

Bioinformatics analyses of significant genes, related pathways and candidate prognostic biomarkers in glioblastoma

LINGQI ZHOU^{1,2*}, HAI TANG^{1-3*}, FANG WANG⁴, LIZHI CHEN^{1,2},
SHANSHAN OU^{1,2}, TONG WU^{1,2}, JIE XU^{1,2} and KAIHUA GUO^{1,2}

¹Department of Anatomy and Neurobiology; ²Guangdong Province Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong 510080;

³Guangdong Jiangmen Chinese Traditional Medicine College, Jiangmen, Guangdong 529000;

⁴Department of Molecular Diagnostics, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510060, P.R. China

Received February 1, 2018; Accepted June 6, 2018

DOI: 10.3892/mmr.2018.9411

Abstract. Glioblastoma (GBM) is the most common type of malignant tumor of the central nervous system. The prognosis of patients with GBM is very poor, with a survival time of ~15 months. GBM is highly heterogeneous and highly aggressive. Surgical removal of intracranial tumors does provide a good advantage for patients as there is a high rate of recurrence. The understanding of this type of cancer needs to be strengthened, and the aim of the present study was to identify gene signatures present in GBM and uncover their potential mechanisms. The gene expression profiles of GSE15824 and GSE51062 were downloaded from the Gene Expression Omnibus database. Normalization of the data from primary GBM samples and normal samples in the two databases was conducted using R software. Then, joint analysis of the data was performed. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed, and the protein-protein interaction (PPI) network of the differentially expressed genes (DEGs) was constructed using Cytoscape software. Identification of prognostic biomarkers was conducted using UALCAN. In total, 9,341 DEGs were identified in the GBM samples, including 9,175 upregulated genes and 166 downregulated genes. The top 1,000 upregulated DEGs and all of the downregulated DEGs were selected for GO, KEGG

and prognostic biomarker analyses. The GO results showed that the upregulated DEGs were significantly enriched in biological processes (BP), including immune response, cell division and cell proliferation, and the downregulated DEGs were also significantly enriched in BP, including cell growth, intracellular signal transduction and signal transduction by protein phosphorylation. KEGG pathway analysis showed that the upregulated DEGs were enriched in circadian entrainment, cytokine-cytokine receptor interaction and maturity onset diabetes of the young, while the downregulated DEGs were enriched in the TGF- β signaling pathway, MAPK signaling pathway and pathways in cancer. All of the downregulated genes and the top 1,000 upregulated genes were selected to establish the PPI network, and the sub-networks revealed that these genes were involved in significant pathways, including olfactory transduction, neuroactive ligand-receptor interaction and viral carcinogenesis. In total, seven genes were identified as good prognostic biomarkers. In conclusion, the identified DEGs and hub genes contribute to the understanding of the molecular mechanisms underlying the development of GBM and they may be used as diagnostic and prognostic biomarkers and molecular targets for the treatment of patients with GBM in the future.

Introduction

Glioma is the most common type of tumor in the central nervous system. According to the literature (1,2), the incidence of gliomas is ~12 per 100,000 people and the proportion of intracranial tumors that are gliomas is ~35-60%, which is a serious harm to human health. As stated in the pathological classification of the WHO (2007) (3), the degree of differentiation of tumor cells can be divided into four levels, with the higher the grade of the tumor, the worse the prognosis of the patient with GBM. Level IV of the WHO classification has a very poor prognosis. According to statistics, each year ~77,000 new cases of patients with glioblastoma (GBM) are diagnosed in the United States and Europe, but the follow-up 5-year survival rate is <3% (4). Compared with other pathological types of glioma, the survival of patients with GBM is also calculated weekly and the majority of

Correspondence to: Professor Kaihua Guo or Dr Jie Xu, Department of Anatomy and Neurobiology, Zhongshan School of Medicine, Sun Yat-sen University, 74 Zhongshan Road, Guangzhou, Guangdong 510080, P.R. China
E-mail: xujie@mail.sysu.edu.cn
E-mail: guokh@mail.sysu.edu.cn

*Contributed equally

Key words: glioblastoma, bioinformatics, key genes, pathways, prognosis

patients with GBM die due to tumor recurrence. It is therefore imperative to find a new solution in order to treat patients with GBM.

Due to the popularization and development of supercomputer technology, high-throughput platforms for the analysis of gene expression are growing. Microarray analysis and next-generation sequencing are increasingly valued as essential tools in the field of medical oncology, with great clinical applications such as molecular classification of cancers, molecular diagnosis, prognosis prediction, patient stratification, tumor response prediction and new drug targets discovery (5-7). A number of tumor-related databases have been established, such as Gene Expression Omnibus (GEO) (8) and TCGA (9). Further, the uses of bioinformatics analysis are changing constantly. The 2016 edition of the WHO pathological grading criteria was the first to include guidelines for the molecular markers of glioma, indicating that genetic molecular diagnosis and gene analysis of the disease play an increasingly important role (10).

In the present study, original data (GSE15824 and GSE51062) were downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). A combined analysis was conducted using limma package (11) in R software (version 3.3.2). In total, 74 GBM samples and 5 normal samples were selected to identify differentially expressed genes (DEGs). Subsequently, the DEGs were screened using Morpheus online software (<https://software.broadinstitute.org/morpheus/>), followed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. By analyzing the biological functions and pathways of the DEGs, the present study provides insight into the development of GBM at the molecular level and identifies candidate biomarkers for the future treatment, diagnosis and prognosis of patients with GBM.

Materials and methods

Microarray data. GSE15824 and GSE51062 were downloaded from the GEO database. Both datasets were based on the Affymetrix GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). In total, 5 normal samples and 22 primary GBM samples from GSE15824 and 52 primary GBM samples from GSE51062 were selected.

Data quality control. Affy package (12) R software was used to perform quality control of all of the microarray CEL files and RNA degradation plots were drawn at the same time. Good uniformity of the curve due to the quality of the RNA extraction with a small level of error laid the foundation for the merging operation of the two datasets.

Identification of DEGs. Analysis of gene expression was conducted using limma package (11) in R software. The Affy package (12) Robust Multichip Average (13) algorithm was used to identify DEGs with a fold-change ≥ 2 and a P-value cutoff of < 0.05 was defined as statistically significant. A heatmap of the DEGs was produced using Morpheus online software and it showed the top 10 upregulated and downregulated genes, respectively. All significant DEGs were shown in a volcano plot generated using R software.

GO and pathway enrichment analyses of DEGs. GO analysis is a common and useful method for annotating genes and gene products and for identifying characteristic biological attributes of high-throughput genome or transcriptome data (14,15). KEGG (<http://www.genome.jp/>) is a well-known database for systematic analysis of gene functions in biological pathways, which links genomic information with higher-order functional information (16). The DEGs were mapped to the relevant biological annotation using the DAVID online database (<https://david.ncifcrf.gov/>) (17). DAVID is an essential online tool for high-throughput gene functional analysis, which provides the functionality to perform simultaneous GO and KEGG analysis. $P < 0.05$ was considered to indicate a statistically significant difference.

Integration of protein-protein interaction (PPI) network and module analysis. Search Tool for the Retrieval of Interacting Genes (18) (STRING, version 10.0, <http://string-db.org/>) was used to conduct PPI network and module analysis. STRING (version 10.0) includes 9.6 million proteins from 2,031 organisms. These proteins have 184 million interactions. The top 1,000 upregulated DEGs and all of the downregulated DEGs were uploaded to the search tool. Only experimentally-validated interactions with a combined score > 0.4 were selected as significant. The whole giant network was then inputted into Cytoscape (19) software (version 3.4) to evaluate primary modules of the PPI sub-network using the plug-in Molecular Complex Detection (MCODE). The criteria were set as default. GO analysis and KEGG pathway enrichment analyses were performed for DEGs in the modules using DAVID. $P < 0.05$ was considered to indicate a statistically significant difference.

