093/nar/gky900>
PMID:[30289549](#)
40. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–50.
41. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010; 26:1572–73.
<https://doi.org/10.1093/bioinformatics/btq170>
PMID:[20427518](#)
42. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018; 24:1550–58.
<https://doi.org/10.1038/s41591-018-0136-1>
PMID:[30127393](#)
43. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One.* 2007; 2:e1195.
<https://doi.org/10.1371/journal.pone.0001195>
PMID:[18030330](#)
44. Nagashima K, Sato Y. Information criteria for firth's penalized partial likelihood approach in cox regression models. *Stat Med.* 2017; 36:3422–36.
<https://doi.org/10.1002/sim.7368>
PMID:[28608396](#)
45. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA.* 2017; 318:1377–84.
<https://doi.org/10.1001/jama.2017.12126>
PMID:[29049590](#)

SUPPLEMENTARY MATERIALS

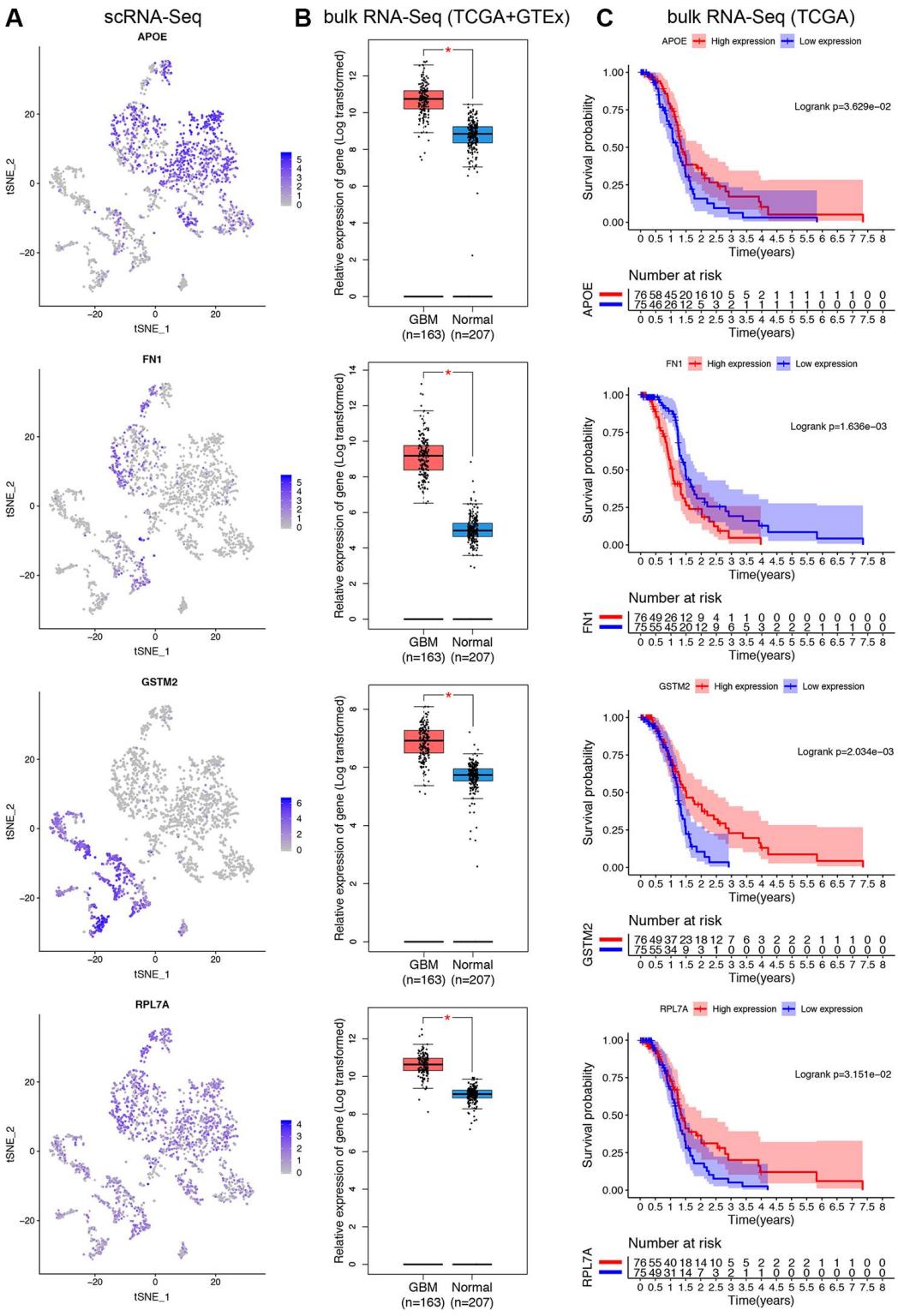
Supplementary Figures



Supplementary Figure 1. The top 20 significantly correlated genes in PC_1 to PC_4 are displayed using dot plots and heatmaps. (A) The dot plots show the top 20 significantly correlated genes in each component. The X axis represents the correlation coefficient. (B) The heatmaps show the expression patterns of the top 20 significantly correlated genes in each component. The colors from purple to yellow indicate the gene expression levels from low to high.



