Am J Transl Res 2020;12(8):4669-4682
www.ajtr.org /ISSN:1943-8141/AJTR0113150

Original Article

Identification of three glioblastoma subtypes and a six-gene prognostic risk index based on the expression of growth factors and cytokines

Beiquan Hu[1,2*], Yushan Ruan[3*], Feng Wei[2*], Gang Qin[2], Xianlun Mo[2], Xiangyu Wang[1], Donghua Zou[4]

[1]Department of Neurosurgery, The First Affiliated Hospital, Jinan University, Guangzhou 510630, Guangdong, People's Republic of China; [2]Department of Neurosurgery, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning 530022, Guangxi, People's Republic of China; [3]Department of Neurosurgery, The Second Affiliated Hospital of Guangxi Medical University, Nanning 530000, Guangxi, People's Republic of China; [4]Department of Neurology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning 530022, Guangxi, People's Republic of China. [*]Equal contributors.

Received April 23, 2020; Accepted July 22, 2020; Epub August 15, 2020; Published August 30, 2020

Abstract: Glioblastoma multiforme (GBM) is the most common and invasive tumor of the central nervous system. Growth factors and cytokines (GFCKs) play a crucial role in tumor invasion. In the present study, GFCK expression profiles from GBM patients in the Chinese Glioma Genome Atlas were used to perform sample clustering with non-negative matrix factorization. Three GBM subtypes were identified based on differences in GFCK expression, and the subtypes differed in characteristics and prognosis. A prognostic risk index (RI) comprising six GFCKs (BMP2, CCN3, GKN1, LIF, MDK, and SEMA3G) was defined using univariate Cox hazard analysis and multivariate stepwise Cox regression. The RI was validated in two independent data sets and may be independent of some known prognostic factors. Our results suggest that GBM occurs as different subtypes expressing different patterns of GFCKs and that these expression patterns can be captured in an RI that can predict prognosis.

Keywords: Glioblastoma multiforme, growth factor, cytokine, molecular subtyping, prognosis

Introduction

Glioblastoma (GBM) accounts for 16% of all primary brain tumors, making it the most common primary malignant brain tumor [1]. Risk factors for GBM remain poorly understood; the only well-known cause is ionizing radiation [2]. Similarly, little progress has been made in prolonging survival of patients with GBM. The disease remains one of the deadliest types of tumor worldwide; the rate of recurrence is high and prognosis is dismal, with a 5-year survival rate of approximately 5% [3]. Standard management for patients with GBM is surgical resection followed by radio- and chemotherapy [4, 5]. Decades of research into new therapies have largely failed because GBM is so heterogeneous and invasive [6-9].

More effective treatment of GBM may depend on defining tumor subtypes that may be particularly susceptible to certain therapies [10-13]. We wondered whether it would be possible to define GBM subtypes based on expression of growth factors and cytokines (GFCKs), since these molecules drive tumor cell invasion [17]. Such invasion makes it difficult to achieve curative effects even with extensive surgical resection [14, 15]: cancer cells may already have infiltrated unresected areas of the brain, where they may form new tumors after surgery [16]. Inhibiting GFCK receptors can transiently slow progression of lung cancer [18], colorectal cancer [19] and hepatocellular carcinoma [20]. Already one anti-GFCK therapy against GBM has been approved by the US Food and Drug Administration: the therapeutic antibody bevacizumab targets the vascular endothelial growth factor (VEGF) receptor [21].

We hypothesized that analyzing the expression pattern of GFCKs in GBM might allow us to dif-

ferentiate subtypes, which may allow stratification of patients by prognosis and response to therapies.

Materials and methods

Data processing

The glioma data sets mRNAseq_693 and mRNAseq_325 (last update: November 28, 2019) containing mRNA sequencing (mRNA-seq) data and clinical information were downloaded from the Chinese Glioma Genome Atlas (http://cgga.org.cn/index.jsp). A third GBM data set in The Cancer Genome Atlas (TCGA, https://www.cancer.gov/) was downloaded from UCSC Xena (http://xena.ucsc.edu/public); this data set contained clinical information and gene expression profiles based on the Affymetrix Human Genome U133a array platform.

Low-grade glioma samples were removed from the mRNAseq_693 and mRNAseq_325 data sets. Samples without survival time were not included in the survival analysis. The mRNAseq_693 data set was used for molecular subtyping and creating a risk index (RI) for predicting prognosis. The mRNAseq_325 and TCGA_ array data sets were used for validating the prognostic value of the RI. GFCKs in each sample were determined from the Molecular Signature database (version 7, https://www.gsea-msigdb.org). The workflow of the present study is shown in **Figure 1**.

Nonnegative matrix factorization (NMF) and identification of marker genes

Nonnegative matrix factorization (NMF), similar to principal component analysis or independent component analysis, aims to explain the observed data using a limited number of basic components. NMF has been used to computational biology [22], predict cis-regulatory elements [23], characterize gene function [24], predict phenotype from data across different microarray platforms [25], and define cancer subtypes based on gene expression profiles [26]. NMF can be performed faster than principal component or independent component analyses, and it can generate more stable clusters [27]. Thus, NMF was applied in the present study.

GFCKs were ranked by variance based in the mRNAseq_693 data set, and the top 20% of GFCKs showing highest variance were subjected to NMF using the *NMF* package in R (https://cran.r-project.org/package=NMF). The *extractFeatures* function in R was used to define marker genes for the resulting GBM subtypes.

Gene set enrichment analysis (GSEA)

GSEA [28, 29] was applied to explore the potential biological characteristics of each GBM subtype. The GSEA Java software (version 4.0.3) was used, and the reference gene set was the hallmark gene set *h.all.v7.0.symbols.gmt* [30]. Gene sets with a false discovery rate < 0.25 after 1,000 permutations were considered to be significantly enriched.

GFCK expression-based prognostic RI for GBM

Univariate Cox hazard analysis was applied to the mRNAseq_693 data set in order to identify individual GFCKs that might affect overall survival. Then multivariate stepwise Cox regression was performed to establish a GFCK expression-based prognostic RI for each GBM as follows: $RI = Expression_{gene1} \times \beta_{gene1} + Expression_{gene2} \times \beta_{gene2} + Expression_{gene3} \times \beta_{gene3}...$, where β is the estimated regression coefficient of a certain GFCK and is derived from the multivariable Cox regression. Patients with GBM were divided into low- or high-risk groups according to median RI. Time-dependent receiver operating characteristic analysis was used to assess the prognostic value of the RI using the *survivalROC* package in R [31]. We also benchmarked the RI's prognostic value against that of routine clinicopathologic characteristics, including age [32], chemotherapy [4], radiotherapy [33], mutation in the isocitrate dehydrogenase (IDH) gene [32], and 1q19p co-deletion [34].

Statistical analysis

All analyses were performed using R version 3.6.1 (https://www.r-project.org/). Gene expression differences between groups were assessed for significance using the unpaired Student's *t* test. Univariate and multivariate Cox proportional hazard modeling was used to identify prognosis-related variables. Survival was compared between groups of patients using Kaplan-Meier analysis and the log-rank method. Differences associated with $P < 0.05$ were considered significant.