Identification of prognostic biomarker. UALCAN (20) (<http://ualcan.path.uab.edu>) was used to identify prognostic biomarkers in the top 100 regulated DEGs. UALCAN was used to construct an algorithm based on the TCGA level 3 RNA-seq database. All data were divided into 31 types according to the type of cancer. The top 100 upregulated and downregulated DEGs, respectively, were uploaded to the UALCAN GBM database. $P < 0.05$ was considered to indicate a statistically significant difference. The influence of genes and sex as an influencing factor on the prognosis of disease were retrieved. Genes with insufficient data were automatically eliminated.

Results

Microarray chip quality control. Affy package in R software was used to conduct quality control of all of the microarray CEL files. The RNA degradation plot shows that the two datasets exhibit a good consistency in the slope of the line. All of the samples have a lower numerical value in the 5' end, which is shown in Fig. 1.

Identification of DEGs. A total 74 primary GBM samples and five normal samples were analyzed. By using R software and the Morpheus online tool, 9,341 DEGs were identified in the GBM samples, which included 9,175 upregulated genes and 166 downregulated genes. A heatmap of the top 10 upregulated and downregulated DEGs, respectively, was produced using the Morpheus online tool. A volcano plot of all of the

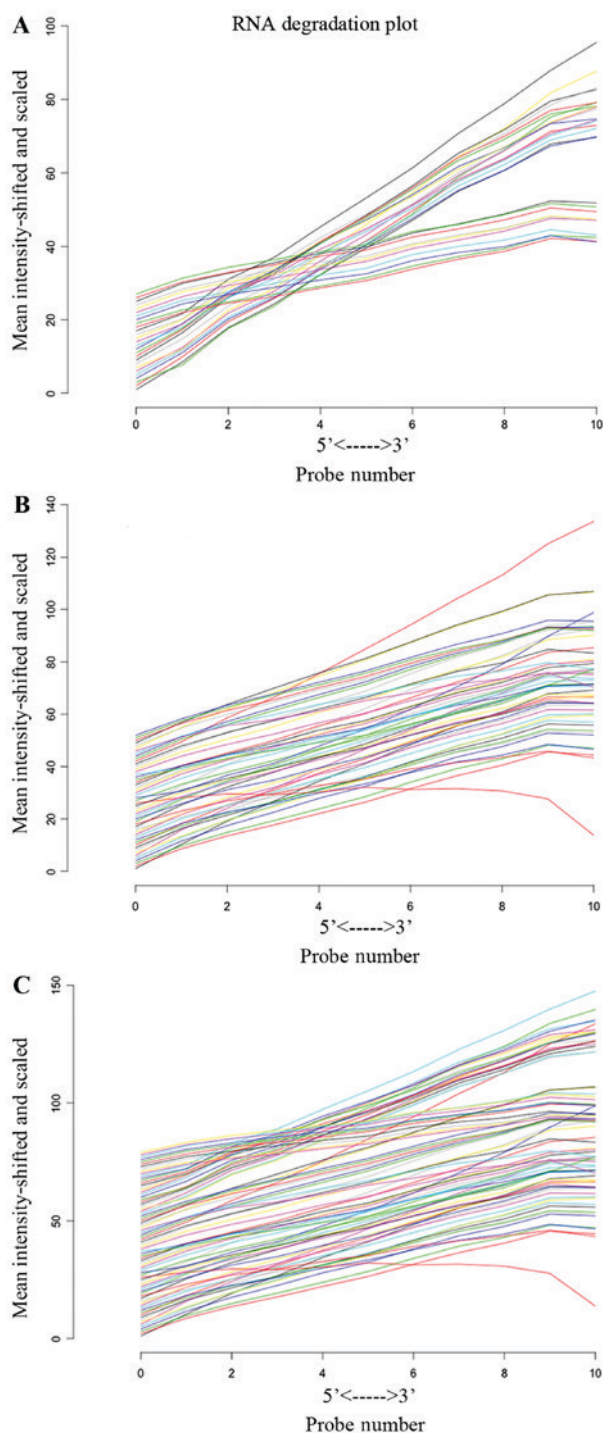


Figure 1. RNA degradation plot of GEO dataset. (A) RNA degradation plot of GSE15824; (B) RNA degradation plot of GSE51062; (C) RNA degradation plot of GSE51062 and GSE15824. GEO, Gene Expression Omnibus.

DEGs was generated using ggplot2 package and R software. The heatmap and volcano plot are shown in Figs. 2 and 3, respectively.

GO term enrichment analysis. The online software DAVID was used to identify overrepresented GO categories and KEGG pathways. The top 1,000 upregulated DEGs and all of the downregulated DEGs were uploaded to the gene list separately. GO analysis showed that the upregulated DEGs were significantly enriched in biological processes (BP), including

defense response, nervous system development and cell development. The downregulated DEGs were also significantly enriched in BP, including regulation of signaling, positive regulation of transport and regulation of cell communication. As for molecular function, the upregulated DEGs were enriched in molecular transducer activity, receptor activity and transmembrane receptor activity and the downregulated DEGs were enriched in cytokine receptor binding, metal ion binding and growth factor activity. In addition, with regards to cell components, the upregulated DEGs were enriched in intrinsic component of plasma membrane, integral component of plasma membrane and extracellular space, and the downregulated DEGs were enriched in regulation of signaling, positive regulation of transport and regulation of cell communication. The results are shown in full in Fig. 4.

KEGG pathway analysis. The enrichment of the top 1,000 upregulated DEGs and all of the downregulated DEGs was identified using KEGG analysis separately. The top 1,000 upregulated DEGs were enriched in circadian entrainment, cytokine-cytokine receptor interaction and maturity onset diabetes of the young, while all of the downregulated DEGs were enriched in TGF- β signaling pathway, MAPK signaling pathway and pathways in cancer. The results are shown in full in Table I.

Module screening from the PPI network. All of the downregulated DEGs and the top 1,000 upregulated DEGs were uploaded to STRING to draw the PPI network (data not shown). Cytoscape was used to evaluate primary modules of the PPI sub-network using the plug-in MCODE. A total of 1,222 nodes and 5,223 edges were analyzed in Cytoscape and the top two sub-networks are displayed in Fig. 5. There are 42 genes involved in sub-network 1, 13 genes involved in sub-network 2 and 47 genes involved in sub-network 3. GO and KEGG analyses were conducted separately for genes in the three top-score sub-networks. All of the results are shown in Tables II and III. In addition, the top 10 hub genes were identified using STRING software and were as follows: AURKB, RACGAP1, MCM2, MCM5, TGFBR1, SMAD3, RPN1, RPN2, VEGFC and FLT4.

Identification of prognostic biomarkers. After verification, the top four genes related to survival were identified as prognostic biomarkers of GBM and were as follows: RANBP17, ZNF734, NLRP2 and GPR1. In addition, the top four prognostic genes affected by sex were also identified and were as follows: CCDC81, NLRP2, SH3RF1 and TM7SF4. All of the results are shown in Figs. 6 and 7.