Supplementary Figure 2. The prognosis-related GDRGs were screened with LASSO and multivariate Cox regression analyses. (A) The coefficient profile plot was produced against the $\log(\lambda)$ sequence. A vertical line was drawn at the value selected using ten-fold cross-validation, where an optimal lambda value resulted in ten features with nonzero coefficients. (B) Optimal parameter (λ) selection in the LASSO model used ten-fold cross-validation via minimum criteria. The partial likelihood deviance (binomial deviance) curve was plotted versus the $\log(\lambda)$ value. Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria. (C) Following LASSO regression analysis, 4 GDRGs were identified with multivariate Cox regression analysis.



Supplementary Figure 3. Expression and survival analyses of the four GDRGs in the prediction model. (A) The expression of the four GDRGs in scRNA-seq data is shown. The colors from gray to blue indicate the gene expression levels from low to high. (B) The expression of the four GDRGs in bulk RNA-seq data, including 163 GBM (TCGA) and 207 normal (GTEX) samples, is shown. Asterisks indicate $p < 0.05$. (C) K-M survival analyses demonstrated that the expression of GDRGs was associated with OS in TCGA GBM patients.

Supplementary Tables

Supplementary Table 1. The marker genes of type I and type II GBM cell subsets.

GBM cell subsets	GBM cell differentiation-related genes (GDRGs)
Type I GDRGs (n=265)	A2M, ABL2, ACSL1, ADAM28, ADAP2, ADORA3, AIF1, AKAP13, ALOX5, ALOX5AP, ANXA1, APBB1IP, APOC1, APOE, ARAP1, ARHGAP18, ARHGAP4, ARHGDB, ARNT2, ARPC1B, ARPC2, ATP6V0E2, B3GNT5, B4GALT1, BAALC, BCL2A1, C10orf54, C1orf162, C1orf38, C1QA, C1QB, C1QC, C3, C3AR1, C5AR1, CAPG, CCL3, CCL4, CD14, CD163, CD300A, CD4, CD53, CD68, CD74, CD83, CD84, CD86, CD93, CDKN1A, CHL1, CLEC7A, CORO1A, CPE, CSF1R, CSF3R, CTSB, CTSC, CTSH, CTSI1, CTSS, CXCR4, CYBA, CYBB, CYTH4, DAG1, DDR1, DENND3, DKK3, DOCK2, DOCK8, DSE, ELL2, ENG, ETS2, ETV1, EVI2B, F13A1, FCER1G, FCGBP, FCGR1A, FCGR2A, FCGR3A, FCGRT, FERMT3, FGD4, FGL2, FHL1, FPR1, FTH1, FTL, FXYD5, FYB, FYN, GATM, GK, GMFG, GNA13, GOLM1, GPM6B, GPNMB, GPR183, GPRC5B, GPX1, HAMP, HAVCR2, HCK, HCLS1, HLA-B, HLA-DMA, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