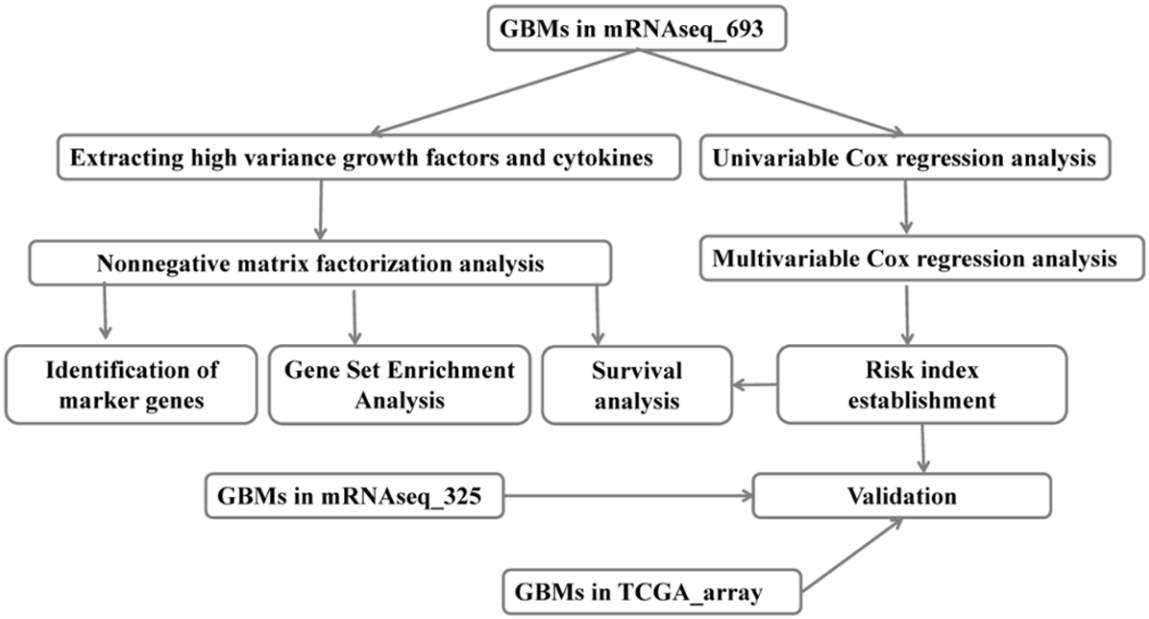


Figure 1. Workflow of the present study. GBM, glioblastoma; TCGA, The Cancer Genome Atlas.

Results

Heterogeneity in GFCK expression may be a basis for defining GBM subtypes

A total of 453 GFCKs were collected from the Molecular Signature database, while 340 GFCKs were detected in the mRNAseq_693 data set. The expression of 68 GFCKs with highest variance were analyzed using NMF (**Figure 2A**). The consensus matrix showed good classification (**Figure 2B**): 237 patients with GBM were divided into three subtypes (**Figure 2C**). Type II GBM was associated with higher frequency of IDH mutations and X1q19p co-deletion than types I or III (**Figure 2D**).

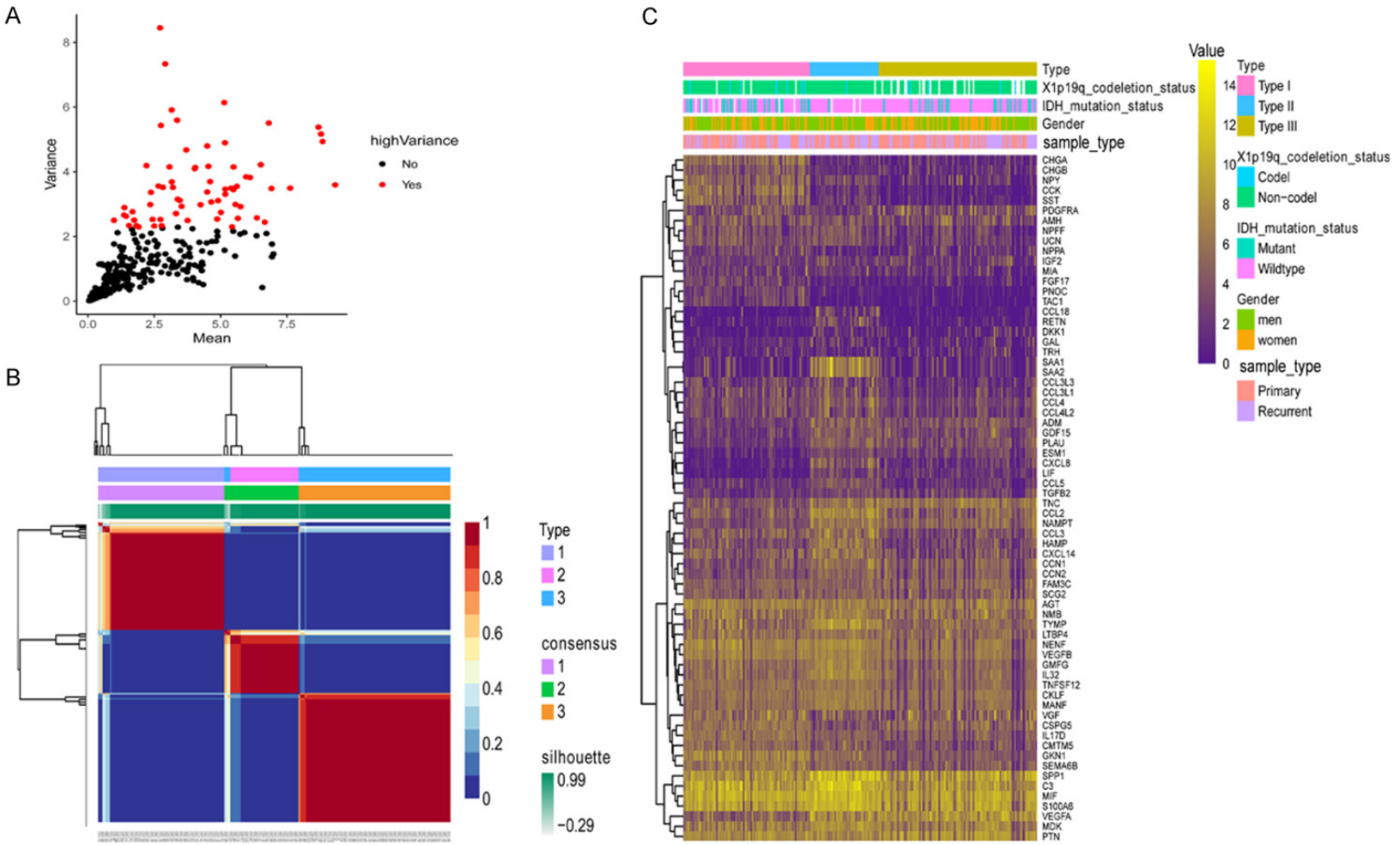
GFCK-based GBM subtypes present different phenotypes

Type I GBM was associated with the marker genes TAC1, CCK, CHGA, PNOC, SST, FGF17, CHGB, and NPY (**Figure 3A**). Type II was associated with the marker genes SAA1/2, RETN, and CXCL8 (**Figure 3B**), and Type III was associated with PDGFRA, ESM1, and CSPG5 (**Figure 3C**). The marker genes identified for Types I and II were expressed at significantly higher levels in those subtypes than in Type III, but the same was not true for marker genes of Type III. Thus, the marker genes for Type III should be considered tentative.

The potential biological characteristics of each GBM subtype were investigated using GSEA. Type I GBM showed significant enrichment in the KRAS signaling down (DN) hallmark gene set (**Figure 4A**), while Type II was enriched in several hallmark gene sets: the epithelial-mesenchymal transition, KRAS signaling up, mTORC1 signaling, P53 pathway, and PI3K/AKT/mTOR signaling (**Figure 4B**; Table S1). Type III GBM showed enrichment in the hallmark gene sets of E2F targets and G2M checkpoint (**Figure 4C**). These differences among the three subtypes suggest different phenotypes. Indeed, we found that prognosis was poorer for Type III patients than for patients with other subtypes (**Figure 4D**).

