

# Identification of Potential Key Genes Associated with the Pathogenesis and Prognosis of Glioblastoma Based on Integrated Bioinformatics Analysis

**Youbing Tu**

Shenzhen People's Hospital

**Sijia Zhou**

Southern Medical University Nanfang Hospital

**Hui Zhang**

Shanghai Tenth People's Hospital

**Jing Lv**

Taihe Hospital

**Dengfeng Ding** (✉ [df\\_ding@sina.com](mailto:df_ding@sina.com))

Shenzhen People's Hospital <https://orcid.org/0000-0003-0615-8741>

**Zhongjun Zhang**

Shenzhen People's Hospital

---

## Primary research

**Keywords:** Glioblastoma, bioinformatics, differentially expressed genes, survival, biomarker, GEO, TCGA

**Posted Date:** April 20th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-20089/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Despite striking advances in multimodality management, the low survival rate of Glioblastoma (GBM) patients has not been significantly improved and identifying novel diagnostic and prognostic biomarkers is urgently demanded. The present study aimed to identify potential key genes associated with the pathogenesis and prognosis of GBM.

**Methods:** Differentially expressed genes between GBM and normal brain tissue samples were screened by an integrated analysis of multiple gene expression profile datasets. Key genes related to the pathogenesis and prognosis of GBM were identified by employing protein–protein interaction network and Cox proportional hazards model analyses.

**Results:** We identified nine hub genes (TP53, FN1, EGFR, MYC, RRM2, EZH2, FOXM1, CD44 and MMP2) which might be closely associated with the pathogenesis of GBM. A prognostic gene signature consisted of RAB33B, KIAA1199, TEK, EVC, SOD2, CXCR4, hCG\_40738, CHD9, GCSH, SUHW1, RPS6KA5, PDCD4, ZG16, KCNG1, DECR1, PPCS, SERPINF1, TMSB10, NAT10, HIC2, PIR and OR2W1 was constructed with a good performance in predicting overall survivals (OS).

**Conclusions:** The findings of present study would provide certain reference for further predicting the diagnostic and prognostic biomarkers to facilitate the molecular targeting therapy of GBM.

## Background

Glioblastoma (GBM), also known as WHO Grade IV astrocytoma, is the most common and highly malignant type of brain tumor that arises from glial cells or their precursors. An estimated 2-5 GBM cases are newly diagnosed per 100,000 persons each year, which accounts for 12-15% of all brain tumors and approximately 70-80% of all diagnosed gliomas.[1-4] Glioblastoma (GBM) is the most aggressive brain tumor with rapid relapse, and responds poorly to current therapies. Despite tremendous advances in knowledge and therapies over the decades, the survival rate of patients diagnosed with GBM has not significantly improved, with an average survival time of only 12 months following diagnosis, only 2% of GBM patients survive to 3 years [5, 6] and 5.1% of glioblastoma patients can survive to 5 years.[7] In spite of significant advances in understanding the epidemiology, pathology, and molecular mechanisms of GBM and in implementing arising therapeutic options such as targeted and immune-based therapies, not all patients react to existing molecularly targeted agents developed for specified acknowledged biomarkers and no improvement is observed in the overall survival rate of GBM.[8] Existing studies still lack of a comprehensive understanding of the underlying pathophysiological mechanisms leading to invasiveness nature of GBM hinders the development of therapies for GBM. Therefore, comprehensive understanding the molecular mechanism of GBMs is of great necessity for locating a completely unique and effective therapeutic strategy.[9-11]

Recently, microarray and high-throughput genomic sequencing technology has significantly facilitated the detection of critical genetic or epigenetic alternations in carcinogenesis and the identification of novel

promising diagnostic, prognostic biomarkers and therapeutic targets.[12-15] Over the past years, expanding data and accumulating results from computational analyses furthering the understanding of underlying pathophysiological mechanisms and prognosis of GBM.[16, 17] Moreover, in order to overcome the restricted or inconsistent results due to the application of either different technological platforms or a small sample size, integrated bioinformatics strategies have been adopted in cancer research and a large amount of valuable biological information has been uncovered.[18]

Herein, we initially performed an integrated analysis and identified differentially expressed genes (DEGs) by using microarray and RNA sequencing data in human GBM and normal GBM tissue samples from multiple studies. Additionally, functional enrichment analysis was further conducted to analyze the main biological functions modulated by the DEGs. Ultimately, key genes affecting the pathogenesis and prognosis of GBM patients were identified by utilizing protein–protein interaction (PPI) network and survival analyses.

## **Materials And Methods**

### **Gene Expression Profile Data Collection and Pre-processing**

Microarray data on gene expression (GSE16011, GSE111260, GSE50161, GSE61335, GSE66354, GSE15824, GSE94349, GSE86574 and GSE68015) were downloaded from Gene Expression Omnibus (GEO) [<http://www.ncbi.nlm.nih.gov/geo/>]. All enclosed datasets meet the subsequent criteria: (1) Human brain tissue samples were employed. (2) Case-control groups included. (3) No less than ten samples contained. Small sample size is one of the main challenges of microarray analysis, while a large sample size may reliably reveal the DEGs or non-coding RNAs, and integrated bioinformatics studies also prefer to use datasets with a relatively large sample size.[19, 20] Therefore, the GEO datasets which contained no less than ten samples were chosen for further study. Raw RNA sequencing data containing 528 GBM samples and 10 matched non-cancerous samples were obtained from The Cancer Genome Atlas (TCGA) [<https://cancergenome.nih.gov/>].

### **Integrated Analysis of Microarray Datasets**

Limma package in R software was applied to perform the normalization for the matrix data of each GEO dataset, and the DEGs were also screened between tumor and normal tissues.[21] The DEGs identified from the 9 datasets were reliably integrated by an R package “RobustRankAggreg” [22] based on a robust rank aggregation (RRA) method. This RRA method screens genes ranked consistently better than expected based on the comparison of actual data with a null model that assumes random order of input lists.[22] Meanwhile, bonferroni correction was performed for avoiding false positive results and leave-one-out cross-validation was applied to assess the stability of acquired p-values. Therefore, we did not integrate the gene expression values of samples varies from different datasets. And as with many published papers based on the RobustRankAggreg package [23, 24], we have not performed batch effect

correction as well.  $|\log_2FC| > 1$ ,  $P\text{-value} < 0.05$  and  $\text{adjust } P\text{-value} < 0.05$  were considered statistically significant for the DEGs.

### **DEGs Validation by TCGA**

The integrated analysis results of GEO datasets were further validated with the RNA sequencing data in the TCGA GBM dataset. The data were normalized and analyzed by the Limma package.[21] Genes were considered to be significantly differentially expressed at a significance level of log fold change ( $|\log_2FC| > 1$ ,  $P\text{-value} < 0.05$  and  $\text{adjust } P\text{-value} < 0.05$ ). Overlapping DEGs between the integrated microarray data and RNA sequencing data analyses were retained for further study.

### **Functional and Pathway Enrichment Analysis of DEGs**

In order to figure out the potential biological functions correlated with the overlapping DEGs, Gene Ontology(GO) Term Enrichment analysis was first conducted based the 512 DEGs utilizing the clusterProfiler package [25], which including enrichment for GO 'Biological Process', 'Molecular Function' and 'Cellular Component' terms. Moreover, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis based these genes were also implemented by clusterProfiler [25] to expound promising signaling pathways associated with the overlapping DEGs. At a significance level of  $P\text{-value} < 0.05$  and  $\text{adjust } P\text{-value} < 0.05$  were defined as the cut-off criteria.

### **PPI Network and Module Analysis for Co-expression Network of Predicted Target Genes.**

The STRING database [26] has been widely applied to explore potential protein-protein interactions (PPI) among the overlapping DEGs. PPIs with a confidence score  $\geq 0.4$  [27] were retained and further imported to Cytoscape v3.7.1 software [28] for visualizing and constructing the PPI network of overlapping DEGs. In addition, to detect hub genes with large degrees (connectivity degree  $> 5$ ) and high betweenness centrality [29] in the established PPI network, we performed module analysis using Molecular Complex Detection (MCODE) [30] app with default parameters in Cytoscape. Besides, we also conducted GO and KEGG pathway enrichment analyses for significant modules as well.

### **Survival Analysis**

Both preprocessed RNA-sequencing data and clinical data of GBM were downloaded from TCGA. Among the clinical data, 525 sample files that contained status and survival time of patients were chosen for subsequent survival analysis. Univariate Cox proportional hazards regression analysis was utilized to predict the candidate genes that were strongly associated with survival.[31, 32] Subsequently, the candidate genes with  $P\text{-value} < 0.05$  were selected for further identification of the prognostic gene markers with multivariate Cox proportional hazards regression analysis.[33] These GBM patients were divided into either low- or high- risk groups according to the median prognostic risk score. Moreover, the screened prognostic gene markers were then enrolled into prognostic risk score model employing a least absolute shrinkage and selection operator (LASSO) regression to further narrow the correlated key genes

for predicting prognosis of GBM.[34] In addition, we performed time-dependent receiver operating characteristic (ROC) curve analysis by using an R package “survivalROC” to evaluate the predictive accuracy of the prognostic signature for time-dependent cancer death.[35] The area under the curve (AUC) was calculated to measure the predictive ability of the gene signature for clinical outcomes.

## Statistical Analysis

The univariate and multivariate Cox proportional hazards regression analyses were performed using an R package “survival”. Meanwhile, hazard ratio (HR) and 95% confidence interval (CI) were calculated to identify protective (HR < 1) or risky genes (HR > 1). Besides, a survival curve made by Kaplan–Meier method was implemented to estimate the differences in survival time between the high- and low-risk patients. All the statistical analyses were conducted with R (version 3.4.3) (<https://www.r-project.org/>).

## Results

Identification of robust and reliable DEGs could provide immense help for understanding molecular pathologies and mechanisms underlying complex disease. To facilitate the identification of DEGs between GBMs and control samples, p-values were estimated to identify genes that were differentially expressed. Thus we integrated the DEGs of the nine datasets downloaded from GEO database and these DEGs were further validated by TCGA database. The information of the nine included GEO datasets in this study was shown in **Table 1** and the detailed information of each sample was displayed in **Supplementary Table 1**. A total of 670 DEGs comprising 293 down-regulated and 377 up-regulated genes were screened after the integrated analysis of nine GEO datasets (**Supplementary Table 2**). The DEGs obtained from the TCGA GBM dataset consist of 962 down-regulated and 1777 up-regulated genes (**Supplementary Table 3**). We further identified 512 overlapping DEGs which consisted of 212 down-regulated and 300 up-regulated genes by intersecting the results of integrated microarray and RNA sequencing data analyses (**Figure 1** and **Supplementary Table 4**).