HMHA1, IER3, IFI30, IFNGR1, IL10RA, IL13RA1, IL18, IL8, INPP5D, IRF8, ITGAM, ITGAX, ITGB2, ITGB8, ITM2C, ITPRIP, KCTD12, KLF6, LAIR1, LAPTM5, LAT2, LCP1, LCP2, LGALS9, LILRB4, LIMS1, LPCAT2, LST1, LY86, LYN, LYZ, MAGED1, MANBA, MAP3K8, MEF2C, MERTK, METRNL, MGAT1, MPP1, MS4A4A, MS4A6A, MS4A7, MSR1, MYO1F, MYO9B, NAV2, NCF4, NFKB1D, NINJ1, NPC2, NR4A2, NR4A3, NRCAM, NTRK2, OLR1, OTUD1, PABPC1, PABPC4, PARVG, PFN1, PIK3AP1, PIK3R5, PLAUR, PLCB2, PLEK, PLIN2, PLXDC2, PMP2, PTAFR, PTGER4, PTPN6, PTPRC, PTPRZ1, RANBP2, RAPGEF1, REL, RGS1, RGS10, RBDF2, RHOG, RILPL2, RIN3, RNASE6, RNASET2, RNF144B, RNF149, RPL23, RPS23, RPS24, RPS3, RPS6, S100A11, S100A4, SAMSN1, SAT1, SCIN, SCRG1, SDC3, SELPLG, SERPINA1, SERPINB9, SERPINE2, SH3BGRL3, SH3TC1, SIGLEC10, SLA, SLC11A1, SLC1A2, SLC22A17, SLC2A5, SLC31A2, SLCO2B1, SOD2, SORBS1, SPARCL1, SPP1, SRGN, STAB1, STAT6, SYK, SYNGR2, SYT11, TBXAS1, TCIRG1, TFRC, TGFBI, THBS2, TLR2, TNFRSF1B, TPM3, TPT1, TREM1, TREM2, TRIM9, TRO, TSPAN3, TSPAN7, TYMP, TYROBP, UCHL1, UCP2, UPP1, USP36, VAMP8, VSIG4, ZBTB20, ZNF331
Type II GDRGs (n=193)	ANPEP, RNASE1, ADAM8, GPNMB, MARCO, FTL, FCGR2B, C15orf48, LSP1, EGR1, TGFBI, S100A10, S100A4, LYZ, TMSB10, LGALS1, S100A9, CD109, SLC16A10, FTH1, ANXA2, VIM, TPT1, CSTB, IFI30, CYTIP, FBP1, BCAT1, GAPDH, SH3BGRL3, LCP1, GPR56, LDHA, RPS18, NDRG1, FLNB, TGM2, ERO1L, PLIN2, VCAN, TNFRSF14, SLC1A3, PAPSS2, RPS24, ENO1, NCF2, TCIRG1, BTG1, TXNIP, LGALS3, AHNAK, CTSD, S100A6, UPP1, ALOX5AP, BRI3, HK2, TYROBP, CLEC5A, TREM1, CD163, RPS6, DBI, FN1, PKM2, FOS, ANXA1, LAPTM5, CD44, S100A11, RPS2, RPL37, TUBGCP2, RPS9, TAGLN2, YBX1, METRNL, RPL37A, RPS3, TRIM25, RPL27A, ALDOA, RPL31, C1orf61, CNN3, ST14, CXCL2, PTN, CLU, CD68, RPL23, CKB, RPS4Y1, TYMP, PTK2B, EEF1A1, MGAT1, RPL9, RPL13, JARID2, S100A8, NFIA, MAP1B, ARHGAP18, RPS12, MAPK13, RPL32, MMP19, NCF1, ITGAM, RPL12, CAPG, PTPRZ1, PON2, GPM6B, UBA52, SPP1, BNIP3, TBC1D2, TJP1, EIF4EBP1, NFIB, PGK1, OAZ1, GFAP, GSTO1, RPLP0, ATP1B2, EEF1G, RPS23, ID2, GPM6A, G0S2, CALM2, HMOX1, FNDC3B, TUBB2B, EGFR, DMXL2, GSTM2, BNIP3L, CALD1, TNS1, RPS8, CTSB, H2AFY, C3, LRRFIP1, BTG2, PMP2, FEZ1, ITGB8, CHL1, ARPC2, P4HB, NRCAM, GPI, SLC39A8, RIPK2, FGR, ANGPTL4, AASS, CTSL1, ITM2C, ITGAX, SLC11A1, RPSA, SRI, RPL8, SCD