ORIGINAL ARTICLE

WILEY

Gene regulation network analysis reveals core genes associated with survival in glioblastoma multiforme

Lan Jiang[1,2] (iD) | Min Zhong[1,2] | Tianbing Chen[1,2] | Xiaolong Zhu[1,2] | Hui Yang[1,2] | Kun Lv[1,2]

[1]Central Laboratory, Yijishan Hospital of Wannan Medical College, Wuhu, China

[2]Key Laboratory of Non-coding RNA Transformation Research of Anhui Higher Education Institution, Wannan Medical College, Wuhu, China

Correspondence
Kun Lv, Central Laboratory, Yijishan Hospital of Wannan Medical College, Wuhu 241001, China.
Email: lvkun@yjsyy.com

Abstract

Glioblastoma multiforme (GBM) is a very serious mortality of central nervous system cancer. The microarray data from GSE2223, GSE4058, GSE4290, GSE13276, GSE68848 and GSE70231 (389 GBM tumour and 67 normal tissues) and the RNA-seq data from TCGA-GBM dataset (169 GBM and five normal samples) were chosen to find differentially expressed genes (DEGs). RRA (Robust rank aggregation) method was used to integrate seven datasets and calculate 133 DEGs (82 up-regulated and 51 down-regulated genes). Subsequently, through the PPI (protein-protein interaction) network and MCODE/ cytoHubba methods, we finally filtered out ten hub genes, including FOXM1, CDK4, TOP2A, RRM2, MYBL2, MCM2, CDC20, CCNB2, MYC and EZH2, from the whole network. Functional enrichment analyses of DEGs were conducted to show that these hub genes were enriched in various cancer-related functions and pathways significantly. We also selected CCNB2, CDC20 and MYBL2 as core biomarkers, and further validated them in CGGA, HPA and CCLE database, suggesting that these three core hub genes may be involved in the origin of GBM. All these potential biomarkers for GBM might be helpful for illustrating the important role of molecular mechanisms of tumorigenesis in the diagnosis, prognosis and targeted therapy of GBM cancer.

KEYWORDS
CCNB2, CDC20, glioblastoma, MYBL2, RRA

1 | INTRODUCTION

Glioblastoma multiforme (GBM) is an incurable malignancy.[1] Almost all GBMs recur within the first year following diagnosis, the recurrence rate of GBM is particularly high, they can be surgically resected again.[2] Heterogeneity is the key point for the treatment of glioma.[3] It is still very difficult to deal with the recurrence of GBM during the treatment. Based on the histology, gliomas would be classified into WHO (World Health Organization) grades I, II, III and IV.[4] The natural course of low-grade glioma (WHO Grade II) is to transform or to dedifferentiate into high-grade glioma (WHO grade III–IV), and they recur after surgical resection frequently. The limited information on the pathogenesis, development, reproduction and molecular mechanisms of GBM has hindered the research and development of precise treatment of available drugs. Therefore, it is urgent to clarify the relevant

molecular mechanisms of GBM and actively develop new therapeutic strategies.

Many studies have illustrated the numerous candidate hub genes involved in GBM from RNA-seq data. Biomarkers could combine with different diseases and display a high value which lead to the development of a robust, effective take on GBM therapy.[5] Numerous recent studies have discovered the biomarkers in GBM from molecular biology to proteomics, such as circulating tumour DNA (ctDNA), DNA, microRNA (miRNA), lncRNA (long non-coding RNA) and protein.[6-8] ctDNAs in cerebrospinal fluid better reflected the sequential change of those tumour drivers than in plasma, which served as biomarkers to improve patients' outcomes.[9] GFAP (Glial Fibrillary Acidic Protein) and EGFR (Epidermal Growth Factor Receptor) were considered to be increased and as potential therapeutic markers in GBM patients.[10,11] CDK1 (Cyclin Dependent Kinase 1) and BUB1 (BUB1 Mitotic Checkpoint Serine/Threonine Kinase) were significantly connected with carcinogenesis of GBM.[12,13] Hsa-miR-21 and hsa-miR-10b were discovered as GBM-specific miRNAs, which lead to the development of a robust take on GBM therapy.[5] The blood biomarkers (LRG1 (Leucine Rich Alpha-2-Glycoprotein 1), CRP (C-Reactive Protein) and C9 (Complement C9)) revealed significant positive correlations with tumour size.[14] Our previous study also reported that HMG-box family and related ceRNA (competing endogenous RNA) established the significance of SOX6 (SRY-Box Transcription Factor 6) in the malignant progression of glioblastoma.[6]