**TABLE1** Information for the nine GEO datasets included in the current study.

Dataset	Reference	Platform	samples(Tumor/Control)
GSE16011	Gravendeel et al.,2009	[CDF: Hs133P_Hs_ENTREZG.cd]Affymetrix GeneChip Human Genome U133 Plus 2.0 Arr	284(276/8)
GSE111260	Jeanmougin et al.,2018	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]	70(67/3)
GSE50161	Donson et al.,2013	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	47(34/13)
GSE61335	Joy et al.,2014	[HG-U133A] Affymetrix Human Genome U133A Array [Custom Brainarray v18 ENTREZG	62(48/14)
GSE66354	Donson et al.,2015	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	32(19/13)
GSE15824	Morin et al., 2011	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	17(12/5)
GSE94349	Donson et al.,2017	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	37(10/27)
GSE86574	Donson et al.,2016	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	25(15/10)
GSE68015	Donson et al.,2015	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	27(11/16)

**FIGURE 1 |** Identification of DEGs. (A) The heat map of top 20 up- and down-regulated DEGs in the integrated microarray analysis. Each column and each row represents one GEO dataset and one gene respectively. The value in each rectangle was log<sub>2</sub>-transformed. The gradual color ranging from blue to red represents the changing process from down- to up-regulation. (B) Venn diagrams of the overlapping DEGs between the integrated nine GEO datasets and the TCGA GBM dataset. (C) The heat map of 512

overlapping DEGs in GBM and normal brain tissues (TCGA dataset). Each column represents one sample and each row represents one gene. The gradual color ranging from blue to red represents the changing process from down- to up-regulation.

### Functional Enrichment Analysis of the overlapping DEGs

To further expound the potential biological functions of 512 overlapping DEGs in GBM, the GO functional and KEGG pathway enrichment analyses were performed. In terms of the 212 down-regulated genes, they were significantly enriched in multiple biological processes related to modulation of chemical synaptic transmission including regulation of trans-synaptic signaling, synaptic vesicle cycle, regulation of vesicle-mediated transport, signal release and regulation of neuron projection development (**Figure 2A and Supplementary Table 5**). As for the 300 up-regulated genes, they showed a close correlation with extracellular matrix and cell cycle regulation, such as extracellular structure organization, extracellular matrix organization, skeletal system development and cell cycle G1/S phase transition, nuclear division, organelle fission, G1/S transition of mitotic cell cycle, regulation of cell cycle phase transition, regulation of mitotic cell cycle phase transition and so on. (**Figure 2A and Supplementary Table 5**). According to KEGG pathway enrichment analysis, the down-regulated genes mainly participated in synaptic chemical signal transduction associated signaling pathways, like cAMP signaling pathway, Neuroactive ligand-receptor interaction, Synaptic vesicle cycle, Endocytosis, Glutamatergic synapse, Adrenergic signaling in cardiomyocytes, GABAergic synapse and so on (**Figure 2B and Supplementary Table 6**). As for the upregulated genes, they mainly regulated pathways correlated with cell cycle, such as Pathways in cancer, PI3K-Akt signaling pathway, Cell cycle, p53 signaling pathway, Cellular senescence and so on (**Figure 2B and Supplementary Table 6**).

**FIGURE 2 |** Functional enrichment analysis of the overlapping DEGs. (A) GO enrichment analysis of the overlapping DEGs. The y-axis displays the significantly enriched GO biological process terms, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a GO term among all the overlapping DEGs. (B) KEGG pathway enrichment analysis of the overlapping DEGs. The y-axis displays significantly enriched KEGG pathways, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a KEGG pathway among all the overlapping DEGs.

### PPI Network and Module Analysis

The PPI network of 512 overlapping DEGs consisted of 197 nodes and 1727 interactions (**Figure 3A and Supplementary Table 7**). The candidate hub nodes were identified according to the degree [36] and betweenness[37], the higher the two topological features quantitative values of a gene, the more important it is in current network. As a result, nine candidate hub nodes, the degree and betweenness of which were all more than four-fold of the corresponding median values, were identified, namely, tumor protein p53(TP53), Fibronectin(FN1), epidermal growth factor receptor (EGFR), Myc(MYC), Ribonucleoside-diphosphate reductase subunit M2(RRM2), Enhancer of zeste homolog 2(EZH2), Forkhead box protein M1(FOXM1), CD44 antigen(CD44), matrix metalloproteinase-2(MMP2) (**Supplementary Table 8**). Additionally, to detect significant clustering modules in present PPI network we

conducted module analysis and obtained top three modules with high scores (**Figures 3B–D**). The eight candidate hub nodes except TP53 and MYC were contained in the three modules, which implied that the three modules might remarkably represent the key biological characteristics of this PPI network, and thereby the eight nodes were defined as major hub nodes in the PPI network (**Figure 4**). At the aspect of GO enrichment analysis, module 1 was closely correlated with nuclear division, organelle fission, chromosome segregation, mitotic, nuclear division, nuclear chromosome segregation, mitotic sister chromatid segregation, sister chromatid segregation, microtubule cytoskeleton organization and microtubule cytoskeleton organization involved in mitosis; module 2 was highly connected to extracellular matrix organization, extracellular structure organization, post-translational protein modification, positive regulation of cell adhesion and regulation of cell-cell adhesion; module 3 was intimately associated with platelet degranulation and leukocyte migration (**Figure 5A and Supplementary Table 9**). With respect to KEGG pathway enrichment analysis, the genes in module 1 were mainly enriched in Cell cycle, Human T-cell leukemia virus 1 infection, Progesterone-mediated oocyte maturation, Oocyte meiosis, Cellular senescence, DNA replication and p53 signaling pathway; the genes in module 2 mainly participated in Human papillomavirus infection, AGE-RAGE signaling pathway in diabetic complications, ECM-receptor interaction, Focal adhesion, PI3K-Akt signaling pathway, Relaxin signaling pathway, Protein digestion and absorption, Amoebiasis, Proteoglycans in cancer, MicroRNAs in cancer and so on; the genes in module 3 were mainly implicated in PI3K-Akt signaling pathway and Regulation of actin cytoskeleton (**Figure 5B, Supplementary Table 10**).

Our data presented that once some DEGs were overexpressed the signaling pathways that they involved in may be dysregulated. For instance, highly up-regulated DEPDC1, KIF14, CEP55 and MCM3 in GBM tissues might be responsible for the dysfunction of Cell cycle, Progesterone-mediated oocyte maturation, DNA replication, Human T-cell leukemia virus 1 infection, Oocyte meiosis, Cellular senescence, and p53 signaling pathway; TLR3, EGFR, CD44, MMP2 and CCL2 in Human papillomavirus infection, AGE-RAGE signaling pathway in diabetic complications, ECM-receptor interaction, Focal adhesion, PI3K-Akt signaling pathway and Relaxin signaling pathway were overexpressed as well. Additionally, up-regulated genes (CD63, ANGPT2 and CXCR4) in module 3 significantly enriched in Regulation of actin cytoskeleton, PI3K-Akt signaling pathway, Complement and coagulation cascades and HIF-1 signaling pathway. All these KEGG pathways were enriched by the up-regulated genes in the three modules, only ANXA1 and ANXA2 was significantly enriched in GO terms (regulation of interleukin-8 production and positive regulation of endocytosis) identified in module analysis.

**FIGURE 3** | Protein–protein interaction (PPI) network and hub clustering modules. (A) The PPI network of overlapping DEGs. (B) Module 1 (MCODE score = 38.8). (C) Module 2 (MCODE score = 9.385). (D) Module 3 (MCODE score = 4). The gradual node color ranging from blue to red represents the changing process from low to high degrees of each gene.

**FIGURE 4** | Expression of the nine hub DEGs in GBM and normal brain tissues (TCGA dataset). Expression values of these genes are log<sub>2</sub>-transformed.

**FIGURE 5** | Functional enrichment analysis of the overlapping DEGs in the 3 modules. (A) GO enrichment analysis of the overlapping DEGs in the 3 modules. The y-axis displays significantly enriched GO biological process terms, and the x-axis displays different modules. Count refers to the number of genes enriched in a GO term among all the overlapping DEGs. (B) KEGG pathway enrichment analysis of the overlapping DEGs in the 3 modules. The y-axis displays significantly enriched KEGG pathways, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a KEGG pathway among all the overlapping DEGs.

## Survival Analysis

A total of 2018 genes significantly correlated with survival time ( $P < 0.05$ ) were preliminarily identified by the univariate Cox proportional hazards regression model (Supplementary Table 11). A total of 261 patients with the risk scores larger than the median risk score (1.060) were divided into the high-risk group, whereas the other 264 patients were divided into the low-risk group. The risk score result of the TCGA GBM dataset was presented in **Figure 6A**. Moreover, a prognostic gene signature composed of 526 genes was further developed after using the multivariate Cox proportional hazards regression model. In addition, we performed LASSO analysis to eliminate possible interactions between genes which might arise from the factor that small sample size with large number of genes. As a result, 22 candidate prognostic genes were identified, namely, RAB33B, KIAA1199, TEK, EVC, SOD2, CXCR4, hCG\_40738, CHD9, GCSH, SUHW1, RPS6KA5, PDCD4, ZG16, KCNG1, DECR1, PPCS, SERPINF1, TMSB10, NAT10, HIC2, PIR, OR2W1. Among these 22 genes, RAB33B, KIAA1199, EVC, SOD2, hCG\_40738, CHD9, GCSH, SUHW1, RPS6KA5, KCNG1, DECR1, PPCS, NAT10, OR2W1, HIC2 with  $HR < 1$  were identified as protective prognostic genes, whereas CXCR4, TMSB10, TEK, PDCD4, ZG16, PIR and SERPINF1 with  $HR > 1$  were identified as risky prognostic genes. As shown in **Figure 6B**, a highly significant difference in OS was detected between the high- and low-risk groups ( $P < 0.0001$ ). The prognostic gene signature shown a good performance in survival prediction, as the AUC was 0.756, 0.722, and 0.797 for 1-, 3-, and 5-year OSs (**Figure 6C**), respectively. The expression level distribution of the 22 genes in low- and high-risk groups was shown in **Figure 7**.