Discussion

Given the heterogeneity of GBM, which includes a variety of gene expression patterns and imaging-based phenotypes, understanding the molecular mechanism of GBM is of critical importance for the diagnosis and treatment of patients with GBM (21). As high-throughput sequencing and microarray analyses provide data on thousands of expressed genes in the human genome simultaneously, these techniques have been widely used to predict the potential therapeutic targets for

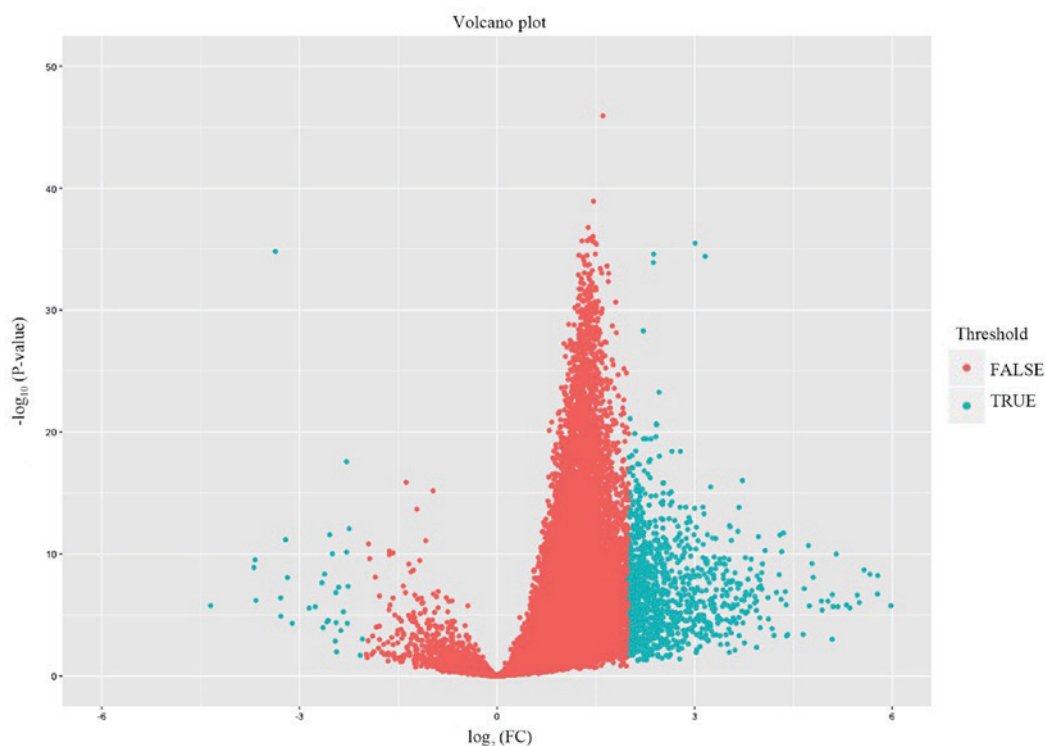


Figure 2. Volcano plot of all significant DEGs, including 9,175 upregulated genes and 166 downregulated genes. TRUE means genes alteration over the threshold by default set. FALSE means genes alteration under the threshold by default set. The default set of threshold was foldchange ≥ 2 , $P \leq 0.05$. DEGs, differentially expressed genes.

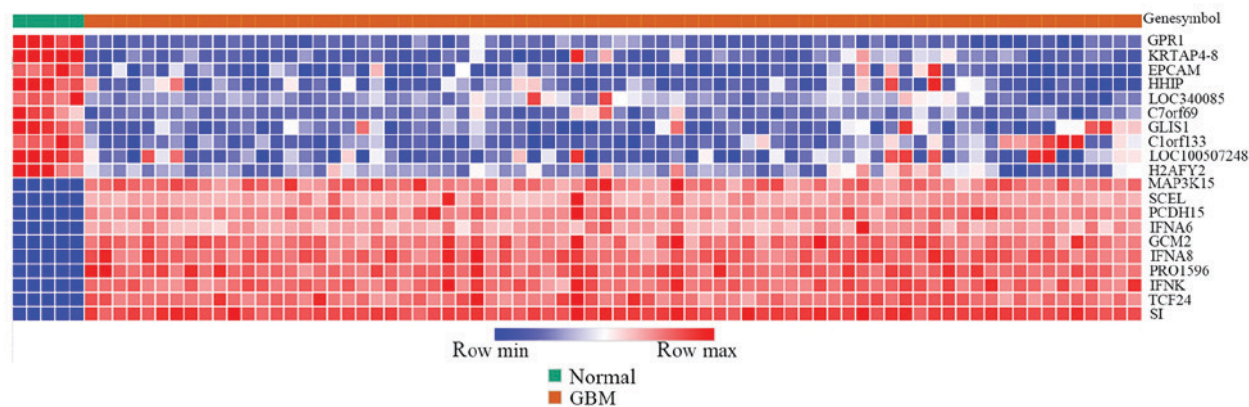


Figure 3. Heatmap of top 10 upregulated DEGs and top 10 downregulated DEGs. Red, upregulation; Blue, downregulation. The value of expression intensity are based on the gene expression level analysis by R software. DEGs, differentially expressed genes.

GBM. In the present study, data from two datasets, GSE15824 and GSE51062, was extracted to identify the DEGs between GBM and normal brain control samples. The results showed that there are 9,175 upregulated and 166 downregulated DEGs between the GBM and normal brain control samples by using bioinformatics analysis. According to the function annotation results of the data, these DEGs were mainly involved in the defense response, cell development, regulation of signaling and of cell communication. By analyzing the PPI network, a number of key genes were identified that may provide new approaches for therapeutic studies of GBM.

In the present study, microarray data of 74 GBM samples and 5 normal brain control samples were downloaded from the GEO datasets under the accession numbers GSE15824 and

GSE51062. Using a series of essential packages in R software, all raw CEL data from the two datasets was extracted and microarray chip quality control was conducted. It is worth noting that there were many more upregulated than downregulated genes. In order to understand the interactions of the DEGs, GO and KEGG pathway analyses were performed. GO term analysis of the top 1,000 upregulated genes and all of the downregulated genes showed that the upregulated DEGs were mainly involved defense response, nervous system development and cell development, and the downregulated DEGs were involved in regulation of signaling, positive regulation of transport and regulation of cell communication. Genes that control tumor progression and enhancement of tumor defense were upregulated and genes that control tumor signaling and cell communication were