5, GPRC5B, EGR3, RPL7A, SLCO4A1, RPL13A, GRB2, SCRG1, CX3CR1, GLUL, BCAN, ZBTB20, DDX5, DTNA, DHRS3, SDS, MEST, F3, IL4I1, MT3, FHL1, DDR1, TCF4, COTL1, FXYD6, PFN2, ABCA1, XIST, TUBA1A, CBX5, IER2, MAGED1, RAMP1, PIM1, SDC2, GSTM3, HTRA1, CRYAB, SLC2A1, MARCKSL1, SPRY2, MDK, KIAA1274, CHD7, GATM, CH25H, OLFML3, DDX41, TSPAN3, LONP1, PADI2, SPTBN1, TMEM9, HSPA6, NDUFA4, RCN2, NDRG2, WRB, ZFP36, LIMA1, STMN1, DPYSL2, SBDS, NGFRAP1, FAM46A, JUNB, SPARC, SKP1, SERPING1, FOSB, SRGAP1, SYT11, GADD45A, SDC3, GABARAPL2, AHCYL1, NUPR1, METTL9, DUSP1, EGR2, ITPR2, CBR1, PGRCMC1, S100B, CCND2, TUBA1B, TSPAN31, PRDX2, TSC22D4, LPL, MLF2, TMX2, 7-Sep, RBBP4, PEPB1, RHBDD2, EIF4E, NR4A3, SPARCL1, ST6GAL1, RTN3, CCL4, EID1, CASC4, ATP5A1, TSC22D1, LDHB, SOD1, CCT6A, CCL2, CDK4, KPNA2, RASSF4, CLDND1, A2M, SORL1, LGALS3BP, CNP, CREB5, GBAS, HSPB1, IL1B, CLIC4, NR4A1
Total GDRGs (n=498)	ANPEP, RNASE1, ADAM8, GPNMB, MARCO, FTL, FCGR2B, C15orf48, LSP1, EGR1, TGFBI, S100A10, S100A4, LYZ, TMSB10, LGALS1, S100A9, CD109, SLC16A10, FTH1, ANXA2, VIM, TPT1, CSTB, IFI30, CYTIP, FBP1, BCAT1, GAPDH, SH3BGRL3, LCP1, GPR56, LDHA, RPS18, NDRG1, FLNB, TGM2, ERO1L, PLIN2, VCAN, TNFRSF14, SLC1A3, PAPSS2, RPS24, ENO1, NCF2, TCIRG1, BTG1, TXNIP, LGALS3, AHNAK, CTSD, S100A6, UPP1, ALOX5AP, BRI3, HK2, TYROBP, CLEC5A, TREM1, CD163, RPS6, DBI, FN1, PKM2, FOS, ANXA1, LAPTM5, CD44, S100A11, RPS2, RPL37, TUBGCP2, RPS9, TAGLN2, YBX1, METRNL, RPL37A, RPS3, TRIM25, RPL27A, ALDOA, RPL31, C1orf61, CNN3, ST14, CXCL2, PTN, CLU, CD68, RPL23, CKB, RPS4Y1, TYMP, PTK2B, EEF1A1, MGAT1, RPL9, RPL13, JARID2, S100A8, NFIA, MAP1B, ARHGAP18, RPS12, MAPK13, RPL32, MMP19, NCF1, ITGAM, RPL12, CAPG, PTPRZ1, PON2, GPM6B, UBA52, SPP1, BNIP3, TBC1D2, TJP1, EIF4EBP1, NFIB, PGK1, OAZ1, GFAP, GSTO1, RPLP0, ATP1B2, EEF1G, RPS23, ID2, GPM6A, G0S2, CALM2, HMOX1, FNDC3B, TUBB2B, EGFR, DMXL2, GSTM2, BNIP3L, CALD1, TNS1, RPS8, CTSB, H2AFY, C3, LRRFIP1, BTG2, PMP2, FEZ1, ITGB8, CHL1, ARPC2, P4HB, NRCAM, GPI, SLC39A8, RIPK2, FGR, ANGPTL4, AASS, CTSL1, ITM2C, ITGAX, SLC11A1, RPSA, SRI, RPL8, SCD5, GPRC5B, EGR3, RPL7A, SLCO4A1, RPL13A, GRB2, SCRG1, CX3CR1, GLUL, BCAN, ZBTB20, DDX5, DTNA, DHRS3, SDS, MEST, F3, IL4I1, MT3, FHL1, DDR1, TCF4, COTL1, FXYD6, PFN2, ABCA1, XIST, TUBA1A, CBX5, IER2, MAGED1, RAMP1, PIM1, SDC2, GSTM3,