**TABLE 2 Prognostic value of the 22 genes in the GBM patients of the TCGA cohort.**

Gene symbol	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P-value	HR (95% CI)	P-value	Coefficient
RAB33B	1.205(1.036-1.4)	0.015	0-0	<0.0001	0.00E+00
KIAA1199	1.097(1.012-1.188)	0.024	0-0	<0.0001	0.00E+00
TEK	0.753(0.57-0.995)	0.046	1.87E+44(1.38E+44-2.52E+44)	<0.0001	1.87E+44
EVC	1.42(1.061-1.899)	0.018	0-0	<0.0001	0.00E+00
SOD2	1.124(1.042-1.213)	0.003	0-0	<0.0001	0.00E+00
CXCR4	1.122(1.037-1.215)	0.004	4.52E+13(4.11E+13-4.97E+13)	<0.0001	4.52E+13
hCG_40738	0.701(0.53-0.927)	0.013	0-0	<0.0001	0.00E+00
CHD9	0.825(0.706-0.964)	0.016	0-0	<0.0001	0.00E+00
GCSH	0.827(0.719-0.952)	0.008	0-0	<0.0001	0.00E+00
SUHW1	0.538(0.313-0.923)	0.024	0-0	<0.0001	0.00E+00
RPS6KA5	0.854(0.744-0.981)	0.026	0-0	<0.0001	0.00E+00
PDCD4	0.802(0.664-0.969)	0.022	1.38E+74(1.13E+74-1.69E+74)	<0.0001	1.38E+74
ZG16	0.47(0.269-0.822)	0.008	1.2E+148(6.70E+147-2.06E+148)	<0.0001	1.18E+148
KCNG1	0.725(0.551-0.955)	0.022	0-0	<0.0001	0.00E+00
DECR1	0.804(0.668-0.967)	0.02	0-0	<0.0001	0.00E+00
PPCS	1.267(1.117-1.436)	0	0-0	<0.0001	0.00E+00
SERPINF1	1.116(1.04-1.197)	0.002	1.14E+11(1.05E+11-1.23E+11)	<0.0001	1.14E+11
TMSB10	1.505(1.209-1.873)	0	2.97E+7(2.31E+7-3.83E+7)	<0.0001	2.97E+07
NAT10	1.231(1.032-1.468)	0.021	0-0	<0.0001	0.00E+00
HIC2	0.692(0.517-0.928)	0.014	0.21(0.15-0.3)	<0.0001	2.10E-01
PIR	1.089(1.015-1.168)	0.018	1.18E+6(1.08E+6-1.28E+6)	<0.0001	1.18E+06
OR2W1	0.413(0.179-0.951)	0.038	0-0	<0.0001	0.00E+00

**FIGURE 6 |** Prognostic gene signature of the 22 genes in the GBM patients (TCGA dataset). (A) From top to bottom is the risk score distribution, patients' survival status distribution, and the heat map of the 22 genes for low- and high-risk groups. Each column and each row represents one GEO dataset and one gene respectively, and the gradual color ranging from blue to red represents the changing process from down- to up-regulation. (B) The Kaplan–Meier curves for low- and high-risk groups. (C) The ROC curves for predicting OS in GBM patients by the risk score.

**FIGURE 7 |** Expression of the 22 prognostic genes in low- and high-risk groups (TCGA dataset). Expression values of these genes are log2-transformed.

## Discussion

Integrated bioinformatics analysis has been extensively applied to identify potential biomarkers closely associated with the diagnosis, treatment, and prognosis of GBM. For example, Bo et al identified potential key genes associated with the progression of GBM from GSE31262 gene expression profile data using a bioinformatics approach incorporating DEG screen, pathway enrichment analysis and co-expression network construction [14]; Long et al identified hub genes for diagnosis and prediction of GBM from GSE90886 dataset by developing an integrated method including DEG screen, pathway enrichment analysis, PPI network construction and survival analysis [15]; Tang et al identified new insights into molecular mechanisms underlying GBM and suggested the candidate targets for the diagnosis and treatment of GBM by developing an integrated method including DEG integration, hierarchical clustering, Functional category enrichment analysis, transcription factor and target miRNAs enrichment analysis,

and construction of gene/protein interaction network and analysis.[17] Bo et al further identified key genes in the recurrence of GBM from GSE32466 dataset using a bioinformatics approach incorporating differentially expressed miRNAs screen, hierarchical clustering analysis, target gene prediction of DEMs and construction of regulatory co-expression network [13]; Li et al identified hub DEGs and pathways for GBM by a network-based method that extracted from KEGG pathway database [16]; Kunkle et al revealed potential gene–environment interactions, and generated new data for hypothesis generation in GBM through a comprehensive bioinformatics method using genetic variations (copy number variations and small-scale variations) and environmental data integration that links with Glioblastoma (GEG) [38]; Celiku et al even reported a database called Glioblastoma Bio Discovery Portal (GBM-BioDP) which is a free web-accessible resource that hosts a subset of the glioblastoma TCGA data and enables an intuitive query and interactive display of the resultant data. Moreover, this resource provides visualization tools for the exploration of gene, miRNA, and protein expression, differential expression within the subtypes of GBM, and potential associations with clinical outcome, which are useful for virtual biological validation. [39] Whereas, most of these studies mainly focused on a single database with small sample size or a single bioinformatics approach which results may not that convincing. To overcome these limitations, the current study not only integrated microarray data with relative large sample size from multiple GEO datasets and RNA sequencing data from TCGA, but also constructed PPI networks and a Cox proportional hazards model to identify potential diagnostic and prognostic biomarkers in GBM.

In the present study, nine microarray datasets were integrated with RNA sequencing data downloaded from TCGA, and 512 DEGs between GBM and normal samples were identified, consisting of 212 down-regulated and 300 up-regulated genes. The functional enrichment analysis showed that the down-regulated genes were primarily implicated in various biological processes associated with modulation of chemical synaptic transmission including regulation of trans-synaptic signaling, synaptic vesicle cycle, regulation of vesicle-mediated transport, signal release and regulation of neuron projection development. For the up-regulated genes, they mainly played important roles in extracellular matrix and cell cycle regulation, such as extracellular structure organization, extracellular matrix organization, skeletal system development and cell cycle G1/S phase transition, nuclear division, organelle fission, G1/S transition of mitotic cell cycle, regulation of cell cycle phase transition, and regulation of mitotic cell cycle phase transition. Particularly, many up-regulated genes were enriched in pathways in cancer, such as PI3K-Akt signaling pathway, Cell cycle, p53 signaling pathway, and Cellular senescence, which suggested these genes might be important in carcinogenesis and metastasis of GBM. While the down-regulated genes mainly participated in synaptic chemical signal transduction associated signaling pathways, like cAMP signaling pathway, Neuroactive ligand-receptor interaction, Synaptic vesicle cycle, Endocytosis, Glutamatergic synapse, Adrenergic signaling in cardiomyocytes, and GABAergic synapse. Our findings in the functional enrichment analysis are in line with previous works.[40-42]

We also identified nine hub genes in the PPI network, namely, TP53, FN1, EGFR, MYC, RRM2, EZH2, FOXM1, CD44 and MMP2, and all of them were up-regulated genes in GBM. The TP53 gene mutations are a common mechanism for glial cell neoplasms in the 18-45-year age group but are unrelated to progression and advanced histological grade.[43] Moreover, the mutant TP53 is not only strongly

associated with a poor prognosis for overall survival in patients with glioblastoma, but also may decrease the chemosensitivity of glioblastoma to temozolomide by increasing O<sup>6</sup>-methylguanine DNA-methyltransferase (MGMT) expression.[44] Besides, miR-25 and -32 might serve as positive regulators of p53, underscoring their role in tumorigenesis in glioblastoma.[45] The FN1 gene has been identified as a key gene of GBM [15] and one of the cis-expression quantitative trait locus (eQTL).[46] It's also reported that FN1 is closely related to cell adhesion and invasion in glioblastoma.[47] EGFR is known to contribute to the malignant properties of glioblastoma multiforme (GBM), such as uncontrolled cell proliferation and evasion of apoptosis.[48] The EGFR signaling pathway is thought to play a crucial role in GBM pathogenesis as well, initiating the early stages of tumor development, sustaining tumor growth, promoting infiltration, and mediating resistance to therapy.[49] c-Myc, a crucial regulator of the Warburg effect [50], was regulated by an unexpected Akt-independent role for mTORC2 in GBM metabolic reprogramming. Importantly, overexpression of c-Myc is associated with significantly shortened survival in GBM patients.[51] It has been demonstrated that Myc-mediated regulation of glioblastoma multiforme cell differentiation [52], which might attribute to either promote or reinforce an undifferentiated phenotype required for glioma cells to respond to the oncogenic effects of elevated Ras and Akt activity.[53] Meanwhile, the Myc oncogene has been suggested to cooperatively regulate multiple downstream targets leading to changes in chromosome stability, gene mutations, and/or modulation of tumor growth in GBM. [54] Overexpression of RRM2 has been confirmed to enhance the proliferation, migration, and invasion of cancer cells in GBM [55, 56] and a potential prognostic biomarker with functional significance in GBM patients.[57] Recent study identified a novel role of BRCA1 as a transcriptional co-activator of RRM2, whereby BRCA1-mediated RRM2 expression protects GBM cells from endogenous replication stress, DNA damage and apoptosis.[58] Furthermore, RRM2 was determined to be correlated with glioma grade.[59] EZH2, the lysine methyl transferase of Polycomb repressive complex 2, is a negative prognostic factor and exhibits pro-oncogenic activity in glioblastoma.[60] EZH2 depletion strongly represses the expression of c-myc [61] and reverses the silencing of Polycomb target genes and diminishes STAT3 activity thus suppressing the tumorigenicity of glioblastoma stem-like cells.[62] Additionally, miR-101 and miR-137, are associated with a poor prognostic phenotype of GBM patients, were down-regulated in glioblastoma resulting in EZH2-induced proliferation, migration, and angiogenesis.[63, 64] FoxM1 plays a critical role in the development and progression of GBM by regulating key factors involved in cell proliferation, epithelial to mesenchymal transition (EMT), invasion, angiogenesis and upregulating Wnt/ $\beta$ -catenin signalling.[65, 66] Strongly elevated expression of FoxM1 is associated with poor prognosis, the regulation of metastasis and the preservation of neural, progenitor, and GBM stem cells.[66] Further, FoxM1 maintains the self-renewal and tumorigenicity of glioma stem-like cells through up-regulating the PDGF-A and STAT3 pathway [67], and promotes clonogenic growth and radiation resistance of GBM via FoxM1-Sox2 signaling axis.[68, 69] CD44 is a major cell surface hyaluronan receptor and cancer stem cell marker that participates in the progression of GBM. CD44 depletion blocks GBM growth and sensitizes GBM cells to cytotoxic drugs in vivo.[70] Lower levels of CD44 expression particularly correlated with lower survival of GBM patients which is a promising candidate for further development as a prognostic and therapeutic tool.[71] Specifically, CD44s-targeted treatment with specific mAb may prevent the progression of highly invasive GBMs.[72] It has been demonstrated that the complex balance between TIMP-2 and MMP-2 is a

critical determinant of glioblastoma invasion.[73] Additionally, MMP2 is essential for cancer neovascularization and cancer invasion in promoting extracellular matrix degradation. [74]