GFCK-based prognostic RI for GBM

Based on the mRNAseq_693 data set, univariate Cox hazard analysis identified 29 GFCKs as associated with survival (Table S2), and multivariate stepwise Cox regression identified six GFCKs (BMP2, CCN3, GKN1, LIF, MDK, and SEMA3G) as independent prognostic factors. This led us to define a prognostic RI as RI = $\text{Expression}_{\text{BMP2}} \times (-0.3286) + \text{Expression}_{\text{CCN3}} \times 0.15059 + \text{Expression}_{\text{GKN1}} \times (-0.19827) + \text{Expression}_{\text{LIF}} \times (-0.13721) + \text{Expression}_{\text{MDK}} \times 0.1625 + \text{Expression}_{\text{SEMA3G}} \times (-0.36129)$.


A
Variance
Mean
highVariance
No
Yes
B
Type
1
2
3
consensus
1
2
3
silhouette
0.99
−0.29
C
Type
X1p19q_codeletion_status
IDH_mutation_status
Gender
sample_type
Value
Type
Type I
Type II
Type III
X1p19q_codeletion_status
Codel
Non−codel
IDH_mutation_status
Mutant
Wildtype
Gender
men
women
sample_type
Primary
Recurrent

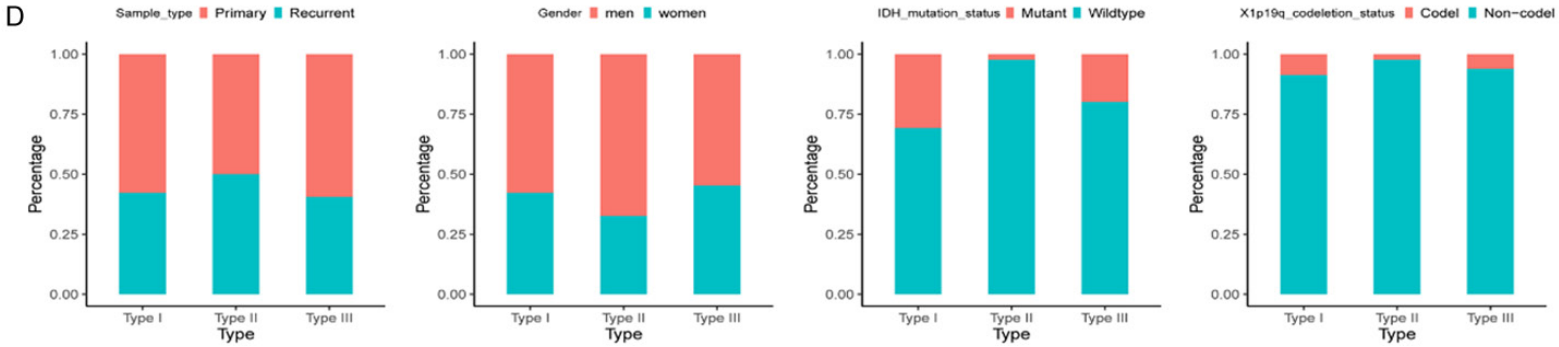


Figure 2. Sample clustering using nonnegative matrix factorization. A. Selection of highly variant genes. B. Consensus clustering matrix. C. Expression heatmap of highly variant genes in the three subtypes of glioblastoma multiforme. D. Proportions of tumor sample types, sex, mutation in isocitrate dehydrogenase (IDH), or 1p19q co-deletion in subtypes I-III.

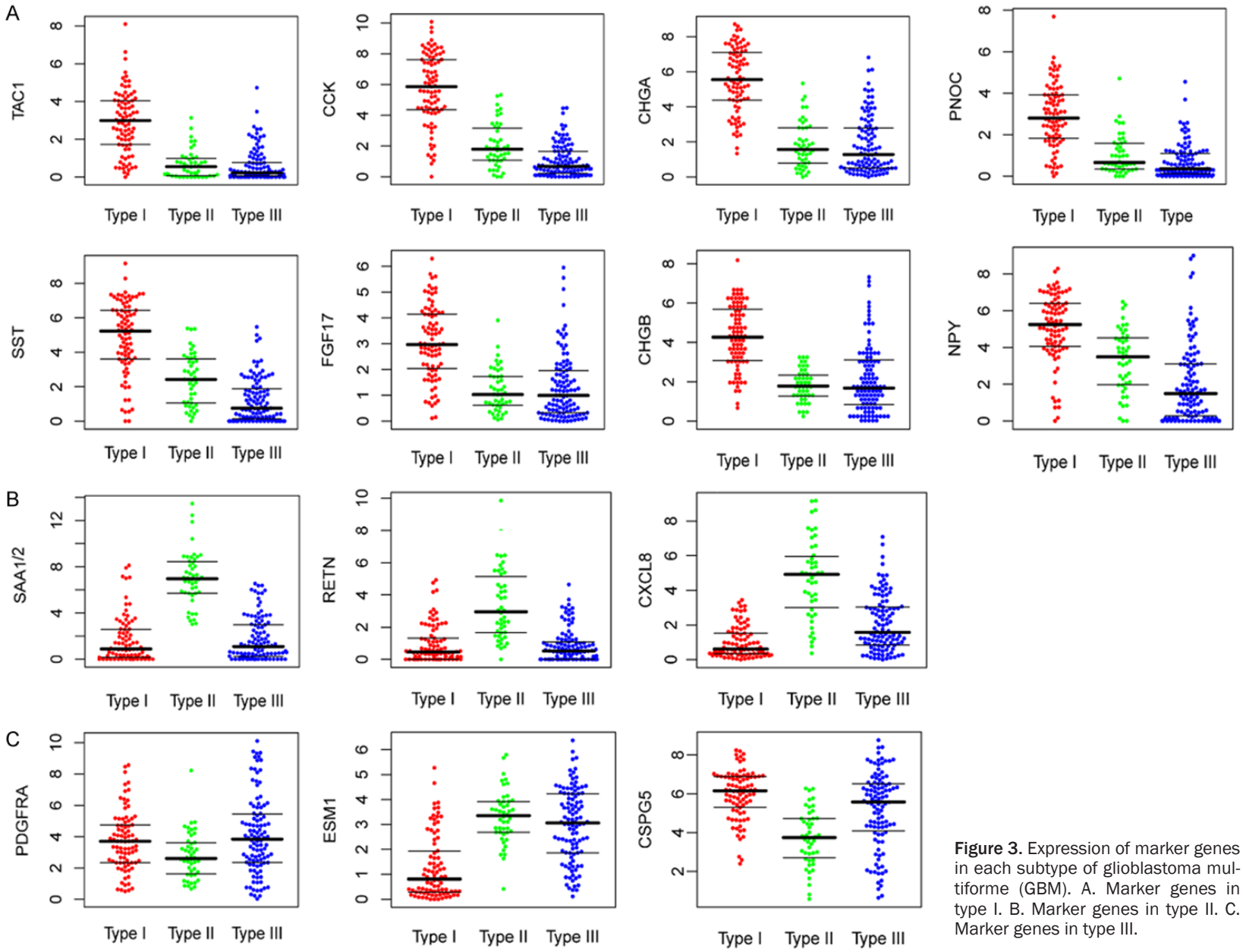


Figure 3. Expression of marker genes in each subtype of glioblastoma multiforme (GBM). A. Marker genes in type I. B. Marker genes in type II. C. Marker genes in type III.