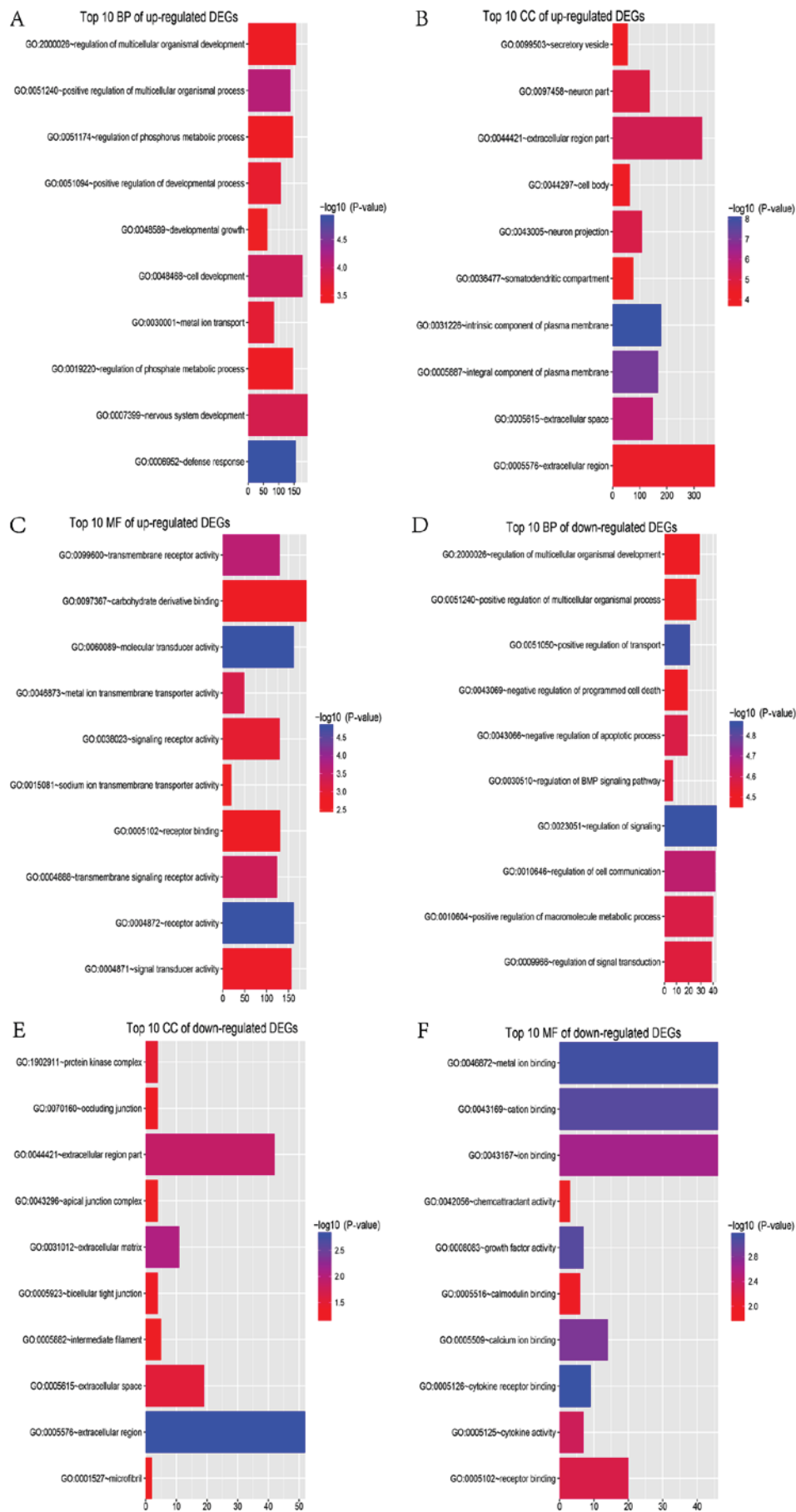


Figure 4. Top 10 GO analysis of top 1,000 upregulated and all downregulated DEGs associated with GBM. (A) Top 10 BP of upregulated DEGs; (B) Top 10 CC of upregulated DEGs; (C) Top 10 MF of upregulated DEGs; (D) Top 10 BP of downregulated DEGs; (E) Top 10 CC of downregulated DEGs; (F) Top 10 MF of downregulated DEGs. GO, Gene ontology; DEGs, differentially expressed genes; BP, biological processes; CC, cellular component; MF, molecular function.

Table I. KEGG pathway analysis of DEGs associated with GBM.

A, Upregulated						
Pathway ID	Name	Gene count	%	P-value	Genes	
hsa04713	Circadian entrainment	18	1.20	4.60E-04	ADCY1, ADCYAP1R1, GRIA3, KCNJ3, KCNJ5, RPS6KA5, GNGT1, FOS, GRIA2, RYR3, RYR1, RYR2, GUCY1A3, GUCY1B3, ADCY10, RASD1, GNG7, MTNR1A	
hsa04060	Cytokine-cytokine receptor interaction	31	2.07	0.0013	OSMR, BMPR2, TNFSF15, CXCL11, PF4V1, IFNA2, IFNA1, IL23A, IFNA6, CXCR4, IL10RB, TNFRSF19, IFNA8, IFNK, IL13RA1, THPO, IL3, IL18RAP, IL5, FLT4, TGFB1, IL9, IL26, TNFSF8, CCR8, TNFSF11, PRLR, PPBP, PDGFRB, IL12B, IL22RA2	
hsa04950	Maturity onset diabetes of the young	8	0.54	0.0021	ONECUT1, FOXA3, IAPP, SLC2A2, PAX4, HNF4G, NR5A2, NKX2-2	
hsa04080	Neuroactive ligand-receptor interaction	34	2.28	0.0035	OPRM1, GLRA1, DRD3, TACR1, ADCYAP1R1, GABBR1, GNRHR, SCTR, AGTR2, S1PR1, GRID2, TAAR1, TAAR2, GABRD, GABRA4, RXFP3, RXFP2, GRIA3, NPY1R, NTSR2, FSHR, P2RX7, P2RY10, ADRB2, GABRR1, PRLR, CHRM3, GRIA2, P2RX3, CHRN4, ADRA1B, ADRA1A, GHSR, MTNR1A	
hsa04630	Jak-STAT signaling pathway	19	1.27	0.0185	PIK3CG, IL3, IL5, OSMR, IL9, IFNA2, IFNA1, IL23A, PRLR, IFNA6, IL10RB, IFNA8, IL12B, IFNK, IL13RA1, MYC, THPO, IL22RA2, IL13RA2	
B, Downregulated						
Pathway ID	Name	Gene count	%	P-value	Genes	
hsa04350	TGF- β signaling pathway	4	2.70	0.0181	BMP4, INHBA, NOG, GDF5	
hsa04010	MAPK signaling pathway	6	4.05	0.0262	CACNA2D1, BDNF, NTF3, RASGRF2, PPP3CB, CHUK	
hsa05200	Pathways in cancer	7	4.73	0.0445	BMP4, VEGFC, CCND1, GNAI1, HHIP, FZD5, CHUK	
hsa05217	Basal cell carcinoma	3	2.03	0.0515	BMP4, HHIP, FZD5	
hsa04068	FoxO signaling pathway	4	2.70	0.0591	CCND1, PLK2, IRS1, CHUK	

GBM, glioblastoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

downregulated, which shows genome alteration consistent with the clinical manifestations. As observed in clinical practice, many patients with low-grade gliomas develop a high-grade glioma after recurrence. Genes controlling tumor progression may be good biomarkers for postoperative recurrence monitoring. A previous study has indicated that MGMT methylation status has great importance in tumor prognosis, especially in patients who have undergone the first surgery (22). Although MGMT methylation status was not reflected in the 2016 edition

of the WHO pathological grading criteria, it is still an influential gene biomarker for the prognosis of patients with GBM.