The current study identified 22 pivotal genes associated with GBM prognosis and constructed a prognostic gene signature comprised of these genes. Among these 22 genes, RAB33B, KIAA1199, EVC, SOD2, hCG\_40738, CHD9, GCSH, SUHW1, RPS6KA5, KCNG1, DECR1, PPCS, NAT10, OR2W1 and HIC2 were identified as protective prognostic genes. The prognostic value of RPS6KA5, DECR1, NAT10 and HIC2 in GBM has been evaluated before. RPS6KA5 has been identified as a protective prognostic gene in GBM based on bioinformatics approach and is significantly associated with overall survival (OS) of GBM patients.[75] DECR1, an enzyme that required for the mitochondrial  $\beta$  oxidation of lipids[76, 77], is highly up-regulated in GBM.[78] Moreover, a latest study based on co-expression network analysis observed that overexpression of DECR1 is relevant to the overall survival time of patients with GBM.[79] NAT10 belongs to a family of N-terminal acetyltransferases (NATs) which are arranged in complexes and believed to target ~80-90 % of all soluble human proteins. Increased expression of NAT10 correlated negatively with patient survival in the group of “all gliomas” patients.[80] HIC2 gene encodes a hypermethylated in cancer 2 protein in human which functions involve in DNA binding, protein C-terminus binding, metal ion binding and nucleic acid binding and so on. Recent study reported that frequent deletion in 9p11.2 (FOXD4L2 and AQP7P3) and 22q11.21 (HIC2) are significantly associated with short-term survivor of GBM patients based on a deep genomic comparative analysis of a cohort of patients receiving standard therapy.[81] However, the prognostic value of RAB33B, KIAA1199, EVC, SOD2, CHD9, KCNG1 and PPCS in GBM has not been validated in previous studies. RAB33B is concordantly overexpressed in GBM patients [82], whereas, the prognostic value of RAB33B has not been validated in previous studies. KIAA1199 is a newly found gene with significance in senescence, progression, distant metastasis and poor prognosis of cancer patients. Importantly, the KIAA1199 gene is located on chromosome 15q25, where shows a strong linkage to familial gliomas.[83] Motaln et al showed that hMSCs were responsible for the impairment of GBM cell proliferation, invasion and growth by affecting KIAA1199 genes.[84] However, the prognostic value of KIAA1199 in GBM remain buried. EVC transcript were identified in GBM and LGG but not in normal brain tissue, and the expression data suggests that EVC is expressed more in GBM and LGG tumors when compared with normal samples, although these data did not pass their stringent 2-fold difference in expression.[85] SOD2, encoding an intramitochondrial free radical scavenging enzyme located in the mitochondria, was identified as a novel primary GBM-specific marker.[86] Although limited to GBMs, there was a significant relationship between iNOS and SOD1 expression. The induction of both iNOS and SOD1 would be functionally and precisely controlled within the iNOS-inducing tumors.[87] CHD9 genes is highly upregulated in glioblastoma in a silicon analysis using the TCGA [88] and correlated with GBM outcomes by contributing to stress response.[89] Besides, KCNG1 and PPCS remains largely uncharacterized genes. KCNG1 encodes a member of the potassium channel, voltage-gated and subfamily G, thus regulating the potassium channel activity, voltage-gated potassium channel activity, ion channel activity and voltage-gated ion channel activity. It is reported that KCNG1 encodes extracellular antigenic epitopes with the highest difference between ovarian tumors and normal tissues. [90] PPCS is an enzyme that catalyzes the chemical reaction which constitutes the second of five steps

involved in the conversion of pantothenate to Coenzyme A. The molecular function of PPCS consists of regulating phosphopantothenate–cysteine ligase activity and ligase activity. Furthermore, the prognostic value of hCG\_40738, GCSH, SUHW1 and OR2W1 has not been absolutely clarified and further studies are still demanded to validate our findings, the importance of these four genes should not be underestimated.

With regard to the seven risky prognostic genes (CXCR4, TMSB10, TEK, PDCD4, ZG16, PIR and SERPINF1), the correlation of TMSB10 with GBM has been investigated before. Existing studies presented that TMSB10 is identified as a negative prognostic gene of GBM via Robust Likelihood-Based Survival Modeling Analysis.[91] CXCR4 is a cell surface chemokine receptor that have been identified as a robust mediator of glioma cell proliferation and invasion.[92] CXCR4 is proven to play a key part in glioma cell migration through HIF-1 $\alpha$  in the pseudopalisading tumor cells.[93] Additionally, CXCL12 has been known to modulate the CXCR4 and CXCR7 activity in human glioblastoma stem-like cells and regulate the tumor microenvironment.[94] It has been reported that inhibition of CXCL12/CXCR4 autocrine/paracrine loop reduces viability of human glioblastoma stem-like cells affecting self-renewal activity. Moreover, recent evidence suggested that miR-663 negatively regulated CXCR4 to inhibit its oncogenic effect. TEK, also known as EC-specific receptor tyrosine kinase Tie2 (EC, endothelial cell), whose activation is positively and negatively modulated by angiopoietin-1 (Ang1) and angiopoietin-2 (Ang2), respectively. The modulation of Tie2/Tek activation by Ang1 and Ang2 contributes to GBM vascularization and overall growth. Recent studies have found an up-regulation in expression of Ang1, Ang2 and Tie2/Tek with increasing malignancy grade of astrocytomas.[95] It has been reported that a robust increase in Tie2/Tek expression, restricted to the EC of GBMs compared to low-grade astrocytomas and normal brain.[95] Similarly, it is found that Tie2 signal transduction plays a significant regulatory role in the pathological vascular growth of GBMs.[96] Additionally, purified ExTek treatment in intracranial models of GBMs increases the overall survival by approximately 35%. PDCD4, a well-known tumor-suppressor gene, has been identified as a functional target of mir-21.[97] It is demonstrated that downregulation of PDCD4 by mir-21 facilitates glioblastoma proliferation.[98, 99] And the loss of PDCD4 expression was closely associated with the progression and prognosis of various tumors, including glioblastomas.[100] Besides, the loss of PDCD4 contributes to enhanced chemoresistance in GBM via de-repression of Bcl-xL translation.[101] As for PIR, it has been identified as a protective prognostic gene and been reported that high expression of PIR ( $p = 0.025$ , HR 0.62 (0.41–0.95); Q1 cut-off) predicts superior overall survival of GBM patients based on univariate analysis.[102] However, unlike the previous study, our results suggest that high expression of PIR ( $p = 0.018$ , HR 1.089(1.015-1.168)) predicts inferior overall survival of GBM patients. This discrepancy may lie in the reasons as follows: (1) Previous study mainly focused on a specific subventricular zone (SVZ) of GBM patients; (2) All clinical data were obtained through review of medical charts and gathered in Department of Neurosurgery (University Hospital, Heidelberg, Germany) between 1998 and 2011; (3) Previous prognostic significance was determined using univariate and multivariate Cox regression analyses and log-rank tests. In terms of ZG16 and SERPINF1 genes, little is known about their prognostic value in GBM. ZG16, which is significantly decreased in colorectal cancer (CRC) samples compared to adjacent non-tumor tissues and associated with prognosis of CRC patients. Overexpression of ZG16 significantly inhibits growth and sphere

formation of stem-like CRC cells.[103] Multiple mutations were found in the SERPINF1 (Serpin Peptidase Inhibitor, Clade F) gene which encodes PEDF (pigment epithelium-derived factor), a potent inhibitor of angiogenesis and known regulator of bone density.[104] Although the status of ZG16 and SERPINF1 and their correlation with prognosis in GBM have seldom been reported before, they could provide helpful evidence for potential prognostic biomarkers in future studies.

The limitations of our study were as follows: (1) we performed our study based on integrated bioinformatics analysis, our results need further biological experiments validation. (2) we obtained the data used in present study from public databases and we cannot evaluate the quality of these data; (3) our study mainly focused on the genes commonly identified as significantly altered ones in multiple datasets, some biological information (characteristic details such as gender, age, race, tumor grade and stage, etc.) may be overlooked in our study.

## Abbreviations

GBM: Glioblastoma; OS: overall survivals; DEGs: differentially expressed genes; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; RRA: robust rank aggregation; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interactions; MCODE: Molecular Complex Detection; LASSO: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; AUC: area under the curve; HR: hazard ratio; TP53: tumor protein p53; FN1: Fibronectin; EGFR: epidermal growth factor receptor; RRM2: Ribonucleoside-diphosphate reductase subunit M2; EZH2: Enhancer of zeste homolog 2; FOXM1: Forkhead box protein M1; CD44:CD44 antigen; MMP2: matrix metalloproteinase-2

## Declarations

### Acknowledgements

Not applicable.

### Authors' contributions

YT: conception, design, and performance of the research and writing of the paper. XL and ZZ: supervision of the research. SZ and HZ: provision of useful suggestions in methodology. JL and HZ: provision of suggestions in figure preparation. All authors read and approved the final version of the manuscript.

### Funding

None

### Availability of data and materials

All data supporting the findings of this study are included in this published article.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

All authors agreed on the manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Department of Anesthesiology, Shenzhen People's Hospital, The first Affiliated hospital of Southern University of Science and Technology, Shenzhen, Guangdong, 518020, China. <sup>2</sup> School of Nursing, Southern Medical University, Guangzhou, Guangdong, 510515, China. <sup>3</sup> Department of Anesthesiology, Shanghai Tenth People's Hospital, Tenth people's hospital of Tongji university, Shanghai, 200072, China. <sup>4</sup> Department of Anesthesiology, Taihe hospital, Hubei University of Medicine, Shiyan, Hubei, 442000, China.

## References

1. Møller HG, Rasmussen AP, Andersen HH, Johnsen KB, Henriksen M, Duroux M: **A systematic review of microRNA in glioblastoma multiforme: micro-modulators in the mesenchymal mode of migration and invasion.** *Molecular neurobiology* 2013, **47**(1):131-144.
2. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M: **Epidemiology and molecular pathology of glioma.** *Nature clinical practice Neurology* 2006, **2**(9):494-503; quiz 491 p following 516.
3. Cloughesy TF, Cavenee WK, Mischel PS: **Glioblastoma: from molecular pathology to targeted treatment.** *Annual review of pathology* 2014, **9**:1-25.
4. Liu T, Zhang T, Zhou F, Wang J, Zhai X, Mu N, Park J, Liu M, Liu W, Shang P *et al*: **Identification of genes and pathways potentially related to PNF20 by gene expression profile analysis of glioblastoma U87 cell line.** *Cancer cell international* 2017, **17**:87.
5. Song H, Zhang Y, Liu N, Wan C, Zhang D, Zhao S, Kong Y, Yuan L: **miR-92b regulates glioma cells proliferation, migration, invasion, and apoptosis via PTEN/Akt signaling pathway.** *Journal of physiology and biochemistry* 2016, **72**(2):201-211.
6. Aldape KD, Bondy M, F. Bondy, M. L. Wrensch, M.: **Molecular epidemiology of glioblastoma.** *Cancer journal (Sudbury, Mass)* 2003, **9**(2):99-106.
7. Ohgaki H: **Genetic pathways to glioblastomas.** *Neuropathology : official journal of the Japanese Society of Neuropathology* 2005, **25**(1):1-7.