Taking into account the individual differences, in this study, we selected the microarray data from the Gene Expression Omnibus (GEO) database and the RNA-seq data from TCGA-GBM (The Cancer Genome Atlas Glioblastoma Multiforme) dataset, using the RRA (RobustRankAggreg) method to identify differentially expressed genes (DEGs) between GBM tissues and normal tissues. Through network analysis and functional analyses, it is possible to predict the pathways and interactions of DEGs. In addition, hub genes related lncRNA, miRNA and transcription factor (TF) were also explored. All of these bioinformatics methods were used to elucidate the comprehensive molecular mechanisms responsible for the development and progression of GBM and to provide potential biomarkers as a treatment for patients in different subgroups of GBM.

2 | METHODS AND MATERIALS

2.1 | Data preparation and RRA method analyses

The microarray data of human tissue from GSE2223,[15] GSE4058,[16] GSE4290,[17] GSE13276,[18] GSE68848[19] and GSE70231[20] were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). (a) The inclusion and exclusion criteria were applied for the selection of GEO datasets (inclusion: datasets containing tumour and normal samples; exclusion: experiments on cell lines, datasets containing serum samples nor tissue biopsies, etc). (b) They were GBM and normal tumours cases based on 456 samples connected with subtypes (389 tumours and

67 normal samples). GSE2223 (29 GBM samples: four normal samples, GPL1833); GSE4058 (30 GBM samples: 3 normal samples, GPL182); GSE4290 (76 GBM samples: 23 normal samples, GPL570); GSE13276 (5 GBM samples: 3 normal samples, GPL96); GSE68848 (228 GBM samples: 28 normal samples, GPL570); GSE70231 (21 GBM samples: 6 normal samples, GPL80). (c) Limma package,[21] Deseq2[22] and edgeR method[23] (fold change > 2 and adjusted P-value (q-value) < .05) were used for identifying GEO data DEGs (differential expressed genes), absolute value of fold change > 2 and adjusted P-value (q-value) < .05 were considered as DEGs. The DEGs result of each dataset was drawn the violin plot by using ggplot2 package,[24] respectively.

Then, we downloaded the RNA-seq data of human tissue from TCGA-GBM dataset (169 GBM and five normal samples)[25] and identified DEGs by intersecting with the results with limma, Deseq2[22] and edgeR method[23] (fold change > 2 and q-value < .05).

Finally, RobustRankAggreg package (RRA method) was designed to sort the multi-gene lists and adopted to gain the robust DEGs.[26] The pheatmap package is used to visualize the top 20 up- and top 20 down- DEGs obtained by the six GEO datasets and TCGA-GBM dataset by RRA method.[27,28]

2.2 | Function and network analyses

Protein-protein interaction (PPI) network was obtained from STRING v11.[29] We used the Cytoscape v 3.7.2 plugin MCODE and cytoHubba to select the hub genes.[30] We used Cytoscape to visualize the PPI networks, and MCODE plugin to screen a significant module from the PPI network with a degree cutoff = 2, node score cutoff = 0.2, node density cutoff = 0.1, Max depth = 100 and K-core = 2. Then, cytoHubba plugin was used to determine the hub genes when the degrees were ≥ 10.