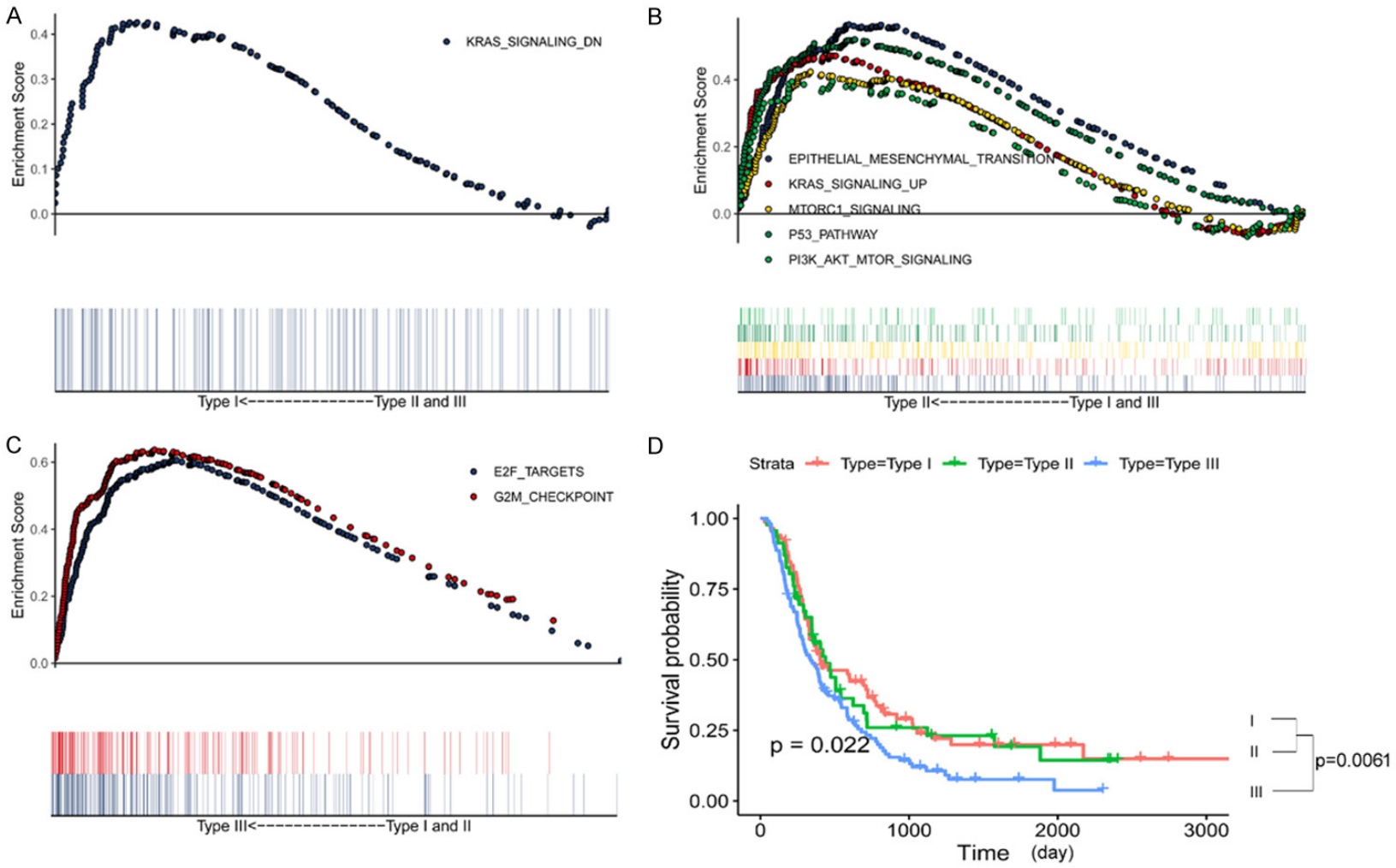


Figure 4. These three glioblastoma multiforme (GBM) subtypes vary in biology and prognosis. A. The KRAS signaling down (DN) hallmark gene set was significantly enriched in type I GBM. B. Multiple hallmark gene sets were significantly enriched in type II GBM. C. The hallmark gene sets E2F targets and G2M checkpoint were significantly enriched in type III GBM. D. Prognosis was worse for Type III than for Type I and II GBM.

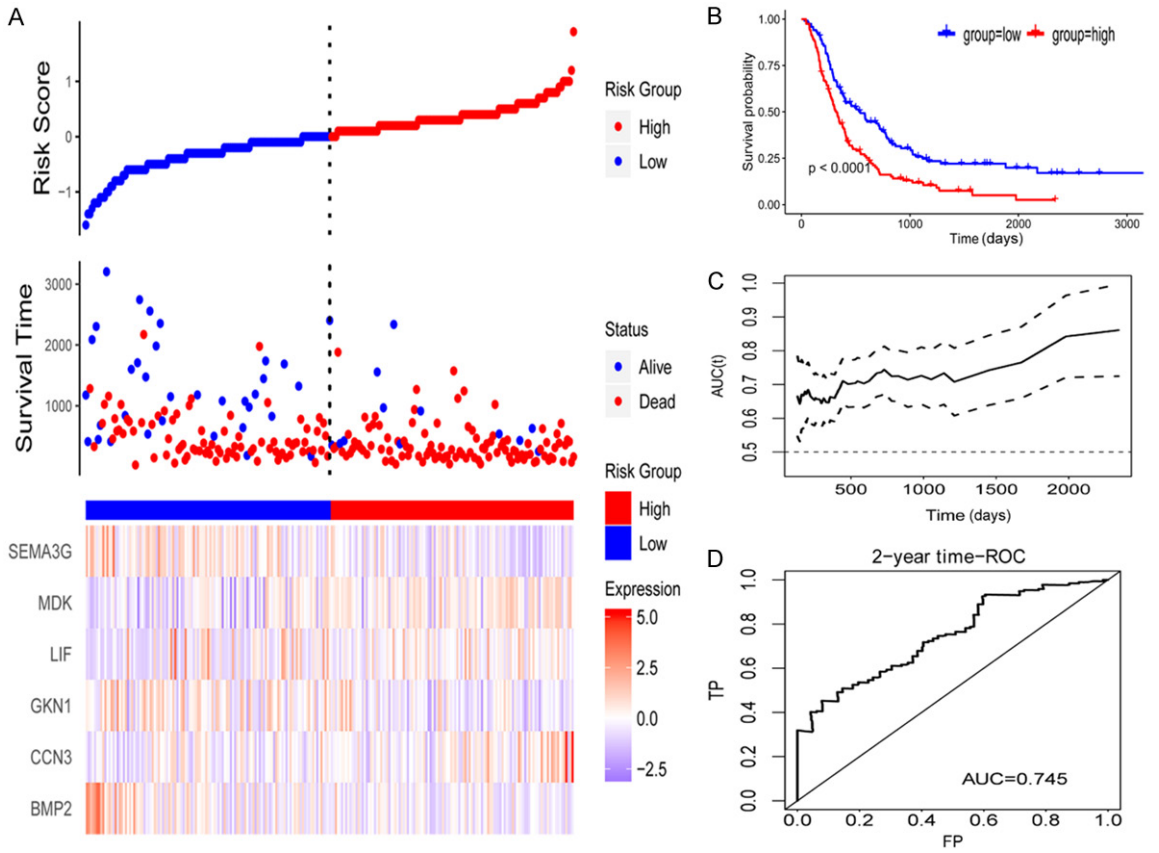


Figure 5. Establishment of a GFCK expression-based prognostic risk index. A. Distribution of prognostic risk index values, survival time, survival status and expression heatmap of the six selected GFCKs in glioblastoma multiforme (GBM) tissues in the CCGA mRNAseq_693 data set. B. Kaplan-Meier curves of patients stratified as low- or high-risk based on the median value of the risk index. C. Area under the receiver operating characteristic curve (AUC) to assess the ability of the risk index to predict survival. D. Time-dependent receiver operating characteristic curve for predicting 2-year survival. TP, true positive; FP, false positive.