8. Desjardins A, Reardon DA, Coan A, Marcello J, Herndon JE, Bailey L, Peters KB, Friedman HS, Vredenburgh JJ: **Bevacizumab and daily temozolomide for recurrent glioblastoma.** *Cancer* 2012, **118**(5):1302-1312.
9. Ernst A, Hofmann S, Ahmadi R, Becker N, Korshunov A, Engel F, Hartmann C, Felsberg J, Sabel M, Peterziel H *et al*: **Genomic and expression profiling of glioblastoma stem cell-like spheroid cultures identifies novel tumor-relevant genes associated with survival.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15**(21):6541-6550.
10. Dieterich LC, Mellberg S, Langenkamp E, Zhang L, Zieba A, Salomäki H, Teichert M, Huang H, Edqvist PH, Kraus T *et al*: **Transcriptional profiling of human glioblastoma vessels indicates a key role of VEGF-A and TGF $\beta$ 2 in vascular abnormalization.** *The Journal of pathology* 2012, **228**(3):378-390.
11. Vastrad B, Vastrad C, Godavarthi A, Chandrashekar R: **Molecular mechanisms underlying gliomas and glioblastoma pathogenesis revealed by bioinformatics analysis of microarray data.** *Medical oncology (Northwood, London, England)* 2017, **34**(11):182.
12. Roh SW, Abell GC, Kim KH, Nam YD, Bae JW: **Comparing microarrays and next-generation sequencing technologies for microbial ecology research.** *Trends in biotechnology* 2010, **28**(6):291-299.
13. Bo LJ, Wei B, Li ZH, Wang ZF, Gao Z, Miao Z: **Bioinformatics analysis of miRNA expression profile between primary and recurrent glioblastoma.** *European review for medical and pharmacological sciences* 2015, **19**(19):3579-3586.
14. Bo L, Wei B, Li C, Wang Z, Gao Z, Miao Z: **Identification of potential key genes associated with glioblastoma based on the gene expression profile.** *Oncology letters* 2017, **14**(2):2045-2052.
15. Long H, Liang C, Zhang X, Fang L, Wang G, Qi S, Huo H, Song Y: **Prediction and Analysis of Key Genes in Glioblastoma Based on Bioinformatics.** *BioMed research international* 2017, **2017**:7653101.
16. Li W, Li K, Zhao L, Zou H: **Bioinformatics analysis reveals disturbance mechanism of MAPK signaling pathway and cell cycle in Glioblastoma multiforme.** *Gene* 2014, **547**(2):346-350.
17. Tang J, He D, Yang P, He J, Zhang Y: **Genome-wide expression profiling of glioblastoma using a large combined cohort.** *Scientific reports* 2018, **8**(1):15104.
18. Liu X, Wu J, Zhang D, Bing Z, Tian J, Ni M, Zhang X, Meng Z, Liu S: **Identification of Potential Key Genes Associated With the Pathogenesis and Prognosis of Gastric Cancer Based on Integrated Bioinformatics Analysis.** *Frontiers in genetics* 2018, **9**:265.
19. Sun M, Song H, Wang S, Zhang C, Zheng L, Chen F, Shi D, Chen Y, Yang C, Xiang Z *et al*: **Integrated analysis identifies microRNA-195 as a suppressor of Hippo-YAP pathway in colorectal cancer.** *Journal of hematology & oncology* 2017, **10**(1):79.
20. Moradifard S, Hoseinbeyki M, Ganji SM, Minuchehr Z: **Analysis of microRNA and Gene Expression Profiles in Alzheimer's Disease: A Meta-Analysis Approach.** *Scientific reports* 2018, **8**(1):4767.
21. Jiang W, Jia P, Hutchinson KE, Johnson DB, Sosman JA, Zhao Z: **Clinically relevant genes and regulatory pathways associated with NRASQ61 mutations in melanoma through an integrative**

- genomics approach. *Oncotarget* 2015, **6**(4):2496-2508.
22. Kolde R, Laur S, Adler P, Vilo J: **Robust rank aggregation for gene list integration and meta-analysis.** *Bioinformatics (Oxford, England)* 2012, **28**(4):573-580.
  23. Yang J, Han S, Huang W, Chen T, Liu Y, Pan S, Li S: **A meta-analysis of microRNA expression in liver cancer.** *PloS one* 2014, **9**(12):e114533.
  24. Shi KQ, Lin Z, Chen XJ, Song M, Wang YQ, Cai YJ, Yang NB, Zheng MH, Dong JZ, Zhang L *et al*: **Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance.** *Oncotarget* 2015, **6**(28):25093-25108.
  25. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics : a journal of integrative biology* 2012, **16**(5):284-287.
  26. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.** *Nucleic acids research* 2017, **45**(D1):D362-d368.
  27. Gonzalez O, Zimmer R: **Contextual analysis of RNAi-based functional screens using interaction networks.** *Bioinformatics (Oxford, England)* 2011, **27**(19):2707-2713.
  28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome research* 2003, **13**(11):2498-2504.
  29. Liu W, Liang Y, Bian C, Jiang L, Zheng G, Dong J: **Gene expression profile analysis of the bone microenvironment in patients with spinal metastases.** *Oncology letters* 2018, **15**(1):61-68.
  30. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC bioinformatics* 2003, **4**:2.
  31. Motakis EI, A. V. Kuznetsov, V. A.: **Data-driven approach to predict survival of cancer patients: estimation of microarray genes' prediction significance by Cox proportional hazard regression model.** *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society* 2009, **28**(4):58-66.
  32. Liu W, Hao XS, Fan Q, Li HX, Song LN, Wang SJ, Wang PZ, Jin Y, Chen Y, Guan LY *et al*: **[Cox proportional hazard model analysis of prognosis in patients with carcinoma of esophagus and gastric cardia after radical resection].** *Zhonghua zhong liu za zhi [Chinese journal of oncology]* 2008, **30**(12):921-925.
  33. Shukla S, Evans JR, Malik R, Feng FY, Dhanasekaran SM, Cao X, Chen G, Beer DG, Jiang H, Chinnaiyan AM: **Development of a RNA-Seq Based Prognostic Signature in Lung Adenocarcinoma.** *Journal of the National Cancer Institute* 2017, **109**(1).
  34. Chen H, Kong Y, Yao Q, Zhang X, Fu Y, Li J, Liu C, Wang Z: **Three hypomethylated genes were associated with poor overall survival in pancreatic cancer patients.** *Aging* 2019, **11**(3):885-897.
  35. Heagerty PJ, Zheng Y: **Survival model predictive accuracy and ROC curves.** *Biometrics* 2005, **61**(1):92-105.

36. Williams O, Del Genio CI: **Degree correlations in directed scale-free networks.** *PloS one* 2014, **9**(10):e110121.
37. Newman MEJ: **A measure of betweenness centrality based on random walks.** *Social Networks* 2005, **27**(1):39-54.
38. Kunkle B, Yoo C, Roy D: **Discovering gene-environment interactions in glioblastoma through a comprehensive data integration bioinformatics method.** *Neurotoxicology* 2013, **35**:1-14.
39. Celiku O, Johnson S, Zhao S, Camphausen K, Shankavaram U: **Visualizing molecular profiles of glioblastoma with GBM-BioDP.** *PloS one* 2014, **9**(7):e101239.
40. Li X, Wu C, Chen N, Gu H, Yen A, Cao L, Wang E, Wang L: **PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma.** *Oncotarget* 2016, **7**(22):33440-33450.
41. Zheng H, Ying H, Yan H, Kimmelman AC, Hiller DJ, Chen AJ, Perry SR, Tonon G, Chu GC, Ding Z *et al*: **p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation.** *Nature* 2008, **455**(7216):1129-1133.
42. Chhipa RR, Fan Q, Anderson J, Muraleedharan R, Huang Y, Ciraolo G, Chen X, Waclaw R, Chow LM, Khuchua Z *et al*: **AMP kinase promotes glioblastoma bioenergetics and tumour growth.** *Nature cell biology* 2018, **20**(7):823-835.
43. Rasheed BK, McLendon RE, Herndon JE, Friedman HS, Friedman AH, Bigner DD, Bigner SH: **Alterations of the TP53 gene in human gliomas.** *Cancer research* 1994, **54**(5):1324-1330.
44. Wang X, Chen JX, Liu JP, You C, Liu YH, Mao Q: **Gain of function of mutant TP53 in glioblastoma: prognosis and response to temozolomide.** *Annals of surgical oncology* 2014, **21**(4):1337-1344.
45. Suh SS, Yoo JY, Nuovo GJ, Jeon YJ, Kim S, Lee TJ, Kim T, Bakacs A, Alder H, Kaur B *et al*: **MicroRNAs/TP53 feedback circuitry in glioblastoma multiforme.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(14):5316-5321.
46. Dong H, Luo L, Hong S, Siu H, Xiao Y, Jin L, Chen R, Xiong M: **Integrated analysis of mutations, miRNA and mRNA expression in glioblastoma.** *BMC systems biology* 2010, **4**:163.
47. Colin C, Baeza N, Bartoli C, Fina F, Eudes N, Nanni I, Martin PM, Ouafik L, Figarella-Branger D: **Identification of genes differentially expressed in glioblastoma versus pilocytic astrocytoma using Suppression Subtractive Hybridization.** *Oncogene* 2006, **25**(19):2818-2826.
48. Rao RD, Mladek AC, Lamont JD, Goble JM, Erlichman C, James CD, Sarkaria JN: **Disruption of parallel and converging signaling pathways contributes to the synergistic antitumor effects of simultaneous mTOR and EGFR inhibition in GBM cells.** *Neoplasia (New York, NY)* 2005, **7**(10):921-929.
49. Zhu H, Acquaviva J, Ramachandran P, Boskovitz A, Woolfenden S, Pfannl R, Bronson RT, Chen JW, Weissleder R, Housman DE *et al*: **Oncogenic EGFR signaling cooperates with loss of tumor suppressor gene functions in gliomagenesis.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(8):2712-2716.
50. Dang CV: **MYC, metabolism, cell growth, and tumorigenesis.** *Cold Spring Harbor perspectives in medicine* 2013, **3**(8).