The analyses of GO (Gene Ontology) enrichment, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway were performed via clusterProfiler package.[31] Gene set enrichment analysis (GSEA) was conducted by clusterProfiler package as well and drawn as emapplot and heatplot to better understand the results for GSEA. Survival analysis was drawn based on GlioVis database (http://www.gliovis.bioinfo.cnio.es). Tumour-infiltrating immune cells were inferred using TIMER (Tumor Immune Estimation Resource).[32] Functional associations of the hub genes (with TF, miRNAs and lncRNAs) were analysed using NetworkAnalyst.[33]

2.3 | Validation of core hub genes

We downloaded the RNA-seq data of human tissues from CGGA, after batching all mRNA-seq matrix and removing the incomplete and duplicate data, we analysed these data by R packages: (a) Prognostic accuracy of the three core hub genes evaluated by ROC (receiver operating characteristics) curve with respect to 1 year, 3 year and 5 year survival of glioma patients by 'survivalROC'

package [34]; (b) Drawing forest plots of univariate and multivariate cox analysis by 'survival' package [35]; (c) The expression of core hub genes in different types by 'beeswarm' package.[36] Next, we validated core hub gene in HPA (The Human Protein Atlas) (GBM U251-MG cell line) (https://www.proteinatlas.org) and CCLE (Cancer Cell Line Encyclopedia) (glioma cell) database (https://portals.broadinstitute.org/ccle).

3 | RESULTS

3.1 | The overall collection of datasets and identification of DEGs

Six gene datasets were collected from NCBI (National Center for Biotechnology Information) GEO database (GSE2223, GSE4058, GSE4290, GSE13276, GSE68848 and GSE70231) and TCGA-GBM dataset (Table 1). A total of 389 GBM tumour and 67 normal tissues were adopted in this study. With t

he criteria of log2 (fold change)> 1 and adjusted P-value < 0.05, 754 up-regulated and 750 down-regulated DEGs were conducted in GSE2223 (Figure 1A); 176 up-regulated and 174 down-regulated DEGs were gained in GSE4058 (Figure 1B); 566 up-regulated and 487 down-regulated DEGs were obtained in GSE4290 (Figure 1C); 177 up-regulated and 51 down-regulated DEGs were gained in GSE13276 (Figure 1D); and 619 up-regulated and 492 down-regulated DEGs were adopted in GSE68848 (Figure 1E); and 377 up-regulated and 244 down-regulated DEGs were obtained in GSE70231 (Figure 1F). In TCGA-GBM dataset, we used three methods to find DEGs, there were 2335 DEGs in edgeR, 1303 DEGs in limma, and 1555 in Deseq2, then we took the intersection of the three methods, and 1255 DEGs were obtained from TCGA-GBM in final.

The RRA method was used to identify genes that are ranked consistently better and to explore the robust DEGs in different seven datasets. Finally, we determined 133 significantly DEGs in these datasets, including 82 up-regulated and 51 down-regulated DEGs (Table S1). The expression heat map of the top 20 up-regulated and top 20 down-regulated DEGs is visualized in Figure 2.

TABLE 1 The selected samples for this current study

GEO accessions	GBM number	Normal number	Reference
GSE2223	29	4	15
GSE4058	29	3	16
GSE4290	77	23	17
GSE13276	5	3	18
GSE68848	228	28	19
GSE70231	21	6	20
TCGA-GBM	169	5	25

3.2 | PPI network and hub gene detection

These 133 DEGs were input to the STRING database for PPI networks, for their potential biological functions. The clusters of sub-networks that were obtained from STRING database with 133 nodes and 547 edges (PPI enrichment P-value < 1.0e-16) (Figure 3).

The top one biological processes (BP), molecular functions (MF) and cellular components (CC) were listed in Table S2 and marked as red, purple and green, respectively. The highest enriched BP, MF, CC terms were 'developmental process', 'protein binding', 'secretory vesicle'. The top one KEGG pathway was 'AGE-RAGE signalling pathway in diabetic complications' marked as yellow. In the glioma pathway (hsa05214), included in CDK4 (Cyclin Dependent Kinase 4), PRKCG (Protein Kinase C Gamma) and EGFR. The highest enriched Reactome pathway was 'Extracellular matrix organization' marked as blue (Figure 3 and Table S2).