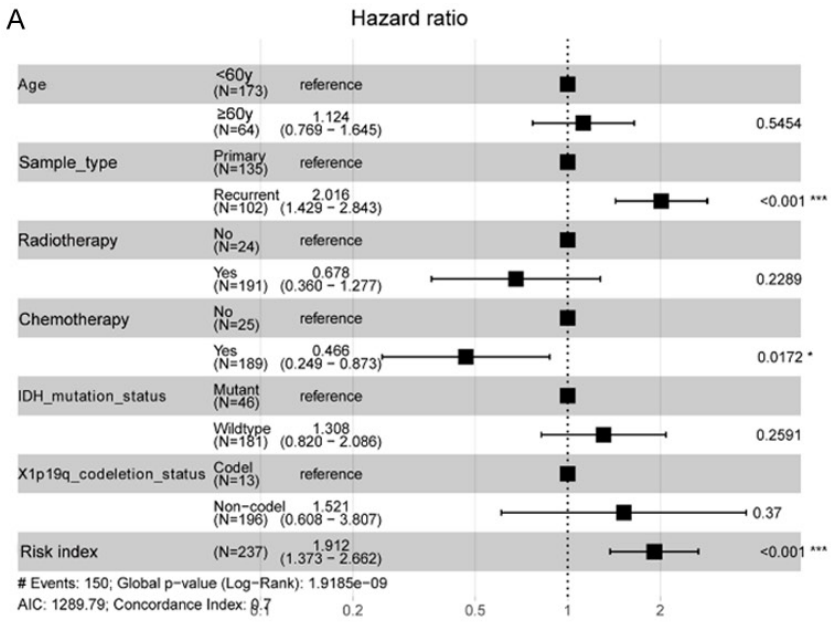
Univariate Cox hazard analysis showed that the RI was significantly associated with survival (hazard ratio 2.277, P = 1.38e-09), and the median RI was used to stratify patients into a low-risk group (< median RI) or high-risk group (≥ median RI) (**Figure 5A**). Overall survival was significantly shorter in the high-risk group (log-rank P < 0.001, **Figure 5B**). Time-dependent ROC curves suggested that the RI predicted survival more accurately over longer than shorter periods (**Figure 5C**). The area under the ROC curve for predicting 2-year survival was 0.745 (**Figure 5D**). The RI predicted prognosis independently of the following known prognostic factors: age, chemotherapy, radiotherapy, IDH mutation, and 1q19p co-deletion (**Figure 6A**).

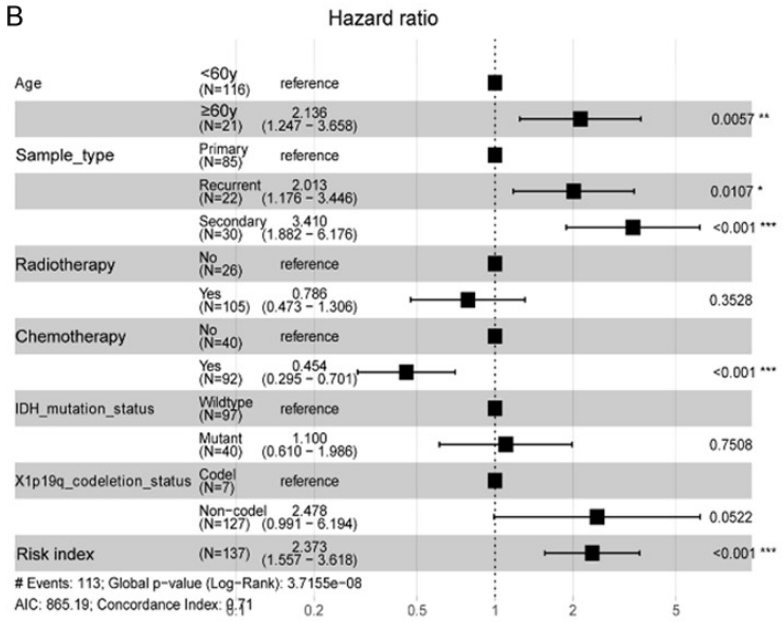
Validation of the GFCK-based prognostic RI

The prognostic RI was validated using the mRNAseq_325 and TCGA_array data sets. In both data sets, univariate Cox hazard analysis showed that the RI was significantly associated with survival (mRNAseq_325, hazard ratio 1.790, P < 0.001; TCGA_array, hazard ratio 1.287, P = 0.004), and it predicted prognosis independently of previously published clinicopathological factors (**Figure 6B**, **6C**). In both data sets, overall survival was shorter for the high-risk than for the low-risk group, although this difference achieved statistical significance only in the mRNAseq_325 data set (**Figure 6D**, **6E**).

Discussion

Personalized medicine plays an increasingly important role in cancer treatment. It is critical to identify "at-risk" patients and help them make decisions according to individual risk levels or subtypes. For GBMs, IDH mutation and O6-methylguanine-methyltransferase (MGMT)


A
Hazard ratio
Age
<60y (N=173) reference
≥60y (N=64) 1.124 (0.769 − 1.645) 0.5454
Sample_type
Primary (N=135) reference
Recurrent (N=102) 2.016 (1.429 − 2.843) <0.001 ***
Radiotherapy
No (N=24) reference
Yes (N=191) 0.678 (0.360 − 1.277) 0.2289
Chemotherapy
No (N=25) reference
Yes (N=189) 0.466 (0.249 − 0.873) 0.0172 *
IDH_mutation_status
Mutant (N=46) reference
Wildtype (N=181) 1.308 (0.820 − 2.086) 0.2591
X1p19q_codeletion_status
Codel (N=13) reference
Non−codel (N=196) 1.521 (0.608 − 3.807) 0.37
Risk index (N=237) 1.912 (1.373 − 2.662) <0.001 ***
Events: 150; Global p−value (Log−Rank): 1.9185e−09
AIC: 1289.79; Concordance Index: 0.7
0.1
0.2
0.5
1
2


B
Hazard ratio
Age
<60y (N=116) reference
≥60y (N=21) 2.136 (1.247 − 3.658) 0.0057 **
Sample_type
Primary (N=85) reference
Recurrent (N=22) 2.013 (1.176 − 3.446) 0.0107 *
Secondary (N=30) 3.410 (1.882 − 6.176) <0.001 ***
Radiotherapy
No (N=26) reference
Yes (N=105) 0.786 (0.473 − 1.306) 0.3528
Chemotherapy
No (N=40) reference
Yes (N=92) 0.454 (0.295 − 0.701) <0.001 ***
IDH_mutation_status
Wildtype (N=97) reference
Mutant (N=40) 1.100 (0.610 − 1.986) 0.7508
X1p19q_codeletion_status
Codel (N=7) reference
Non−codel (N=127) 2.478 (0.991 − 6.194) 0.0522
Risk index (N=137) 2.373 (1.557 − 3.618) <0.001 ***
Events: 113; Global p−value (Log−Rank): 3.7155e−08
AIC: 865.19; Concordance Index: 0.71
0.1
0.2
0.5
1
2
5