HTRA1, CRYAB, SLC2A1, MARCKSL1, SPRY2, MDK, KIAA1274, CHD7, GATM, CH25H, OLFML3, DDX41, TSPAN3, LONP1, PADI2, SPTBN1, TMEM9, HSPA6, NDUFA4, RCN2, NDRG2, WRB, ZFP36, LIMA1, STMN1, DPYSL2, SBDS, NGFRAP1, FAM46A, JUNB, SPARC, SKP1, SERPING1, FOSB, SRGAP1, SYT11, GADD45A, SDC3, GABARPL2, AHCYL1, NUPR1, METTL9, DUSP1, EGR2, ITPR2, CBR1, PGRMC1, S100B, CCND2, TUBA1B, TSPAN31, PRDX2, TSC22D4, LPL, MLF2, TMX2, 7-Sep, RBBP4, PEBP1, RHBDD2, EIF4E, NR4A3, SPARCL1, ST6GAL1, RTN3, CCL4, EID1, CASC4, ATP5A1, TSC22D1, LDHB, SOD1, CCT6A, CCL2, CDK4, KPNA2, RASSF4, CLDND1, A2M, SORL1, LGALS3BP, CNP, CREB5, GBAS, HSPB1, IL1B, CLIC4, NR4A1, ABL2, ACSL1, ADAM28, ADAP2, ADORA3, AIF1, AKAP13, ALOX5, APBB1IP, APOC1, APOE, ARAP1, ARHGAP4, ARHGAP4, ARNT2, ARPC1B, ATP6V0E2, B3GNT5, B4GALT1, BAALC, BCL2A1, C10orf54, C1orf162, C1orf38, C1QA, C1QB, C1QC, C3AR1, C5AR1, CCL3, CD14, CD300A, CD4, CD53, CD74, CD83, CD84, CD86, CD93, CDKN1A, CLEC7A, CORO1A, CPE, CSF1R, CSF3R, CTSC, CTSH, CTSS, CXCR4, CYBA, CYBB, CYTH4, DAG1, DENND3, DKK3, DOCK2, DOCK8, DSE, ELL2, ENG, ETS2, ETV1, EVI2B, F13A1, FCER1G, FCGBP, FCGR1A, FCGR2A, FCGR3A, FCGRT, FERMT3, FGD4, FGL2, FPR1, FXYD5, FYB, FYN, GK, GMFG, GNA13, GOLM1, GPR183, GPX1, HAMP, HAVCR2, HCK, HCLS1, HLA-B, HLA-DMA, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HMHA1, IER3, IFNGR1, IL10RA, IL13RA1, IL18, IL8, INPP5D, IRF8, ITGB2, ITPRIP, KCTD12, KLF6, LAIR1, LAT2, LCP2, LGALS9, LILRB4, LIMS1, LPCAT2, LST1, LY86, LYN, MANBA, MAP3K8, MEF2C, MERTK, MPP1, MS4A4A, MS4A6A, MS4A7, MSR1, MYO1F, MYO9B, NAV2, NCF4, NFKBID, NINJ1, NPC2, NR4A2, NTRK2, OLR1, OTUD1, PABPC1, PABPC4, PARVG, PFN1, PIK3AP1, PIK3R5, PLAUR, PLCB2, PLEK, PLXDC2, PTAFR, PTGER4, PTPN6, PTPRC, RANBP2, RAPGEF1, REL, RGS1, RGS10, RHDF2, RHOG, RILPL2, RIN3, RNASE6, RNASET2, RNF144B, RNF149, SAMSN1, SAT1, SCIN, SELPLG, SERPINA1, SERPINB9, SERPINE2, SH3TC1, SIGLEC10, SLA, SLC1A2, SLC22A17, SLC2A5, SLC31A2, SLCO2B1, SOD2, SORBS1, SRGN, STAB1, STAT6, SYK, SYNGR2, TBXAS1, TFRC, THBS2, TLR2, TNFRSF1B, TPM3, TREM2, TRIM9, TRO, TSPAN7, UCHL1, UCP2, USP36, VAMP8, VSIG4, ZNF331

Supplementary Table 2. The corresponding genes of cell surface markers (CellMarker database) for identifying cell types of the 13 cell clusters.