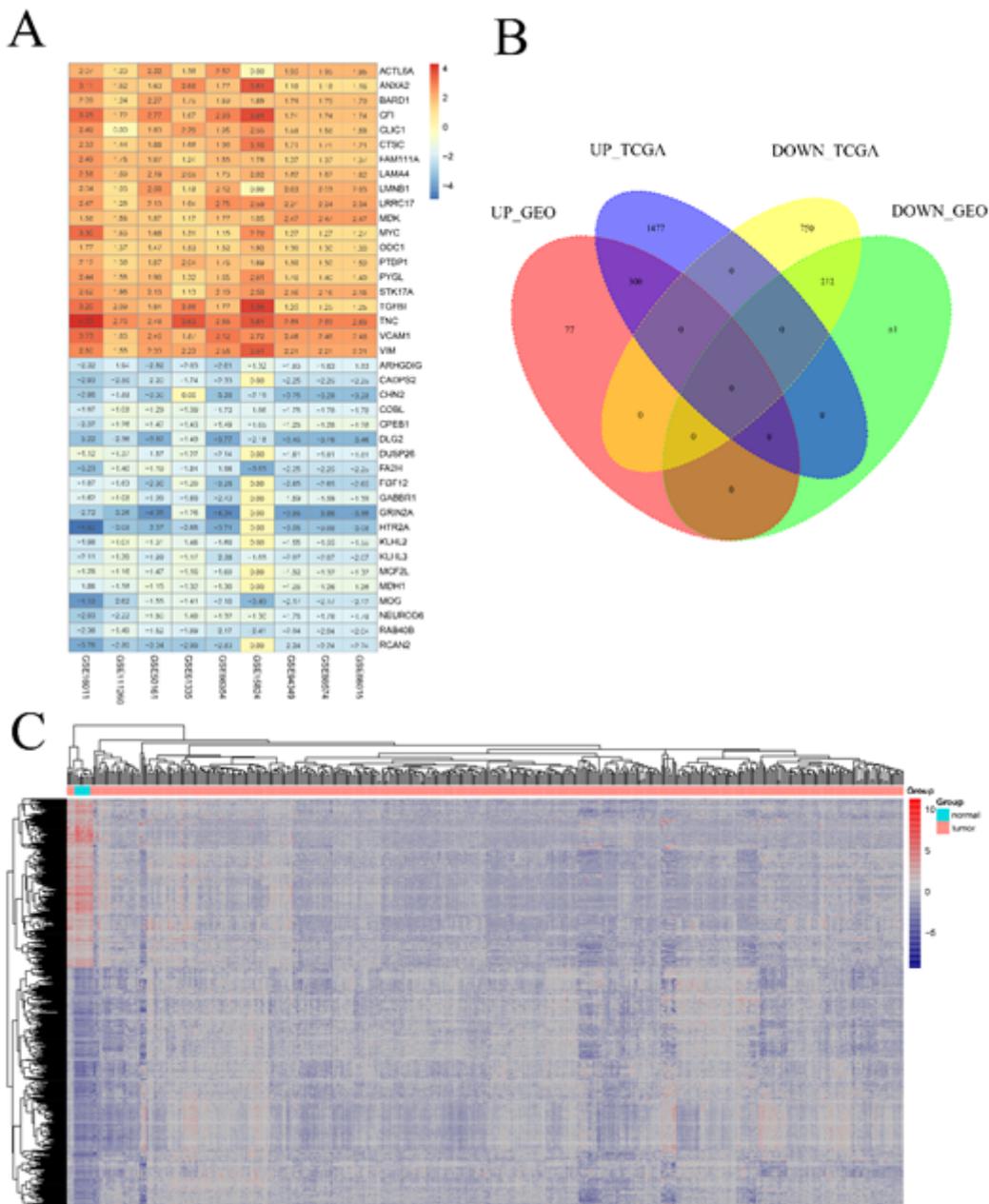
51. Masui K, Tanaka K, Akhavan D, Babic I, Gini B, Matsutani T, Iwanami A, Liu F, Villa GR, Gu Y *et al*: **mTOR complex 2 controls glycolytic metabolism in glioblastoma through FoxO acetylation and upregulation of c-Myc.** *Cell metabolism* 2013, **18**(5):726-739.
52. Panicker SP, Raychaudhuri B, Sharma P, Tipps R, Mazumdar T, Mal AK, Palomo JM, Vogelbaum MA, Haque SJ: **p300- and Myc-mediated regulation of glioblastoma multiforme cell differentiation.** *Oncotarget* 2010, **1**(4):289-303.
53. Lassman AB, Dai C, Fuller GN, Vickers AJ, Holland EC: **Overexpression of c-MYC promotes an undifferentiated phenotype in cultured astrocytes and allows elevated Ras and Akt signaling to induce gliomas from GFAP-expressing cells in mice.** *Neuron glia biology* 2004, **1**(2):157-163.
54. Wang G, Wang J, Zhao H, Wang J, Tony To SS: **The role of Myc and let-7a in glioblastoma, glucose metabolism and response to therapy.** *Archives of biochemistry and biophysics* 2015, **580**:84-92.
55. Li C, Zheng J, Chen S, Huang B, Li G, Feng Z, Wang J, Xu S: **RRM2 promotes the progression of human glioblastoma.** *Journal of cellular physiology* 2018, **233**(10):6759-6767.
56. Yingfei Li ZW, Luoyong Tang, Ronghuan Hu, Long Huang, Jianwu Ding: **RRM2 overexpression in glioblastoma enhances the proliferation and invasion of cancer cells.** *Int J Clin Exp Pathol* 2016, **9**(11):11623-11630.
57. Sun H, Yang B, Zhang H, Song J, Zhang Y, Xing J, Yang Z, Wei C, Xu T, Yu Z *et al*: **RRM2 is a potential prognostic biomarker with functional significance in glioma.** *International journal of biological sciences* 2019, **15**(3):533-543.
58. Rasmussen RD, Gajjar MK, Tuckova L, Jensen KE, Maya-Mendoza A, Holst CB, Møllgaard K, Rasmussen JS, Brennum J, Bartek J *et al*: **BRCA1-regulated RRM2 expression protects glioblastoma cells from endogenous replication stress and promotes tumorigenicity.** *Nature communications* 2016, **7**:13398.
59. Gao YF, Mao XY, Zhu T, Mao CX, Liu ZX, Wang ZB, Li L, Li X, Yin JY, Zhang W *et al*: **COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value.** *Oncotarget* 2016, **7**(43):70494-70503.
60. Zhang J, Chen L, Han L, Shi Z, Zhang J, Pu P, Kang C: **EZH2 is a negative prognostic factor and exhibits pro-oncogenic activity in glioblastoma.** *Cancer letters* 2015, **356**(2 Pt B):929-936.
61. Suvà ML, Riggi N, Janiszewska M, Radovanovic I, Provero P, Stehle JC, Baumer K, Le Bitoux MA, Marino D, Cironi L *et al*: **EZH2 is essential for glioblastoma cancer stem cell maintenance.** *Cancer research* 2009, **69**(24):9211-9218.
62. Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang N, Oh YT, Kim H, Rhee J, Nakano I *et al*: **Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells.** *Cancer cell* 2013, **23**(6):839-852.
63. Smits M, Nilsson J, Mir SE, van der Stoop PM, Hulleman E, Niers JM, de Witt Hamer PC, Marquez VE, Cloos J, Krichevsky AM *et al*: **miR-101 is down-regulated in glioblastoma resulting in EZH2-induced proliferation, migration, and angiogenesis.** *Oncotarget* 2010, **1**(8):710-720.

64. Sun J, Zheng G, Gu Z, Guo Z: **MiR-137 inhibits proliferation and angiogenesis of human glioblastoma cells by targeting EZH2.** *Journal of neuro-oncology* 2015, **122**(3):481-489.
65. Wang Z, Zhang S, Siu TL, Huang S: **Glioblastoma multiforme formation and EMT: role of FoxM1 transcription factor.** *Current pharmaceutical design* 2015, **21**(10):1268-1271.
66. Bowman A, Nusse R: **Location, location, location: FoxM1 mediates beta-catenin nuclear translocation and promotes glioma tumorigenesis.** *Cancer cell* 2011, **20**(4):415-416.
67. Maachani UB, Shankavaram U, Kramp T, Tofilon PJ, Camphausen K, Tandle AT: **FOX M1 and STAT3 interaction confers radioresistance in glioblastoma cells.** *Oncotarget* 2016, **7**(47):77365-77377.
68. Gong AH, Wei P, Zhang S, Yao J, Yuan Y, Zhou AD, Lang FF, Heimberger AB, Rao G, Huang S: **FoxM1 Drives a Feed-Forward STAT3-Activation Signaling Loop That Promotes the Self-Renewal and Tumorigenicity of Glioblastoma Stem-like Cells.** *Cancer research* 2015, **75**(11):2337-2348.
69. Lee Y, Kim KH, Kim DG, Cho HJ, Kim Y, Rheey J, Shin K, Seo YJ, Choi YS, Lee JI *et al*: **FoxM1 Promotes Stemness and Radio-Resistance of Glioblastoma by Regulating the Master Stem Cell Regulator Sox2.** *PloS one* 2015, **10**(10):e0137703.
70. Xu Y, Stamenkovic I, Yu Q: **CD44 attenuates activation of the hippo signaling pathway and is a prime therapeutic target for glioblastoma.** *Cancer research* 2010, **70**(6):2455-2464.
71. Wei KC, Huang CY, Chen PY, Feng LY, Wu TW, Chen SM, Tsai HC, Lu YJ, Tsang NM, Tseng CK *et al*: **Evaluation of the prognostic value of CD44 in glioblastoma multiforme.** *Anticancer research* 2010, **30**(1):253-259.
72. Breyer R, Hussein S, Radu DL, Pütz KM, Gunia S, Hecker H, Samii M, Walter GF, Stan AC: **Disruption of intracerebral progression of C6 rat glioblastoma by in vivo treatment with anti-CD44 monoclonal antibody.** *Journal of neurosurgery* 2000, **92**(1):140-149.
73. Lu KV, Jong KA, Rajasekaran AK, Cloughesy TF, Mischel PS: **Upregulation of tissue inhibitor of metalloproteinases (TIMP)-2 promotes matrix metalloproteinase (MMP)-2 activation and cell invasion in a human glioblastoma cell line.** *Laboratory investigation; a journal of technical methods and pathology* 2004, **84**(1):8-20.
74. Gong J, Zhu S, Zhang Y, Wang J: **Interplay of VEGFa and MMP2 regulates invasion of glioblastoma.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2014, **35**(12):11879-11885.
75. Bao ZS, Zhang CB, Wang HJ, Yan W, Liu YW, Li MY, Zhang W: **Whole-genome mRNA expression profiling identifies functional and prognostic signatures in patients with mesenchymal glioblastoma multiforme.** *CNS neuroscience & therapeutics* 2013, **19**(9):714-720.
76. Thorpe C, Kim JJ: **Structure and mechanism of action of the acyl-CoA dehydrogenases.** *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 1995, **9**(9):718-725.
77. Alphey MS, Yu W, Byres E, Li D, Hunter WN: **Structure and reactivity of human mitochondrial 2,4-dienoyl-CoA reductase: enzyme-ligand interactions in a distinctive short-chain reductase active site.** *The Journal of biological chemistry* 2005, **280**(4):3068-3077.