KEGG analysis of the upregulated genes in the present study identified that they are mainly involved in cytokine relevant pathways and ligand-receptor relevant pathways, including the pathway of neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction and Jak-STAT signaling pathway. The downregulated genes were mainly involved in pathways of classical tumor signaling, including TGF- β signaling

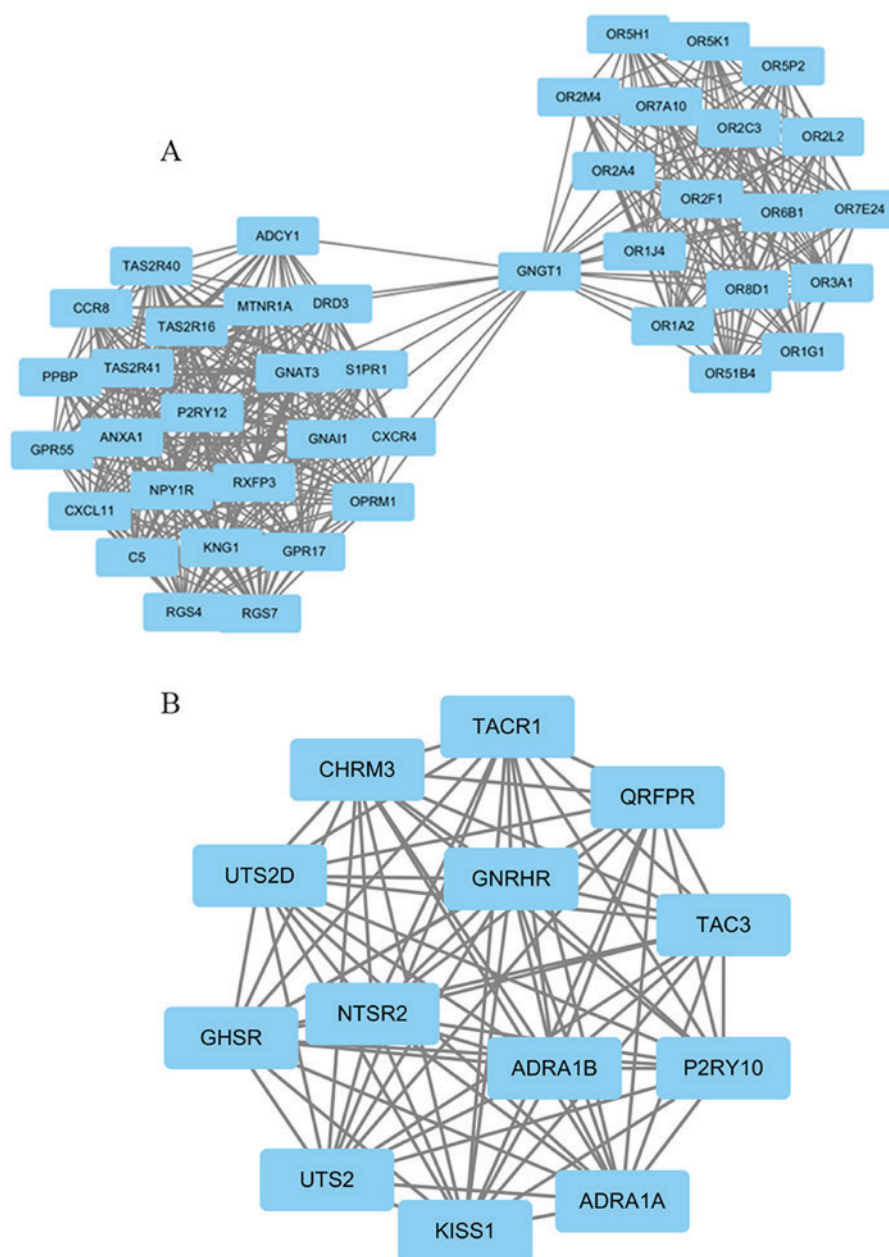


Figure 5. Top 2 primary modules of PPI sub-network by plug-in MCODE in Cytoscape software. (A) module 1; (B) module 2. PPI, protein-protein interaction; MCODE, Molecular Complex Detection.

pathway, MAPK signaling pathway, pathways in cancer, basal cell carcinoma and FoxO signaling pathway (23-28). It is noteworthy that a number of the downregulated genes (BMP4, FZD5 and HHIP) in the GBM samples are involved in the basal cell carcinoma pathway, which may be because GBM originates from the source of nerve epithelium. These three genes may exert some of the same effects in both basal cell carcinoma and GBM. On the other hand, it is also notable that the majority of the downregulated genes are not involved in KEGG pathways and therefore exist in a scattered form.

The PPI network of the DEGs was constructed using the plug-in MCODE in Cytoscape and the top degree hub genes were listed as follows: AURKB, RACGAP1, MCM2, MCM5, TGFBR1, SMAD3, RPN1, RPN2, VEGFC and FLT4. These genes were identified by the degree of connection of every gene according to the connectivity measured by the prior

analysis with STRING. All of the hub genes are upregulated except VEGFC. AURKB was identified as one of the hub genes which exhibits the highest degree of connectivity. AURKB encodes Aurora B, which is an important kinase that regulates cell mitosis. A recent study claims that Aurora B activity is an important modulator of the taxane response in non-small cell lung cancer cells (29). A 2012 study of GBM showed that inhibition of Aurora kinases enhances chemosensitivity to temozolomide and causes radiosensitization in GBM cells (30). RACGAP1 encodes Rac GTPase activating protein 1, a component of the centralspindlin complex that serves as a microtubule-dependent and Rho-mediated signal required for myosin contractile ring formation during cell cycle cytokinesis. RACGAP1 plays key roles in the control of cell growth and differentiation of hematopoietic cells through mechanisms other than regulating Rac GTPase activity. A

Table II. GO analysis of top 3 primary modules of PPI sub-network genes.

Sub-network	Category	Term/gene function	Gene count	%	P-value
Module 1	GOTERM_BP_FAT	GO:0007186~G-protein coupled receptor signaling pathway	41	97.62	3.95E-44
	GOTERM_BP_FAT	GO:0007606~sensory perception of chemical stimulus	21	50	1.25E-19
	GOTERM_BP_FAT	GO:0050907~detection of chemical stimulus involved in sensory perception	20	47.62	4.32E-19
	GOTERM_MF_FAT	GO:0004930~G-protein coupled receptor activity	31	73.81	7.39E-29
	GOTERM_MF_FAT	GO:0004871~signal transducer activity	35	83.33	6.72E-26
	GOTERM_MF_FAT	GO:0004888~transmembrane signaling receptor activity	31	73.81	6.75E-24
	GOTERM_CC_FAT	GO:0098552~side of membrane	8	19.05	4.08E-06
	GOTERM_CC_FAT	GO:0005834~heterotrimeric G-protein complex	4	9.52	1.84E-05
	GOTERM_CC_FAT	GO:0031226~intrinsic component of plasma membrane	12	28.57	2.00E-05
Module 2	GOTERM_BP_FAT	GO:0007200~phospholipase C-activating G-protein coupled receptor signaling pathway	7	58.33	4.05E-11
	GOTERM_BP_FAT	GO:0007186~G-protein coupled receptor signaling pathway	11	91.67	6.50E-11
	GOTERM_BP_FAT	GO:0006940~regulation of smooth muscle contraction	5	41.67	5.28E-08
	GOTERM_MF_FAT	GO:0004930~G-protein coupled receptor activity	9	75	1.69E-08
	GOTERM_MF_FAT	GO:0004888~transmembrane signaling receptor activity	9	75	3.48E-07
	GOTERM_MF_FAT	GO:0099600~transmembrane receptor activity	9	75	4.76E-07
	GOTERM_CC_FAT	GO:0005887~integral component of plasma membrane	8	66.67	5.03E-05
	GOTERM_CC_FAT	GO:0031226~intrinsic component of plasma membrane	8	66.67	6.55E-05
	GOTERM_CC_FAT	GO:0043005~neuron projection	4	33.34	0.038
Module 3	GOTERM_BP_FAT	GO:0006323~DNA packaging	13	28.26	1.31E-13
	GOTERM_BP_FAT	GO:0006334~nucleosome assembly	11	23.91	5.31E-12
	GOTERM_BP_FAT	GO:0071103~DNA conformation change	13	28.26	6.24E-12
	GOTERM_MF_FAT	GO:0046982~protein heterodimerization activity	12	26.09	5.17E-08
	GOTERM_MF_FAT	GO:0046983~protein dimerization activity	15	32.61	2.25E-06
	GOTERM_MF_FAT	GO:0042393~histone binding	6	13.04	1.75E-04
	GOTERM_CC_FAT	GO:0098687~chromosomal region	13	28.26	1.53E-10
	GOTERM_CC_FAT	GO:0005694~chromosome	18	39.13	2.91E-10
	GOTERM_CC_FAT	GO:0000786~nucleosome	9	19.57	6.45E-10