Cell populations	Gene symbols of cell surface markers
Astrocyte	ALDH1L1, WIF1, NTSR2, GFAP, SOX9, CD40, CD80, CD86, ABCA1, NFIA, S100B, C5AR1, ACSL6, ADCYAP1R1, AGT, ALDOC, ANLN, APOE, ARHGEF26, ATF3, ATP13A4, ATP1B2, BBOX1, BTG2, C1orf61, CA12, CASQ1, CBS, CDC42EP4, CHST9, CLU, CPE, CPNE5, CPVL, CRYAB, CST3, DAND5, DGKG, DKK3, DNJB1, DNASE2, DOK5, DTNA, EDNRB, EEPD1, EFEMP1, EGLN3, EPAS1, EZR, F3, FOS, FOSB, FTH1, FXYD7, GABRB1, GADD45B, CASTOR1, GLUL, GPR75, GRAMD2B, GRIA1, HEPACAM, HEPN1, HEY1, HIF1A, HIF3A, HLA-E, HNMT, HSPB8, ID1, ID4, IER2, IGFBP7, IL11RA, IL33, JUN, JUNB, JUND, KCNH7, KCNIP2, KIF21A, L1CAM, LCAT, LHFPL6, LIX1, LPL, LRIG1, LRRC8A, LYPD1, MAFB, METTL7A, MLC1, MMD2, MT1X, MT2A, MTHFD2, NDRG2, NFKBIA, NHL1, NANOG, NRP1, NTRK2, P2RY1, PAPLN, PEA15, PER1, PFKFB3, PLTP, PON2, PRLHR, PRRT2, PSAP, RASL10A, RASSF4, RFX4, RGMA, RHOB, RND3, SCG2, SCG3, SEMA6A, SERPINA3, SF3A1, SLC1A2, SLC1A3, SLC39A11, SORL1, SPARCL1, SPOCK1, SPON1, SRPX, SSTR2, ST6GAL2, TACR1, TBC1D10A, TIMP3, TMEM151B, TNS1, TOB2, TPCN1, TRIL, TSC22D4, TSPAN12, ZFAND5, ZFP36, ZFP36L1, ZFP36L2, AQP4, BMPR1B, C16orf89, CHRDL1, CTH, CYBRD1, FGFR3, GLI3, GLIS3, HGF, ITGA7, ITGB4, NWD1, PAQR6, PPP1R3C, RNF43, SLC14A1, SLC30A10, SLC4A4, SORCS2, TRIM66
Oligodendrocyte	MBP, OPALIN, RTN4, CSPG4, PDGFRA, OLIG2, ACAT2, ADAMTS4, LHFPL3, ACTG1, AFAP1L2, AMOTL2, ANGPTL2, APOD, ARL4A, ASIC1, ATCAY, ATP5F1E, BIN1, LRRCT5A, C2orf27A, CDH13, CDHR1, CNP, COX7C, CRB1, CXADR, DHCR24, DHCRT7, DLL1, DLL3, EBP, EEF1B2, EEF2, EIF3E, EIF3L, EPN2, FA2H, FABP7, FAU, FDPS, FERMT1, FGFBP3, FXYD6, GAP43, RACK1, GPR17, GPR37L1, GRIA2, GRIA4, GRIK2, HIP1, HIPK2, IFITM10, KCNIP3, KLRC2, LDHB, LIMA1, LIMS2, LMFI, PLPPR1, LRRN1, MAML2, MAP1A, MAP2, MARCKSL1, MEST, MICAL1, MIF, MTSS1, MYT1, NACA, NAP1L1, NEU4, NME1, NME2, NPM1, NPPA, NXPH1, OLIG1, OMG, OPCML, P2RX7, PGRMC1, PHACTR3, PHLDA1, PHLDB1, PID1, POLR2F, RAB2A, RAB33A, RGMB, RGR, RPL13A, RPL31, RPS17, RPS2, RPS23, RPSAP58, RTKN, SCD, SERINC5, SGK1, SHD, SHISA4, SIRT2, SLC1A1, SNX1, SNX22, SOX8, TAGLN3, TCF12, THY1, TM7SF2, TMEFF2, TMEM97, TMSB10, TMSB4X, TNK2, TRAF4, TUBB, TUBB3, TUBB4A, UGT8, UQCRB, VIPR2, WSCD1, ZCCHC24, ZDHHC9, ZEB2, ASPA, CDK18, CLDN11, DAAM2, DPYD, ERMN, GJB1, GPR37, GRM3, GSN, KCNH8, KLK6, LGI3, LPAR1, MAG, MAL, MAP6D1, MOBP, PLEKH1, PLP1, PPP1R14A, SEC14L5, SHC4, MEGF11, PCDH15
Neuron	L1CAM, DCLK3, RBFOX3, GFAP, MAP2, NES, MARK4, ENO2, SYP, ARMH4, CDO1, CNTN4, COBL, DCN, DLX1, DLX2, DLX5, DLX6, DPYSL5, KDR, GLRA2, GRIA3, KCNK1, KIAA1324, LNX1, LRRTM3, NELL1, NFASC, NXPH1, PNOC, RELN, SLC10A4, SLITRK1, SST, TMEM130
Glia cell	GFAP, S100 family
GBM stem cell (GSC)	PROM1, CADM1, CLCC1, HMOX1, SCAMP3, SLC16A1, FUT4, CDH5, NANOG, NES, SOX2, THY1, MSI1, CD133, CD15, L1CAM, CD90, A2B5, POU3F2, OLIG2, SALL2
GBM cell	PARP1, MBTPS2, CD44
Endothelial cell	PECAM1, CDH5, VWF, VCAM1, A2M, APOLD1, FLT1, TM4SF1
B cell	MS4A1, CD19, CD79A
T cell (general)	CD3D, CD3E, CD3G, CD2
Regulatory T (Treg) cell	IL2RA, CD4, FOXP3, CCR8, STAT5B, TGFB1
T helper-1 (Th1) cell	CCR6, CXCR3, TBX21, STAT4, STAT1, IFNG, TNF
T helper-2 (Th2) cell	GATA3, STAT6, STAT5A, IL13, CCR6, CXCR3, TNFRSF8
T helper-17 (Th17) cell	STAT3, IL17A, CCR6, CXCR3
Tumor-associated macrophage (TAM)	CD68, CCL2, IL10, AIF1, ARG2, BHLHE40, CD74, CD93, CIB1, CIITA, CREM, CYBB, CYTH1, DOK3, DSE, EMB, FAM49A, FGR, FOSL2, FPR3, FXYD5, GPR132, GPR65, HLA-DMB, HLA-DQA1, HLA-DRB5, IFITM2, IL1RN, IQGAP1, ITGA4, KYNU, LYZ, METRNL, MS4A6A, MS4A7, MXD1, NFL3, PDE4B, PIM1, PLAC8, PLBD1, PLTP, PQLC3, PTPN7, S100A11, SAMHD1, SH3BGRL, SPINT2, SYNGR2, TGFBI, THBD, TMEM123, TNFSF13, TREM1, VOPP1
M1 macrophage	TSPO, IRF5, PTGS2, NOS2
M2 macrophage	CD163, VSIG4, MS4A4A
Monocyte	CD86, CSF1R
Neutrophil	MPO, CEACAM8, ITGAM, CCR7
Natural killer (NK) cell	KIR2DL1, KIR2DL3, KIR2DL4, KIR3DL1, KIR3DL2, KIR3DL3, KIR2DS4
Dendritic cell (DC)	HLA-DPB1, HLA-DQB1, HLA-DRA, HLA-DPA1, CD1C, NRP1, ITGAX