Next, Cytoscape plugin MCODE were used to find out the hub genes. We found five types from MCODE results, the cluster one with score = 9.7, obtained 21 DEGs, and the other four clusters with score = 6.588, 4, 3 and 3, respectively (Figure 4A-E). Through MCODE analysis results, we used cytoHubba for further analysis in the results of cluster one (21 DEGs), we determined the top 10 DEGs as hub genes (FOXM1 (Forkhead Box M1), CDK4, TOP2A (DNA Topoisomerase II Alpha), RRM2 (Ribonucleotide Reductase Regulatory Subunit M2), MYBL2 (MYB Proto-Oncogene Like 2), MCM2 (Minichromosome), CDC20 (Cell Division Cycle 20), CCNB2 (Cyclin B2), MYC (MYC Proto-Oncogene, BHLH Transcription Factor) and EZH2 (Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit)) (Figure 4F).
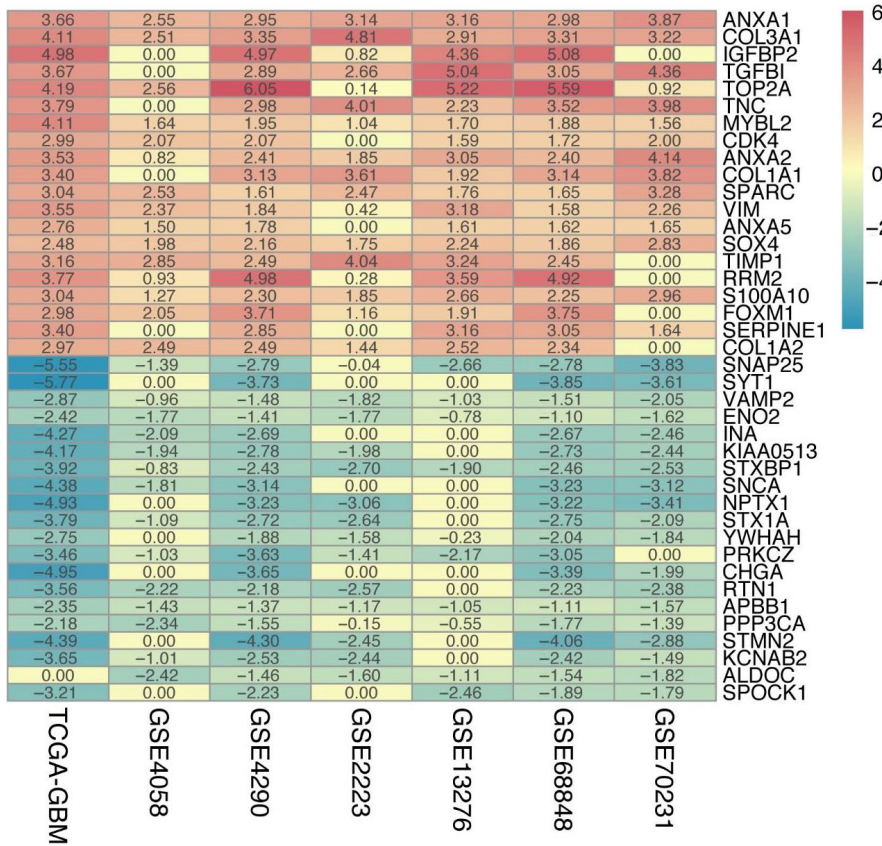
3.3 | Functional and survival analyses

Gene ontology (GO) functional enrichment analyses were used to determine the potential molecular mechanisms employed by these 10 hub genes. The top 10 biological processes (BP), molecular functions (MF) and cellular components (CC) are listed in Figure 5A-C. The highly enriched BP terms were 'G1/S transition of mitotic cell cycle', 'cell cycle G1/S phase transition' and 'negative regulation of mitotic cell cycle'. The markedly enriched MF terms were 'cyclin-dependent protein serine/threonine kinase regulator activity', 'core promoter binding' and 'DNA-dependent ATPase activity'. The predominantly enriched CC terms were 'cyclin-dependent protein kinase holoenzyme complex', 'serine/threonine protein kinase complex'. With adjusted P-value < 0.05, seven pathways were enriched by the 10 DEGs (Figure 5D), many of which were tumour-associated pathways, including the 'Bladder cancer', and the 'p53 signalling pathway' and the 'Small cell lung cancer'. Moreover, CDK4 was predicted to be centralized in the glioma pathway (hsa05214) (Figure 6).