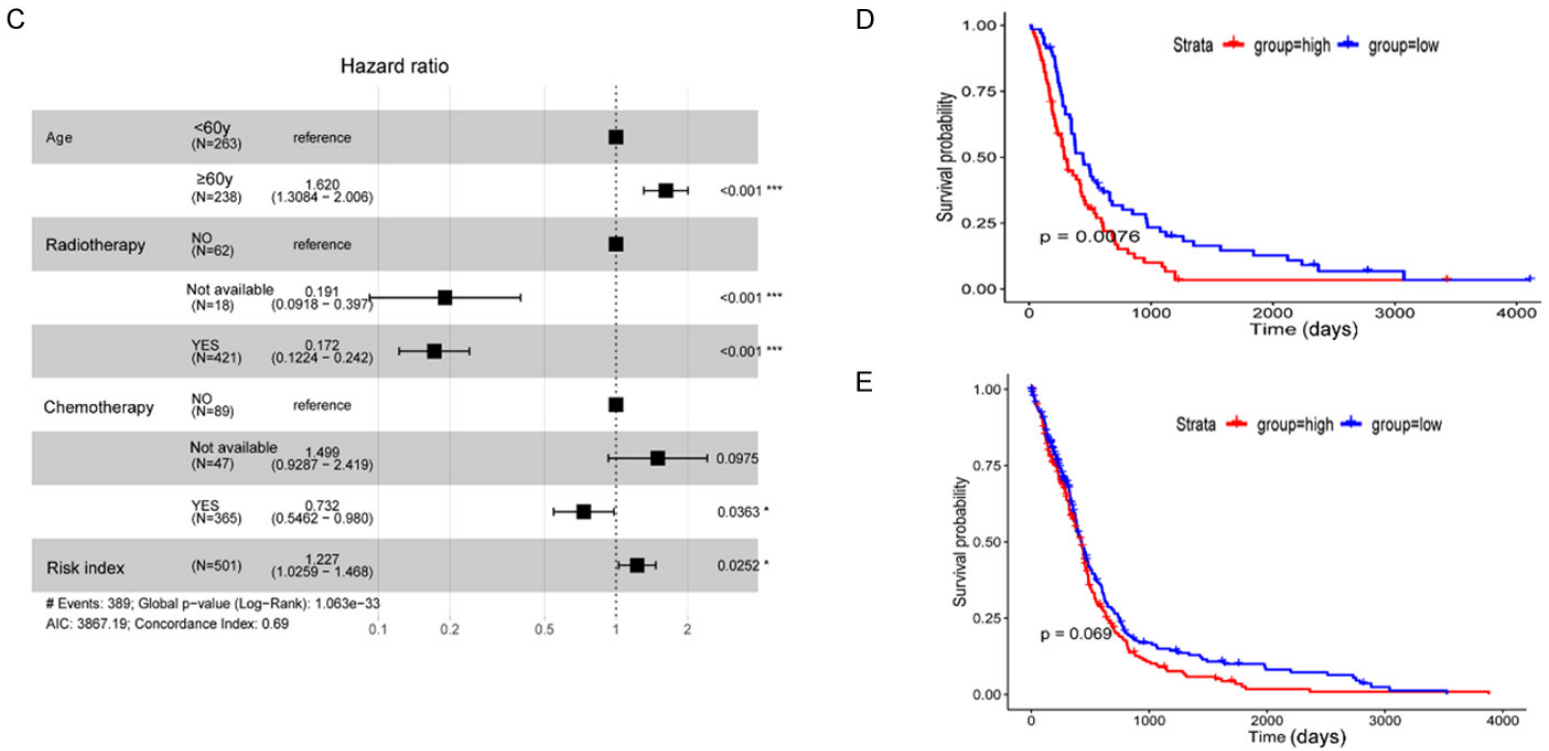


Figure 6. Validation of the GFCK expression-based prognostic risk index (RI). (A-C) Multivariate analyses of clinicopathological features and the RI in glioblastoma multiforme (GBM) tissues (A) in the CCGA mRNAseq_693 data set, (B) CCGA mRNAseq_325 data set, or (C) TCGA_array data set. (D, E) Kaplan-Meier curves of patients stratified as low- or high-risk based on median RI value in the (D) CCGA mRNAseq_325 data set or (E) TCGA_array data set. AIC, Akaike Information Criterion.

promoter methylation have proven to be reliable prognostic biomarkers [9]. However, IDH mutations are found in fewer than 5% of patients with primary GBM [35]. A methylated MGMT promoter is associated with longer survival of GBM patients, in part because of increased sensitivity to temozolomide [36]. The TCGA classification system has defined four subtypes of GBM (classical, mesenchymal, proneural and neural) based on comprehensive analysis of genetic alterations and expression [12]. These subtypes differ in prognosis [37] and response to chemoradiotherapy [12]. However, the classification system has not found widespread use in the clinic because it requires complex data that are costly to collect.

As a potentially more straightforward, accessible alternative, we have developed a system for classifying GBM subtypes based on expression of 68 GFCKs. Three subtypes were identified, each with a unique hallmark gene set and different clinical characteristics and prognosis. Type I GBM is associated with high expression of TAC1, CCK, CHGA, PNOC, SST, FGF17, CHGB, and NPY, and it is enriched for the hallmark gene set down-regulated by KRAS signaling. Type II GBM is associated with high expression of SAA1/2, RETN, and CXCL8, and it is enriched in several typical cancer-related hallmark gene sets: the epithelial-mesenchymal transition, KRAS signaling up, mTORC1 signaling, P53 pathway, and PI3K/AKT/mTOR signaling. Type III GBM, while clearly different from the other subtypes in clinical profile and prognosis, appears more difficult to define genetically: no genes were identified as uniquely up-regulated in tumors of this subtype. Nevertheless, we did find the subtype to be associated with enrichment in two cell cycle-related hallmark gene sets: E2F targets and G2M checkpoint. Further work is needed to understand the genetic basis for this subtype. It may also be that this subtype is defined more by environment-gene interactions than by genomic alterations.

Our GFCK-based classification may help clarify some of the difficulties in treating GBM. For example, treatments targeting the PI3K/AKT/mTOR signaling pathway have proven ineffective in the longer term [38], and our results suggest that this may be, in part, because only one of the three GBM subtypes is likely to respond to it. Our classification suggests the usefulness of targeting mTORC1, consistent with an analysis confirming mTOR as an important therapeutic target [39]. Our results suggest that GBM of Type III involve perturbations of the cell cycle, consistent with the observation that approximately 80% of GBM cases feature dysregulation of the CDK4/6-Rb-E2F axis in the cell cycle [35, 40, 41]. In these ways, our classification scheme may help guide improvements in GBM management.

In addition, we created a prognostic RI based on the expression of six GFCKs (BMP2, CCN3, GKN1, LIF, MDK, and SEMA3G). We validated the RI using two independent data sets from different platforms, suggesting that it may be a reliable prognostic marker. Indeed, it appears to predict prognosis independently of several known prognostic factors, and it can stratify patients into those showing shorter or longer survival, similar to previous multiple studies [42, 43]. However, while the prognostic difference was not significant in the data set from TCGA. This may result from too much censored data in TCGA or racial effects [44].

Unsurprisingly, several studies pointed out that the six GFCKs were associated with GBM. BMP2 may sensitize GBM stem-like cells to temozolomide by affecting HIF-1α stability and MGMT expression [45]. Ablating LIF may potently attenuate GBM growth enhancement [46]. and MDK expression correlates with GBM progression [47]. We also found the expression of CCN3, GKN1 and SEMA3G were associated with prognosis of patients with GBM. These roles of three GFCKs in GBM were needed further investigation.

Although our study provides new insights into GBM, it has several limitations. First, no suitable tool is currently available to predict new data using an NMF model, so our GBM classification system requires validation and perhaps improvement based on another independent data set based on the same platforms. Second, we were limited to publicly available data, so we were unable to examine whether our GBM subtypes also differ in metastatic behavior, response to chemotherapy, or tumor recurrence. Thirdly, it is not clear whether these six GFCKs are causal or merely markers for the prognosis of patients with GBM before furthermore molecular experiment.

Despite these limitations, we were able to define GBM subtypes based on GFCK expression patterns, from which we also derived a prognostic RI. If confirmed in prospective studies, our findings may aid in clinical decision-making to improve GBM management, and they may help guide further research into GBM onset and progression.

Acknowledgements

This study was supported by the Project of Nanning Scientific Research and Technology Development Plan (20193093 and 20183040-2), the Scientific Research Project of Guangxi Health Commission (Z20190376 and Z2019-0684) and the High-Level Medical Expert Training Program of Guangxi "139" Plan Funding (G201903049). This study was also sponsored by the Nanning Excellent Young Scientist Program and Guangxi Beibu Gulf Economic Zone Major Talent Program (RC20190103).

Disclosure of conflict of interest

None.