78. Deighton RF, Le Bihan T, Martin SF, Gerth AMJ, McCulloch M, Edgar JM, Kerr LE, Whittle IR, McCulloch J: **Interactions among mitochondrial proteins altered in glioblastoma.** *Journal of neuro-oncology* 2014, **118**(2):247-256.
79. Yang Y, Yan R, Zhang L, Meng X, Sun W: **Primary glioblastoma transcriptome data analysis for screening survival-related genes.** *Journal of cellular biochemistry* 2019.
80. Mughal AA, Grieg Z, Skjellegrind H, Fayzullin A, Lamkhannat M, Joel M, Ahmed MS, Murrell W, Vik-Mo EO, Langmoen IA *et al*: **Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells.** *Molecular cancer* 2015, **14**:160.
81. Michael Berens BA, Sen Peng, Julianna Ross, Bodour Salhia, Sara Byron, Selene Virk, Harshil Dhruv, Nhan Tran, Andrew Sloan: **GENOMIC CHARACTERIZATION OF SURVIVAL OUTLIERS IN GLIOBLASTOMA MULTIFORME** *Neuro-Oncology* 2014, **16**:96-106.
82. Lo KC, Rossi MR, LaDuca J, Hicks DG, Turpaz Y, Hawthorn L, Cowell JK: **Candidate glioblastoma development gene identification using concordance between copy number abnormalities and gene expression level changes.** *Genes, chromosomes & cancer* 2007, **46**(10):875-894.
83. Paunu N, Lahermo P, Onkamo P, Ollikainen V, Rantala I, Helén P, Simola KO, Kere J, Haapasalo H: **A novel low-penetrance locus for familial glioma at 15q23-q26.3.** *Cancer research* 2002, **62**(13):3798-3802.
84. Zhang Y, Jia S, Jiang WG: **KIAA1199 and its biological role in human cancer and cancer cells (review).** *Oncology reports* 2014, **31**(4):1503-1508.
85. Karunasena E, McIver LJ, Rood BR, Wu X, Zhu H, Bavarva JH, Garner HR: **Somatic intronic microsatellite loci differentiate glioblastoma from lower-grade gliomas.** *Oncotarget* 2014, **5**(15):6003-6014.
86. Reddy SP, Britto R, Vinnakota K, Aparna H, Sreepathi HK, Thota B, Kumari A, Shilpa BM, Vrinda M, Umesh S *et al*: **Novel glioblastoma markers with diagnostic and prognostic value identified through transcriptome analysis.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, **14**(10):2978-2987.
87. Kato S, Esumi H, Hirano A, Kato M, Asayama K, Ohama E: **Immunohistochemical expression of inducible nitric oxide synthase (iNOS) in human brain tumors: relationships of iNOS to superoxide dismutase (SOD) proteins (SOD1 and SOD2), Ki-67 antigen (MIB-1) and p53 protein.** *Acta neuropathologica* 2003, **105**(4):333-340.
88. Ohta S, Kawakami Y, Okano H: **MIF: functions in brain and glioblastoma.** *Oncotarget* 2017, **8**(29):46706-46707.
89. Boyd NH, Walker K, Ayokanmbi A, Gordon ER, Whetsel J, Smith CM, Sanchez RG, Lubin FD, Chakraborty A, Tran AN *et al*: **Chromodomain Helicase DNA-Binding Protein 7 Is Suppressed in the Perinecrotic/Ischemic Microenvironment and Is a Novel Regulator of Glioblastoma Angiogenesis.** *Stem cells (Dayton, Ohio)* 2019, **37**(4):453-462.
90. Zaborowski MP, Lee K, Na YJ, Sammarco A, Zhang X, Iwanicki M, Cheah PS, Lin HY, Zinter M, Chou CY *et al*: **Methods for Systematic Identification of Membrane Proteins for Specific Capture of Cancer-**

- Derived Extracellular Vesicles.** *Cell reports* 2019, **27**(1):255-268.e256.
91. HyungJun Cho AY, Sukwoo Kim, Jaewoo Kang, Seung-Mo Hong: **Robust Likelihood-Based Survival Modeling with Microarray Data.** *Journal of Statistical Software* 2009, **29**(1).
92. Ehtesham M, Winston JA, Kabos P, Thompson RC: **CXCR4 expression mediates glioma cell invasiveness.** *Oncogene* 2006, **25**(19):2801-2806.
93. Ehtesham M, Mapara KY, Stevenson CB, Thompson RC: **CXCR4 mediates the proliferation of glioblastoma progenitor cells.** *Cancer letters* 2009, **274**(2):305-312.
94. Zagzag D, Lukyanov Y, Lan L, Ali MA, Esencay M, Mendez O, Yee H, Voura EB, Newcomb EW: **Hypoxia-inducible factor 1 and VEGF upregulate CXCR4 in glioblastoma: implications for angiogenesis and glioma cell invasion.** *Laboratory investigation; a journal of technical methods and pathology* 2006, **86**(12):1221-1232.
95. Ding H, Roncari L, Wu X, Lau N, Shannon P, Nagy A, Guha A: **Expression and hypoxic regulation of angiopoietins in human astrocytomas.** *Neuro-oncology* 2001, **3**(1):1-10.
96. Zadeh G, Qian B, Okhowat A, Sabha N, Kontos CD, Guha A: **Targeting the Tie2/Tek receptor in astrocytomas.** *The American journal of pathology* 2004, **164**(2):467-476.
97. Lu Z, Liu M, Stribinskis V, Klinge CM, Ramos KS, Colburn NH, Li Y: **MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene.** *Oncogene* 2008, **27**(31):4373-4379.
98. Gaur AB, Holbeck SL, Colburn NH, Israel MA: **Downregulation of Pdc4 by mir-21 facilitates glioblastoma proliferation in vivo.** *Neuro-oncology* 2011, **13**(6):580-590.
99. Chen Y, Liu W, Chao T, Zhang Y, Yan X, Gong Y, Qiang B, Yuan J, Sun M, Peng X: **MicroRNA-21 down-regulates the expression of tumor suppressor PDCD4 in human glioblastoma cell T98G.** *Cancer letters* 2008, **272**(2):197-205.
100. Gao F, Zhang P, Zhou C, Li J, Wang Q, Zhu F, Ma C, Sun W, Zhang L: **Frequent loss of PDCD4 expression in human glioma: possible role in the tumorigenesis of glioma.** *Oncology reports* 2007, **17**(1):123-128.
101. Liwak U, Jordan LE, Von-Holt SD, Singh P, Hanson JE, Lorimer IA, Roncaroli F, Holcik M: **Loss of PDCD4 contributes to enhanced chemoresistance in Glioblastoma multiforme through de-repression of Bcl-xL translation.** *Oncotarget* 2013, **4**(9):1365-1372.
102. Jungk C, Mock A, Exner J, Geisenberger C, Warta R, Capper D, Abdollahi A, Friauf S, Lahrmann B, Grabe N *et al.*: **Spatial transcriptome analysis reveals Notch pathway-associated prognostic markers in IDH1 wild-type glioblastoma involving the subventricular zone.** *BMC medicine* 2016, **14**(1):170.
103. Chen X, Du P, She J, Cao L, Li Y, Xia H: **Loss of ZG16 is regulated by miR-196a and contributes to stemness and progression of colorectal cancer.** *Oncotarget* 2016, **7**(52):86695-86703.
104. Ziff JL, Crompton M, Powell HR, Lavy JA, Aldren CP, Steel KP, Saeed SR, Dawson SJ: **Mutations and altered expression of SERPINF1 in patients with familial otosclerosis.** *Human molecular genetics* 2016, **25**(12):2393-2403.

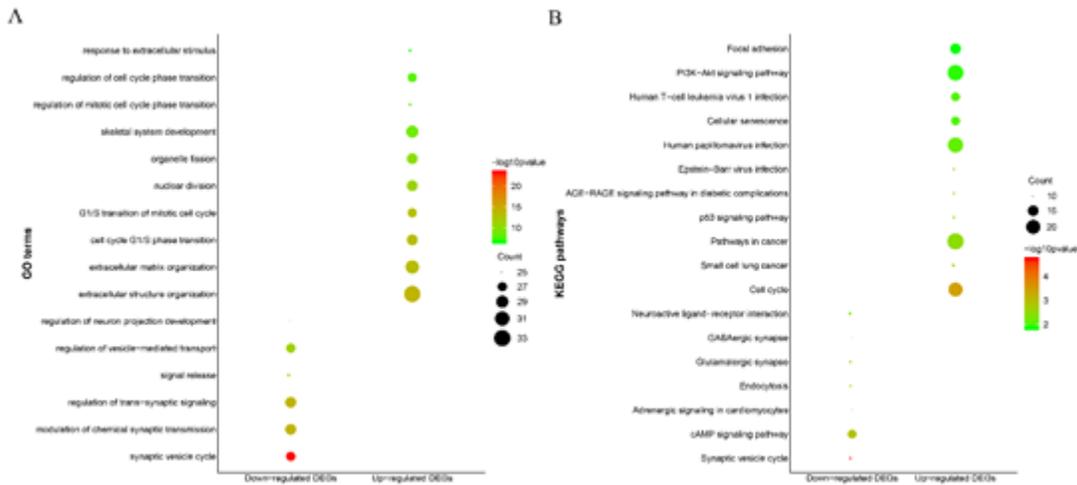
# Figures



**Figure 1**

Identification of DEGs. (A) The heat map of top 20 up- and down-regulated DEGs in the integrated microarray analysis. Each column and each row represents one GEO dataset and one gene respectively. The value in each rectangle was log<sub>2</sub>-transformed. The gradual color ranging from blue to red represents the changing process from down- to up-regulation. (B) Venn diagrams of the overlapping DEGs between the integrated nine GEO datasets and the TCGA GBM dataset. (C) The heat map of 512 overlapping DEGs in GBM and normal brain tissues (TCGA dataset). Each column represents one sample and each row

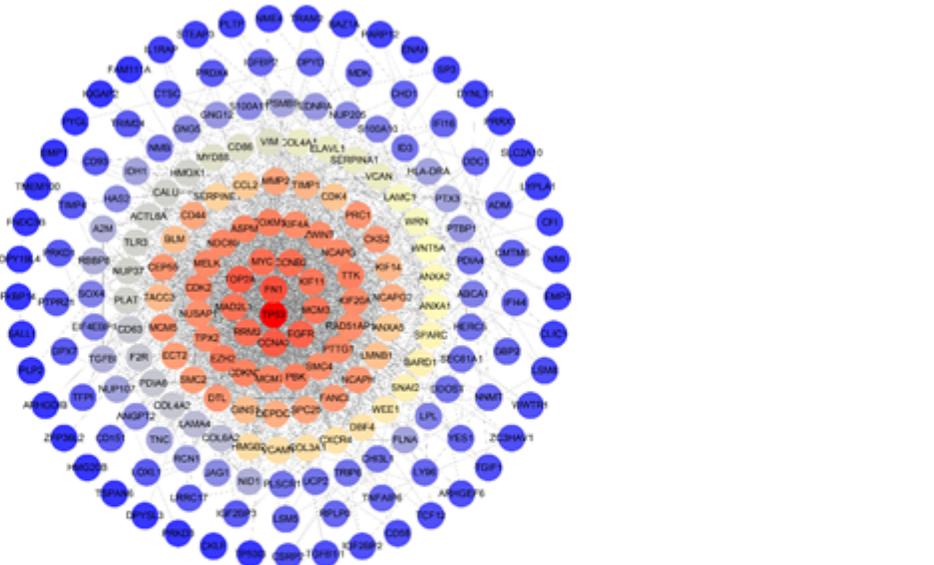
represents one gene. The gradual color ranging from blue to red represents the changing process from down- to up-regulation.