GO, Gene Ontology; PPI, protein-protein interaction; BP, biological processes; CC, cellular component; MF, molecular function.

previous study showed a strong connection in the expression of AURKA and RACGAP1 (31). However, there is a lack of data regarding whether AURKB has a strong connection with the expression of RACGAP1. MCM2 and MCM5 regulate cell proliferation. A study in 2017 showed that MCM2 can be used as an alternative to Ki-67 for measuring breast cancer cell proliferation (32). MCM5 also has a regulatory effect in cell proliferation (33). There is little research regarding MCM2 and MCM5 in GBM. As for TGFBR1, a study in 2016 showed that mi-181c inhibits cell invasion, migration and mesenchymal transition by downregulating TGFBR1 (34). SAMD3

is also a miR-regulated target in GBM (35). As shown in the present study, these six genes are all involved in the cell cycle. TGFBR1, SMAD3, VEGFC and FLT4 are involved in the regulation of epithelial cell proliferation. VEGFC has been shown to have an important role in GBM angiogenesis (36). FLT4 has been found to be differentially expressed in angiosarcoma of the breast and lymphatic filariasis (37,38). However, to the best of our knowledge, there are no studies concerning FLT4 in GBM. Previous studies have shown that little is known about these two genes (VEGFC, FLT4). RPN1 has been demonstrated to function in proteasome-mediated proteolysis (39).

Table III. KEGG pathway analysis of top 3 primary modules of PPI sub-network genes.

Sub-network	Pathway ID	Name	Gene count	%	P-value	Genes
Module 1	hsa04740	Olfactory transduction	17	40.48	2.86E-11	OR5H1, OR2A4, OR5P2, OR1A2, OR5K1, OR1J4, OR7E24, OR2M4, OR2L2, OR1G1, OR6B1, OR2F1, OR3A1, OR8D1, OR51B4, OR2C3, OR7A10
	hsa04062	Chemokine signaling pathway	7	16.67	3.49E-04	GNGT1, CCR8, ADCY1, PPBP, GNAI1, CXCR4, CXCL11
	hsa05032	Morphine addiction	4	9.52	0.01	OPRM1, GNGT1, ADCY1, GNAI1
Module 2	hsa04080	Neuroactive ligand-receptor interaction	8	66.67	1.55E-10	P2RY10, CHRM3, TACR1, ADRA1B, ADRA1A, GNRHR, GHSR, NTSR2
	hsa04020	Calcium signaling pathway	4	33.34	5.54E-04	CHRM3, TACR1, ADRA1B, ADRA1A
Module 3	hsa04970	Salivary secretion	3	25	0.003	CHRM3, ADRA1B, ADRA1A
	hsa05203	Viral carcinogenesis	12	26.09	2.68E-09	HIST1H2BA, MAD1L1, HIST1H4L, HDAC3, KAT2B, HIST1H4A, HIST1H4B, HIST1H2BH, ACTN1, HIST1H4F, ACTN2, HIST1H4C
	hsa05322	Systemic lupus erythematosus	10	21.74	1.36E-08	HIST1H2BA, HIST1H4L, HIST1H4A, HIST1H4B, HIST1H2BH, H2AFY2, ACTN1, HIST1H4F, ACTN2, HIST1H4C
	hsa05034	Alcoholism	9	15.57	2.06E-06	HIST1H2BA, HIST1H4L, HDAC3, HIST1H4A, HIST1H4B, HIST1H2BH, H2AFY2, HIST1H4F, HIST1H4C

KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction.

RPN2 has been shown as a promising biomarker in colorectal cancer (40). Both of these genes have not been studied clearly in GBM; therefore, they are two potential biomarkers which merit further studies.

KEGG analysis of the whole PPI network in the present study shows that the development of GBM is associated with the TGF- β signaling pathway, MAPK signaling pathway, pathways in cancer, basal cell carcinoma, FoxO signaling pathway, neuroactive ligand-receptor interaction, Jak-STAT signaling pathway, circadian entrainment and cytokine-cytokine receptor interaction. In addition, certain genes were also involved in the Wnt pathway, cell cycle and cell proliferation.

As shown in the survival curve, RANBP17, ZNF734, NLRP2, GPR1, CCDC81, SH3RF1 and TM7SF4 are good prognostic biomarkers for GBM. It is noteworthy that the expression of a number of the genes not only affects survival time, but also shows sex differences in the GBM samples. According to a recent study (41), the female sex plays an

important role in the survival of one subtype of GBM. This may mean that sex plays an important role in cancer. The prognosis of patients of different sexes varies greatly. To date, the international diagnostic and treatment criteria for glioma does not include sex differences. The present study reveals that different expression levels of key genes and sex may provide more information with regards to treating patients with GBM.

In conclusion, the large-scale data analysis of the present study provides a comprehensive bioinformatics analysis of DEGs that may be involved in the progress of GBM. Joint analysis of two databases helped to overcome the shortcomings of insufficient samples, and existing data was reused to conduct a novel analysis. This study provides a set of important targets for future investigation into the molecular mechanisms and prognostic biomarkers involved in GBM. However, follow-up molecular biological experiments and verification with many more clinical samples are required to confirm the functions of the identified genes in patients with different subtypes of GBM.