Address correspondence to: Donghua Zou, Department of Neurology, The Fifth Affiliated Hospital of Guangxi Medical University, No. 89 Qixing Road, Nanning 530022, Guangxi, People's Republic of China. Tel: +86-771-2636189; Fax: +86-771-2617-892; E-mail: danvor0922@hotmail.com; Xiangyu Wang, Department of Neurosurgery, The First Affiliated Hospital, Jinan University, 601 West Huangpu Avenue, Guangzhou 510630, Guangdong, People's Republic of China. Tel: +86-2038688888; Fax: +86-2085226089; E-mail: wang_xy123@126.com

References

[1] Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS and Villano JL. Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiol Biomarkers Prev 2014; 23: 1985-1996.

[2] Ron E, Modan B, Boice JD Jr, Alfandary E, Stovall M, Chetrit A and Katz L. Tumors of the brain and nervous system after radiotherapy in childhood. N Engl J Med 1988; 319: 1033-1039.

[3] Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, Wolinsky Y, Kruchko C and Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. Neuro Oncol 2015; 17 Suppl 4: iv1-iv62.

[4] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E and Mirimanoff RO; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005; 352: 987-996.

[5] Clarke J, Butowski N and Chang S. Recent advances in therapy for glioblastoma. Arch Neurol 2010; 67: 279-283.

[6] Wick W, Brandes AA, Gorlia T, Bendszus M, Sahm F, Taal W, Taphoom M, Domont J, Idbaih A, Campone M, Clement P, Stupp R, Fabbro M, Rhun E, Dubois F, Klein M, Platten M, weller M, Golfinopoulos V and Bent M. EORTC 26101 phase III trial exploring the combination of bevacizumab and lomustine in patients with first progression of a glioblastoma. J Clin Oncol 2016; 34: 2001-2001.

[7] Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, Carpentier AF, Hoang-Xuan K, Kavan P, Cernea D, Brandes AA, Hilton M, Abrey L and Cloughesy T. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. N Engl J Med 2014; 370: 709-722.

[8] Nabors LB, Fink KL, Mikkelsen T, Grujicic D, Tarnawski R, Nam DH, Mazurkiewicz M, Salacz M, Ashby L, Zagonel V, Depenni R, Perry JR, Hicking C, Picard M, Hegi ME, Lhermitte B and Reardon DA. Two cilengitide regimens in combination with standard treatment for patients with newly diagnosed glioblastoma and unmethylated MGMT gene promoter: results of the open-label, controlled, randomized phase II CORE study. Neuro Oncol 2015; 17: 708-717.

[9] Alexander BM and Cloughesy TF. Adult Glioblastoma. J Clin Oncol 2017; 35: 2402-2409.

[10] Gan Z, Zou Q, Lin Y, Huang X, Huang Z, Chen Z, Xu Z and Lv Y. Construction and validation of a seven-microRNA signature as a prognostic tool for lung squamous cell carcinoma. Cancer Manag Res 2019; 11: 5701-5709.

[11] 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, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner DD, Van Meir EG, Prados M, Sloan A, Black KL, Eschbacher J, Finocchiaro G, Friedman W, Andrews DW, Guha A, Iacocca M, O'Neill BP, Foltz G, Myers J, Weisenberger DJ, Penny R, Kucher-

Iapati R, Perou CM, Hayes DN, Gibbs R, Marra M, Mills GB, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird PW, Haussler D, Getz G, Chin L and Network TR. The somatic genomic landscape of glioblastoma. Cell 2013; 155: 462-477.

[12] Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM and Hayes DN; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010; 17: 98-110.

[13] Vitucci M, Hayes DN and Miller CR. Gene expression profiling of gliomas: merging genomic and histopathological classification for personalised therapy. Br J Cancer 2011; 104: 545-553.

[14] Kuhnt D, Becker A, Ganslandt O, Bauer M, Buchfelder M and Nimsky C. Correlation of the extent of tumor volume resection and patient survival in surgery of glioblastoma multiforme with high-field intraoperative MRI guidance. Neuro Oncol 2011; 13: 1339-1348.

[15] Roder C, Bisdas S, Ebner FH, Honegger J, Naegele T, Ernemann U and Tatagiba M. Maximizing the extent of resection and survival benefit of patients in glioblastoma surgery: high-field iMRI versus conventional and 5-ALA-assisted surgery. Eur J Surg Oncol 2014; 40: 297-304.

[16] Wilson TA, Karajannis MA and Harter DH. Glioblastoma multiforme: state of the art and future therapeutics. Surg Neurol Int 2014; 5: 64.

[17] Odenthal J, Takes R and Friedl P. Plasticity of tumor cell invasion: governance by growth factors and cytokines. Carcinogenesis 2016; 37: 1117-1128.

[18] Kalykaki A, Agelaki S, Kallergi G, Xyrafas A, Mavroudis D and Georgoulias V. Elimination of EGFR-expressing circulating tumor cells in patients with metastatic breast cancer treated with gefitinib. Cancer Chemother Pharmacol 2014; 73: 685-693.

[19] Musella V, Pietrantonio F, Di Buduo E, Iacovelli R, Martinetti A, Sottotetti E, Bossi I, Maggi C, Di Bartolomeo M, de Braud F, Daidone MG and Cappelletti V. Circulating tumor cells as a longitudinal biomarker in patients with advanced chemorefractory, RAS-BRAF wild-type colorectal cancer receiving cetuximab or panitumumab. Int J Cancer 2015; 137: 1467-1474.

[20] Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, Schwartz M, Porta C, Zeuzem S, Bolondi L, Greten TF, Galle PR, Seitz JF, Borbath I, Häussinger D, Giannaris T, Shan M, Moscovici M, Voliotis D and Bruix J; SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008; 359: 378-390.

[21] Kreisl TN, Kim L, Moore K, Duic P, Royce C, Stroud I, Garren N, Mackey M, Butman JA, Camphausen K, Park J, Albert PS and Fine HA. Phase II trial of single-agent bevacizumab followed by bevacizumab plus irinotecan at tumor progression in recurrent glioblastoma. J Clin Oncol 2009; 27: 740-745.

[22] Devarajan K. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. PLoS Comput Biol 2008; 4: e1000029.

[23] Hutchins LN, Murphy SM, Singh P and Graber JH. Position-dependent motif characterization using non-negative matrix factorization. Bioinformatics 2008; 24: 2684-2690.

[24] Pehkonen P, Wong G and Toronen P. Theme discovery from gene lists for identification and viewing of multiple functional groups. BMC Bioinformatics 2005; 6: 162.

[25] Xu M, Li W, James GM, Mehan MR and Zhou XJ. Automated multidimensional phenotypic profiling using large public microarray repositories. Proc Natl Acad Sci U S A 2009; 106: 12323-12328.

[26] Brunet JP, Tamayo P, Golub TR and Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004; 101: 4164-4169.

[27] Gaujoux R and Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010; 11: 367.

[28] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.

[29] Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D and Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003; 34: 267-273.

[30] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.

[31] Heagerty PJ, Lumley T and Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000; 56: 337-344.

[32] Weller M, Felsberg J, Hartmann C, Berger H, Steinbach JP, Schramm J, Westphal M, Schackert G, Simon M, Tonn JC, Heese O, Krex D, Nikkhah G, Pietsch T, Wiestler O, Reifenberger G, von Deimling A and Loeffler M. Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network. J Clin Oncol 2009; 27: 5743-5750.

[33] Walker MD, Alexander E Jr, Hunt WE, MacCarty CS, Mahaley MS Jr, Mealey J Jr, Norrell HA, Owens G, Ransohoff J, Wilson CB, Gehan EA and Strike TA. Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. A cooperative clinical trial. J Neurosurg 1978; 49: 333-343.

[34] Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, Pekmezci M, Rice T, Kosel ML, Smirnov IV, Sarkar G, Caron AA, Kollmeyer TM, Praska CE, Chada AR, Halder C, Hansen HM, McCoy LS, Bracci PM, Marshall R, Zheng S, Reis GF, Pico AR, O'Neill BP, Buckner JC, Giannini C, Huse JT, Perry A, Tihan T, Berger MS, Chang SM, Prados MD, Wiemels J, Wiencke JK, Wrensch MR and Jenkins RB. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. N Engl J Med 2015; 372: 2499-2508.