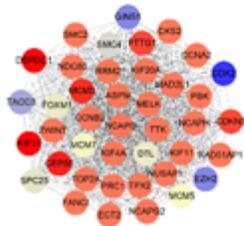
**Figure 2**

Functional enrichment analysis of the overlapping DEGs. (A) GO enrichment analysis of the overlapping DEGs. The y-axis displays the significantly enriched GO biological process terms, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a GO term among all the overlapping DEGs. (B) KEGG pathway enrichment analysis of the overlapping DEGs. The y-axis displays significantly enriched KEGG pathways, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a KEGG pathway among all the overlapping DEGs.

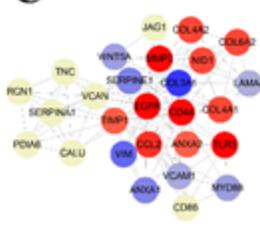
A



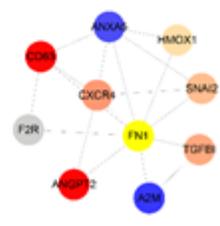
B



C

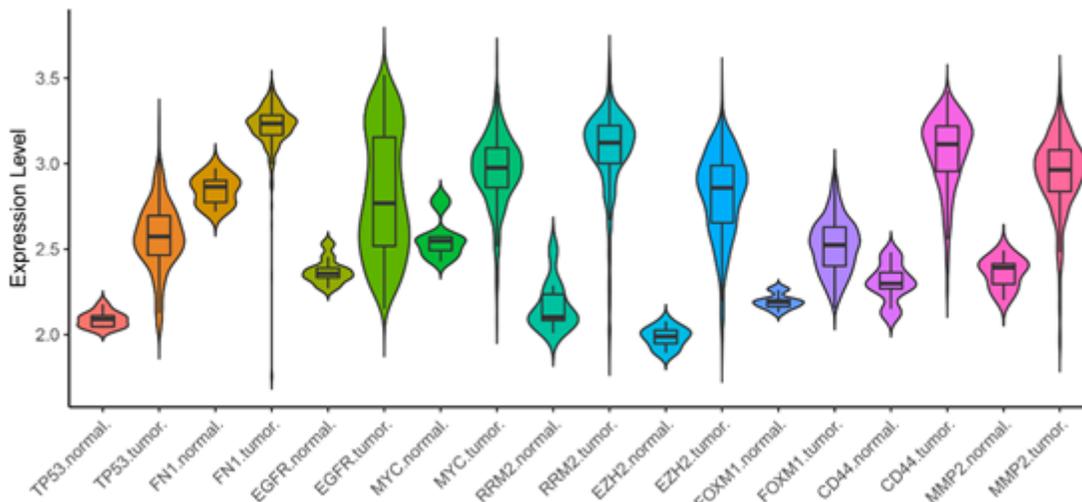


D



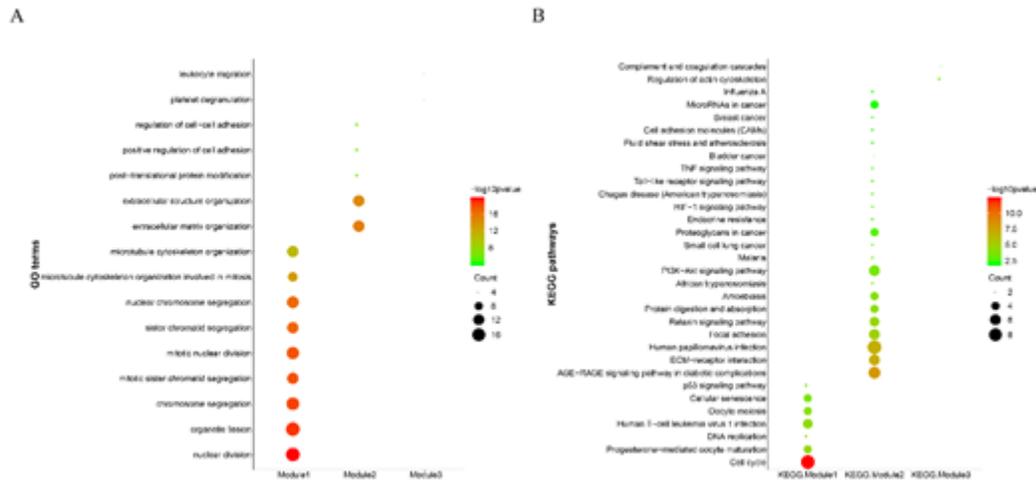
**Figure 3**

Protein–protein interaction (PPI) network and hub clustering modules. (A) The PPI network of overlapping DEGs. (B) Module 1 (MCODE score = 38.8). (C) Module 2 (MCODE score = 9.385). (D) Module 3 (MCODE score = 4). The gradual node color ranging from blue to red represents the changing process from low to high degrees of each gene.



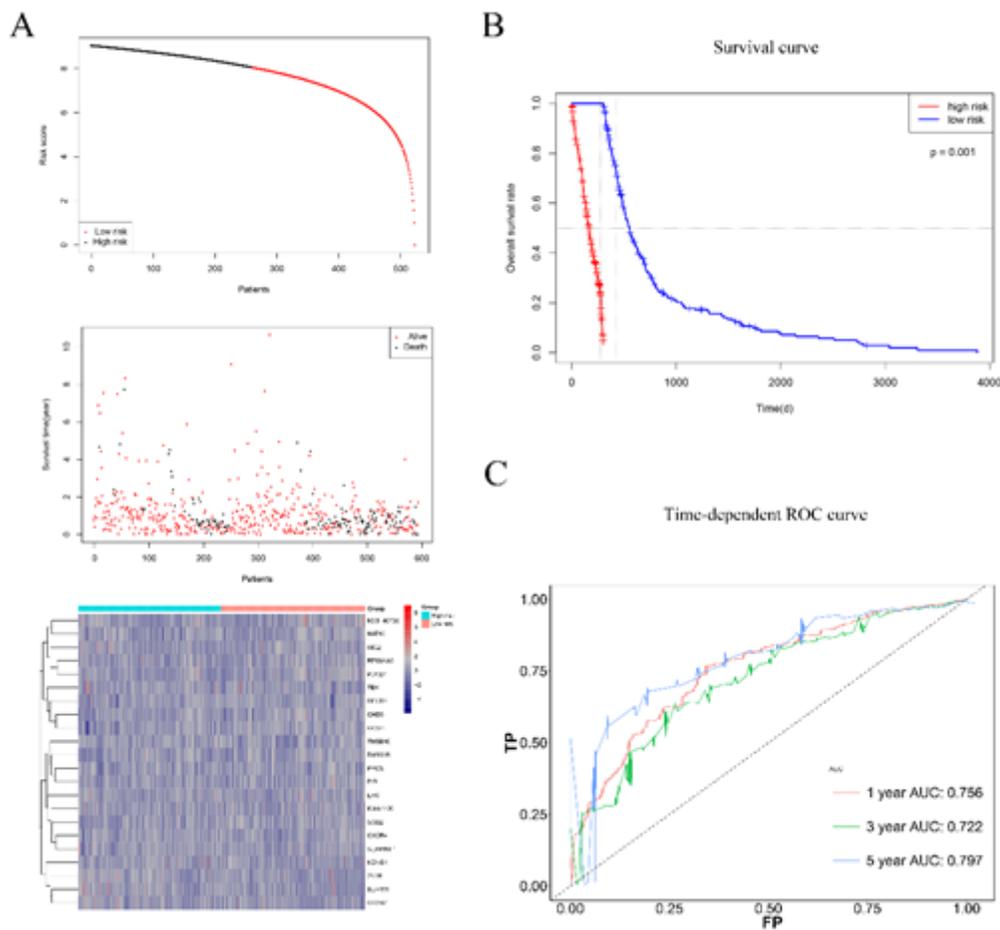
**Figure 4**

Expression of the nine hub DEGs in GBM and normal brain tissues (TCGA dataset). Expression values of these genes are log2-transformed.



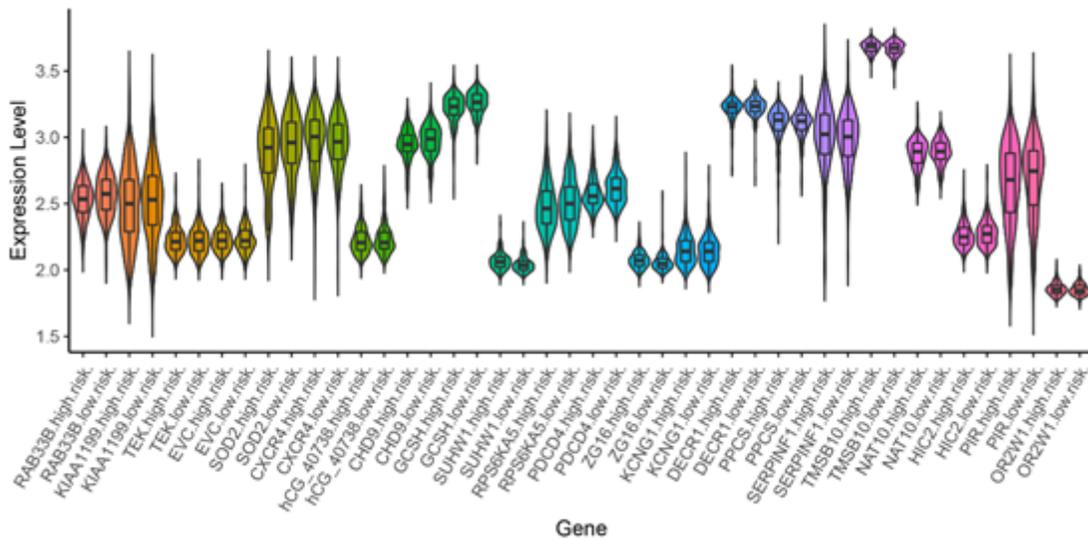
**Figure 5**

Functional enrichment analysis of the overlapping DEGs in the 3 modules. (A) GO enrichment analysis of the overlapping DEGs in the 3 modules. The y-axis displays significantly enriched GO biological process terms, and the x-axis displays different modules. Count refers to the number of genes enriched in a GO term among all the overlapping DEGs. (B) KEGG pathway enrichment analysis of the overlapping DEGs in the 3 modules. The y-axis displays significantly enriched KEGG pathways, and the x-axis displays different gene categories. Count refers to the number of genes enriched in a KEGG pathway among all the overlapping DEGs.



**Figure 6**

Prognostic gene signature of the 22 genes in the GBM patients (TCGA dataset). (A) From top to bottom is the risk score distribution, patients' survival status distribution, and the heat map of the 22 genes for low- and high-risk groups. Each column and each row represents one GEO dataset and one gene respectively, and the gradual color ranging from blue to red represents the changing process from down- to up-regulation. (B) The Kaplan–Meier curves for low- and high-risk groups. (C) The ROC curves for predicting OS in GBM patients by the risk score.



**Figure 7**

Expression of the 22 prognostic genes in low- and high-risk groups (TCGA dataset). Expression values of these genes are log<sub>2</sub>-transformed.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable8.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable10.xlsx](#)
- [SupplementaryTable9.xlsx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable11.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable7.xlsx](#)