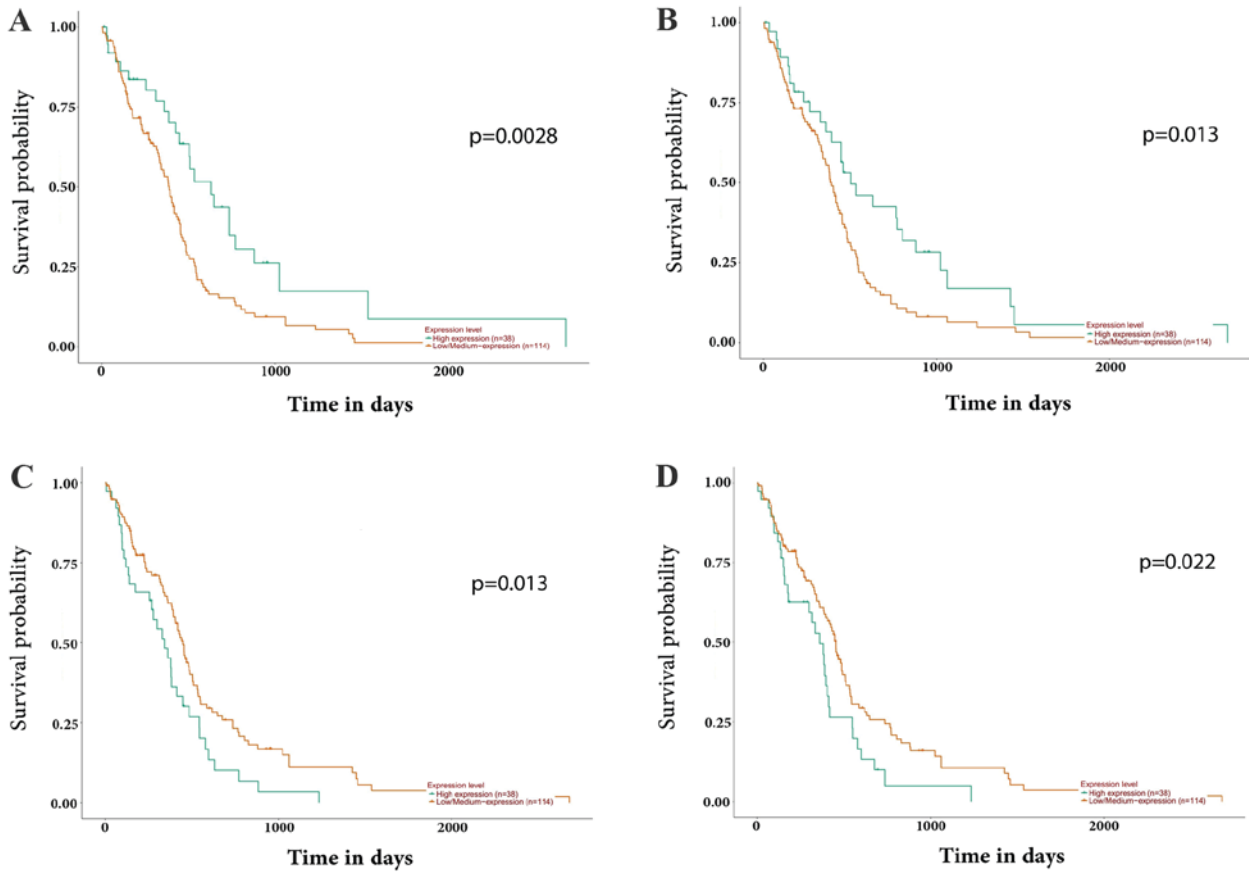


Figure 6. Top 4 genes related to survival of GBM patients. (A) RANBP17; (B) ZNF734; (C) NLRP2; (D) GPR1. GBM, glioblastoma.

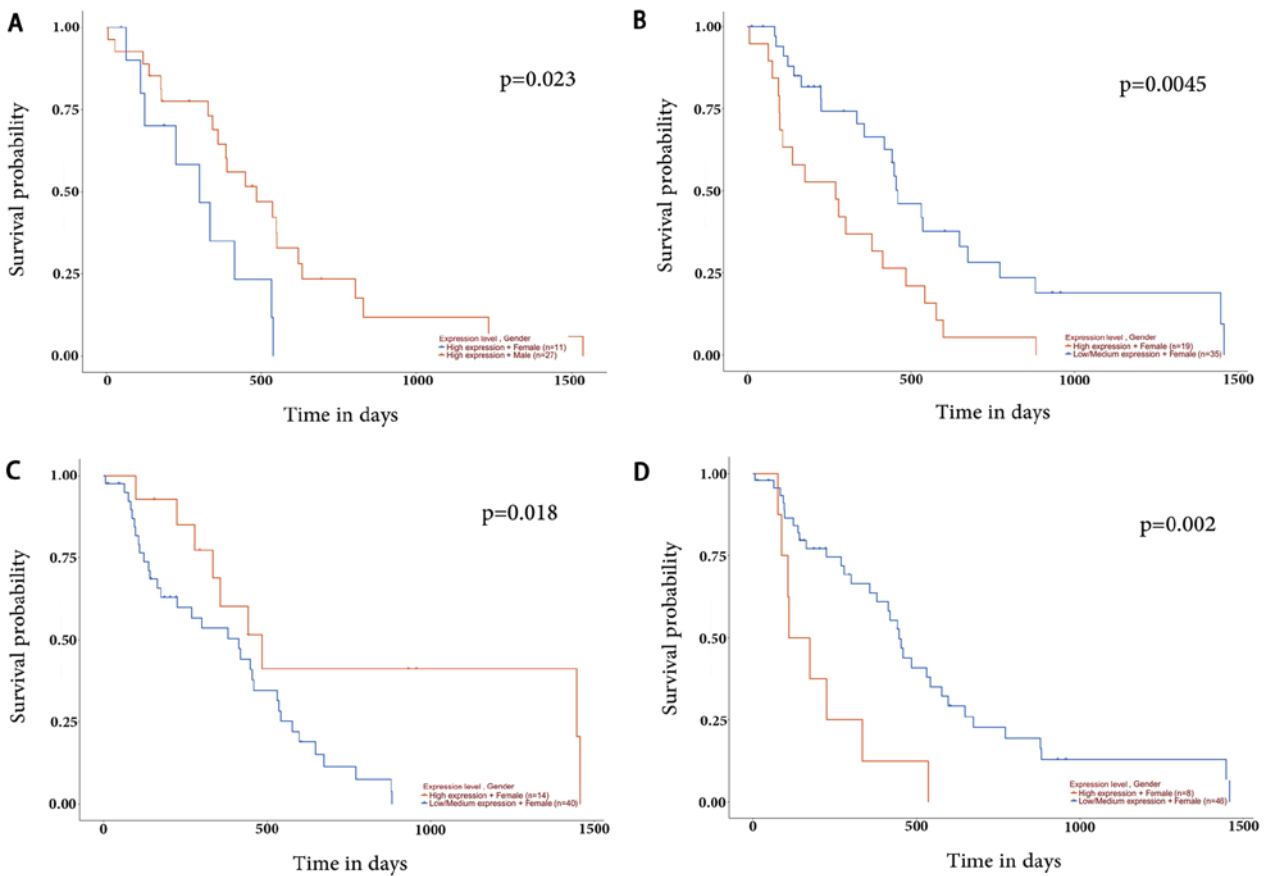


Figure 7. Top 4 prognostic genes affected by gender. (A) CCDC81; (B) NLRP2; (C) SH3RF1; (D) TM7SF4.

Acknowledgements

The present study was supported by International Program for Ph.D. Candidates, Sun Yat-sen University.

Funding

The present study was financially supported by the National Natural Science Foundation of China (grant no. 31360258), the Fundamental Research Funds for the Central Universities (grant no. 17ykpy84), Guangdong Provincial Natural Science Foundation (grant nos. 2015A030313047, 2015A030313077 and 2017A030310192), and Science and Technology Planning Project of Guangdong Province of China (grant no. 2016A020215036).

Availability of data and materials

The datasets analyzed in the present study are available from the GEO repository, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15824> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51062>.