[35] 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, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE and Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. Science 2008; 321: 1807-1812.

[36] Martinez R, Schackert G, Yaya-Tur R, Rojas-Marcos I, Herman JG and Esteller M. Frequent hypermethylation of the DNA repair gene MGMT in long-term survivors of glioblastoma multiforme. J Neurooncol 2007; 83: 91-93.

[37] Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG and Aldape K. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 2006; 9: 157-173.

[38] Sami A and Karsy M. Targeting the PI3K/AKT/mTOR signaling pathway in glioblastoma: novel therapeutic agents and advances in understanding. Tumour Biol 2013; 34: 1991-2002.

[39] Fan Q, Aksoy O, Wong RA, Ilkhanizadeh S, Novotny CJ, Gustafson WC, Truong AY, Cayanan G, Simonds EF, Haas-Kogan D, Phillips JJ, Nicolaides T, Okaniwa M, Shokat KM and Weiss WA. A kinase inhibitor targeted to mTORC1 drives regression in glioblastoma. Cancer Cell 2017; 31: 424-435.

[40] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455: 1061-1068.

[41] Olmez I, Zhang Y, Manigat L, Benamar M, Brenneman B, Nakano I, Godlewski J, Bronisz A, Lee J, Abbas T, Abounader R and Purow B. Combined c-Met/Trk inhibition overcomes resistance to CDK4/6 inhibitors in glioblastoma. Cancer Res 2018; 78: 4360-4369.

[42] Bai X, Lu D, Lin Y, Lv Y and He L. A seven-miRNA expression-based prognostic signature and its corresponding potential competing endogenous RNA network in early pancreatic cancer. Exp Ther Med 2019; 18: 1601-1608.

[43] Lin Y, Liang R, Ye J, Li Q, Liu Z, Gao X, Piao X, Mai R, Zou D and Ge L. A twenty gene-based gene set variation score reflects the pathological progression from cirrhosis to hepatocellular carcinoma. Aging (Albany NY) 2019; 11: 11157-11169.

[44] Patel NP, Lyon KA and Huang JH. The effect of race on the prognosis of the glioblastoma patient: a brief review. Neurol Res 2019; 41: 967-971.

[45] Persano L, Pistollato F, Rampazzo E, Della Puppa A, Abbadi S, Frasson C, Volpin F, Indraccolo S, Scienza R and Basso G. BMP2 sensitizes glioblastoma stem-like cells to Temozolomide by affecting HIF-1alpha stability and MGMT expression. Cell Death Dis 2012; 3: e412.

[46] Inda MM, Bonavia R, Mukasa A, Narita Y, Sah DW, Vandenberg S, Brennan C, Johns TG, Bachoo R, Hadwiger P, Tan P, Depinho RA, Cavenee W and Furnari F. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. Genes Dev 2010; 24: 1731-1745.

[47] Roversi G, Pfundt R, Moroni RF, Magnani I, van Reijmersdal S, Pollo B, Straatman H, Larizza L and Schoenmakers EF. Identification of novel genomic markers related to progression to glioblastoma through genomic profiling of 25 primary glioma cell lines. Oncogene 2006; 25: 1571-1583.

Table S1. Gene sets enriched in type II glioblastoma multiforme

Name	Size	Enrichment score	Adjusted P*
HALLMARK_ALLOGRAFT_REJECTION	186	0.699	0.000
HALLMARK_COMPLEMENT	192	0.591	0.000
HALLMARK_IL2_STAT5_SIGNALING	197	0.582	0.001
HALLMARK_COAGULATION	119	0.573	0.001
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	48	0.682	0.001
HALLMARK_INTERFERON_GAMMA_RESPONSE	193	0.690	0.001
HALLMARK_INFLAMMATORY_RESPONSE	193	0.646	0.002
HALLMARK_IL6_JAK_STAT3_SIGNALING	80	0.697	0.002
HALLMARK_APOPTOSIS	159	0.548	0.002
HALLMARK_TNFA_SIGNALING_VIA_NFKB	196	0.667	0.003
HALLMARK_P53_PATHWAY	195	0.520	0.006
HALLMARK_XENOBIOTIC_METABOLISM	184	0.470	0.009
HALLMARK_HYPOXIA	191	0.548	0.009
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	198	0.565	0.009
HALLMARK_INTERFERON_ALPHA_RESPONSE	95	0.680	0.009
HALLMARK_KRAS_SIGNALING_UP	191	0.473	0.012
HALLMARK_GLYCOLYSIS	189	0.482	0.022
HALLMARK_UV_RESPONSE_UP	156	0.381	0.119
HALLMARK_ESTROGEN_RESPONSE_LATE	194	0.338	0.121
HALLMARK_ADIPOGENESIS	194	0.399	0.136
HALLMARK_FATTY_ACID_METABOLISM	150	0.400	0.137
HALLMARK_PI3K_AKT_MTOR_SIGNALING	102	0.395	0.146
HALLMARK_MTORC1_SIGNALING	197	0.424	0.150
HALLMARK_ANGIOGENESIS	34	0.449	0.159
HALLMARK_MYOGENESIS	192	0.322	0.161
HALLMARK_APICAL_JUNCTION	195	0.323	0.174
HALLMARK_CHOLESTEROL_HOMEOSTASIS	71	0.370	0.174
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	108	0.393	0.196
HALLMARK_PEROXISOME	101	0.337	0.221

*Adjusted by the false discovery rate.

Table S2. Genes associated with survival of glioblastoma multiforme patients in univariate Cox hazard analysis

Gene	β	HR (95% CI)	*P* value
APLN	0.123	1.13 (1.02-1.25)	0.018
BMP2	-0.171	0.843 (0.747-0.951)	0.005
CCK	-0.0564	0.945 (0.894-1)	0.048
CCL19	-0.151	0.86 (0.748-0.989)	0.035
CCN3	0.169	1.18 (1.05-1.34)	0.007
CHGA	-0.066	0.936 (0.881-0.995)	0.033
CHGB	-0.0832	0.92 (0.85-0.996)	0.039
CRH	-0.343	0.709 (0.538-0.935)	0.015
CTF1	0.182	1.2 (1.02-1.41)	0.024
EDN3	-0.166	0.847 (0.722-0.995)	0.044
EGF	0.275	1.32 (1.02-1.7)	0.037
GDF10	-0.247	0.782 (0.649-0.94)	0.009
GKN1	-0.0786	0.924 (0.86-0.994)	0.034
LIF	0.0894	1.09 (1-1.2)	0.049
MDK	0.119	1.13 (1.04-1.22)	0.003
NRG3	-0.144	0.866 (0.755-0.992)	0.039
PDGFA	0.143	1.15 (1.02-1.3)	0.019
PLAU	0.0941	1.1 (1.02-1.19)	0.018
PRLHR	-0.532	0.588 (0.433-0.797)	0.001
SEMA3G	-0.292	0.747 (0.604-0.922)	0.007
SEMA4C	0.256	1.29 (1.04-1.6)	0.020
SPP1	0.0642	1.07 (1-1.13)	0.037
STC1	0.147	1.16 (1.02-1.31)	0.019
TAC1	-0.0944	0.91 (0.829-0.999)	0.047
TGFB2	0.128	1.14 (1.03-1.25)	0.008
TNC	0.154	1.17 (1.08-1.26)	0.000
TNFRSF11B	0.229	1.26 (1.09-1.46)	0.002
TNFSF14	0.246	1.28 (1.1-1.49)	0.001
VEGFA	0.0925	1.1 (1.03-1.17)	0.003

HR, hazard ratio; CI, Confidence Interval.