Authors' contributions

LZ, HT, JX and KG conceived and designed the study; LZ analyzed and interpreted the microarray datasets, and produced the manuscript. LZ, HT, FW, LC, SO and TW analyzed the data. LZ wrote the paper, HT submitted the manuscript, and LZ and HT revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Steiner HH, Herold-Mende C, Bonsanto M, Geletneký K and Kunze S: Prognosis of brain tumors: Epidemiology, survival time and clinical course. *Versicherungsmedizin* 50: 173-179, 1998 (In French).
- Fisher JL, Schwartzbaum JA, Wrensch M and Wiemels JL: Epidemiology of brain tumors. *Neurol Clin* 25: 867-890, 2007.
- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW and Kleihues P: The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114: 97-109, 2007.
- Asthağiri AR, Pouratian N, Sherman J, Ahmed G and Shaffrey ME: Advances in brain tumor surgery. *Neurol Clin* 25: 975-1003, 2007.
- Grzmil M, Morin P Jr, Lino MM, Merlo A, Frank S, Wang Y, Moncayo G and Hemmings BA: MAP kinase-interacting kinase 1 regulates SMAD2-dependent TGF- β signaling pathway in human glioblastoma. *Cancer Res* 71: 2392-2402, 2011.
- Walsh AM, Kapoor GS, Buonato JM, Mathew LK, Bi Y, Davuluri RV, Martinez-Lage M, Simon MC, O'Rourke DM and Lazzara MJ: Sprouty2 drives drug resistance and proliferation in glioblastoma. *Mol Cancer Res* 13: 1227-1237, 2015.
- Gerber NK, Goenka A, Turcan S, Reingold M, Makarov V, Kannan K, Beal K, Omuro A, Yamada Y, Gutin P, *et al*: Transcriptional diversity of long-term glioblastoma survivors. *Neuro Oncol* 16: 1186-1195, 2014.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomaszewsky M, Marshall KA, Phillippy KH, Sherman PM, *et al*: NCBI GEO: Archive for functional genomics data sets-10 years on. *Nucleic Acids Res* 39: D1005-D1010, 2011.
- NIH: The Cancer Genome Atlas. TCGA Data Portal. <https://tcga-data.nci.nih.gov/docs/publications/tcga/>?. Last updated June 30, 2016
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P and Ellison DW: The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol* 131: 803-820, 2016.
- Diboun I, Wernisch L, Orengo CA and Koltzenburg M: Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics* 7: 252, 2006.
- Gautier L, Cope L, Bolstad BM and Irizarry RA: affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307-315, 2004.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP: Exploration, normalization and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264, 2003.
- Gene Ontology Consortium: The gene ontology (GO) project in 2006. *Nucleic Acids Res* 34: D322-D336, 2006.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics* 25: 25-29, 2000.
- Kanehisa M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30, 2000.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA: DAVID: Database for annotation, visualization and integrated discovery. *Genome Biol* 4: P3, 2003.
- Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen LJ: STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41: D808-D815, 2013.
- Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27: 431-432, 2011.
- Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Rodriguez I, Chakravarthi BVSK and Varambally S: UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19: 649-658, 2017.
- Liu TT, Achrol AS, Mitchell LA, Rodriguez SA, Feroze A, Iv M, Kim C, Chaudhary N, Gevaert O, Stuart JM, *et al*: Magnetic resonance perfusion image features uncover an angiogenic subgroup of glioblastoma patients with poor survival and better response to antiangiogenic treatment. *Neuro Oncol* 19: 997-1007, 2017.
- Brandes A, Franceschi E, Tosoni A, Fioravanti A, Agati R, Andreoli A, Mazzocchi V, Morandi L, Bartolini S and Ermani M: Change in MGMT methylation status between first and second surgery for recurrence: Clinical implications. *Eur J Cancer Suppl* 7: 495, 2009.
- Akbari A, Farahnejad Z, Akhtari J, Abastabar M, Mobini GR and Mehbod AS: Staphylococcus aureus enterotoxin B down-regulates the expression of transforming growth factor-beta (TGF- β). Signaling Transducers in Human Glioblastoma. *Jundishapur J Microbiol* 9: e27297, 2016.
- Jiang C, Shen F, Du J, Hu Z, Li X, Su J, Wang X and Huang X: MicroRNA-564 is downregulated in glioblastoma and inhibited proliferation and invasion of glioblastoma cells by targeting TGF- β 1. *Oncotarget* 7: 56200-56208, 2016.
- Li J, Ren S, Liu Y, Lian Z, Dong B, Yao Y and Xu Y: Knockdown of NUPRI inhibits the proliferation of glioblastoma cells via ERK1/2, p38 MAPK and caspase-3. *J Neuro Oncol* 132: 15-26, 2017.
- Kim EH, Song HS, Yoo SH and Yoon M: Tumor treating fields inhibit glioblastoma cell migration, invasion and angiogenesis. *Oncotarget* 7: 65125-65136, 2016.
- Firat E and Niedermann G: FoxO proteins or loss of functional p53 maintain stemness of glioblastoma stem cells and survival after ionizing radiation plus PI3K/mTOR inhibition. *Oncotarget* 7: 54883-54896, 2016.

28. 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 Metab* 18: 726-739, 2013.
29. Al-Khafaji AS, Davies MP, Risk JM, Marcus MW, Koffa M, Gosney JR, Shaw RJ, Field JK and Liloglou T: Aurora B expression modulates paclitaxel response in non-small cell lung cancer. *Br J Cancer* 116: 592-599, 2017.
30. Borges KS, Castro-Gamero AM, Moreno DA, da Silva Silveira V, Brassesco MS, de Paula Queiroz RG, de Oliveira HF, Carlotti CG Jr, Scrideli CA and Tone LG: Inhibition of Aurora kinases enhances chemosensitivity to temozolomide and causes radiosensitization in glioblastoma cells. *J Cancer Res Clin Oncol* 138: 405-414, 2012.
31. Bornschein J, Nieltz J, Drozdov I, Selgrad M, Wex T, Jechorek D, Link A, Vieth M and Malfertheiner P: Expression of aurora kinase A correlates with the Wnt-modulator RACGAP1 in gastric cancer. *Cancer Med* 5: 516-526, 2016.
32. Yousef EM, Furrer D, Laperriere DL, Tahir MR, Mader S, Diorio C and Gaboury LA: MCM2: An alternative to Ki-67 for measuring breast cancer cell proliferation. *Mod Pathol* 30: 682-697, 2017.
33. Su Z, Zheng X, Zhang X, Wang Y, Zhu S, Lu F, Qu J and Hou L: Sox10 regulates skin melanocyte proliferation by activating the DNA replication licensing factor MCM5. *J Dermatol Sci* 85: 216-225, 2017.
34. He X, Liu Z, Peng Y and Yu C: MicroRNA-181c inhibits glioblastoma cell invasion, migration and mesenchymal transition by targeting TGF- β pathway. *Biochem Biophys Res Commun* 469: 1041-1085, 2016.
35. Wu ZB, Cai L, Lin SJ, Lu JL, Yao Y and Zhou LF: The miR-92b functions as a potential oncogene by targeting on Smad3 in glioblastomas. *Brain Res* 1529: 16-25, 2013.
36. Carpenter RL, Paw I, Zhu H, Sirkisoon S, Xing F, Watabe K, Debinski W and Lo HW: The gain-of-function GLI1 transcription factor TGLI1 enhances expression of VEGF-C and TEM7 to promote glioblastoma angiogenesis. *Oncotarget* 6: 22653-22655, 2015.
37. Sheik Y, Qureshi SF, Mohammed B and Nallari P: FOXC2 and FLT4 gene variants in lymphatic filariasis. *Lymphat Res Biol* 13: 112-119, 2015.
38. Cornejo KM, Deng A, Wu H, Cosar EF, Khan A, St Cyr M, Tomaszewicz K, Dresser K, O'Donnell P and Hutchinson L: The utility of MYC and FLT4 in the diagnosis and treatment of postradiation atypical vascular lesion and angiosarcoma of the breast. *Hum Pathol* 46: 868-875, 2015.
39. Rani N, Aicheh A, Schmidtke G, Kreft SG and Groettrup M: FAT10 and NUBIL bind to the VWA domain of Rpn10 and Rpn1 to enable proteasome-mediated proteolysis. *Nat Commun* 3: 749, 2012.
40. Zhang J, Yan B, Späth SS, Qun H, Cornelius S, Guan D, Shao J, Hagiwara K, Van Waes C, Chen Z, *et al*: Integrated transcriptional profiling and genomic analyses reveal RPN2 and HMGB1 as promising biomarkers in colorectal cancer. *Cell Biosci* 5: 53, 2015.
41. Liao CT, Chou WC, Wei KC, Chang CN, Toh CH and Jung SM: Female sex, good performance status and bevacizumab-induced hypertension associated with survival benefit in Asian patients with recurrent glioblastoma treated with bevacizumab. *Asia Pac J Clin Oncol* 14: e8-e14, 2018.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.