GSEA heatplot of enriched DEGs list on each term was shown in Figure 7 and Figure S2. There are four types (primary malignant neoplasm of brain, malignant peripheral nerve sheath tumour, malignant neoplasm of brain, ewings sarcoma-primitive neuroectodermal
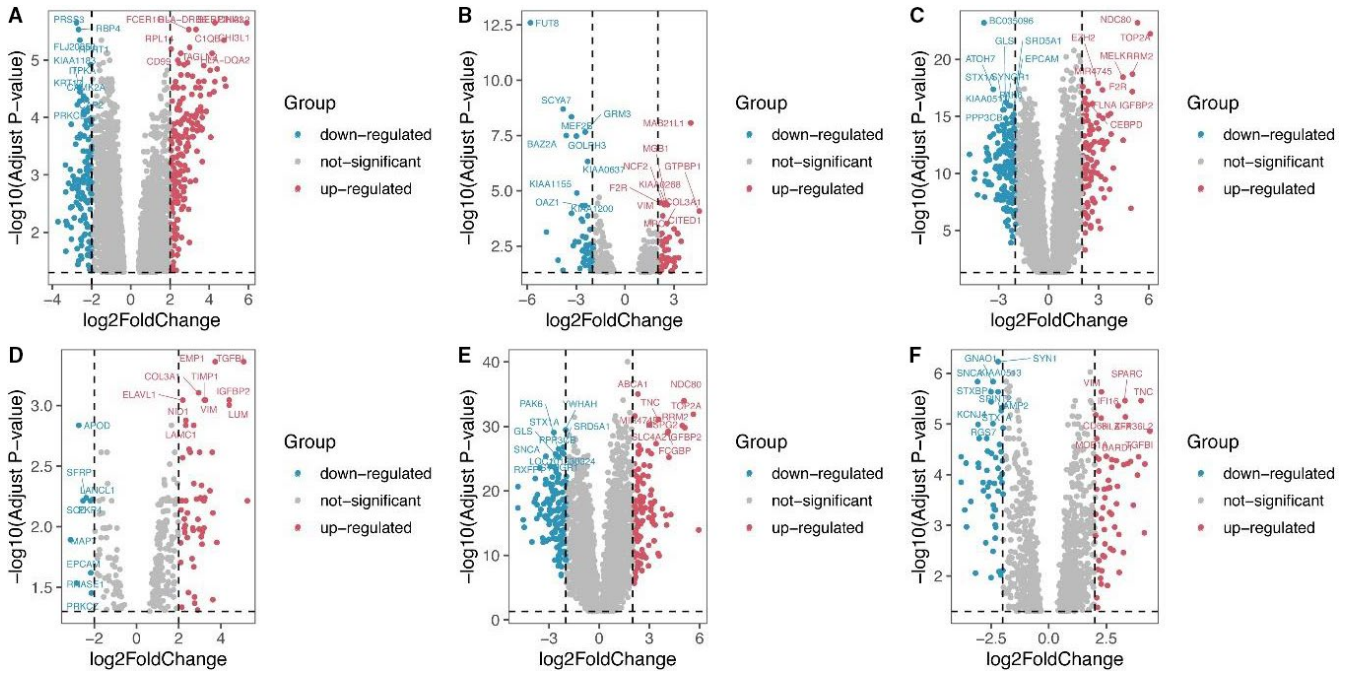
FIGURE 1 Volcano plots for DEGs in GBM and normal tissues based on data from the GEO datasets. A. GSE2223, B. GSE4058, C. GSE4290, D. GSE13276, E. GSE68848 and F. GSE70231

TCGA-GBM	GSE4058	GSE4290	GSE2223	GSE13276	GSE68848	GSE70231	
3.66	2.55	2.95	3.14	3.16	2.98	3.87	ANXA1
4.11	2.51	3.35	4.81	2.91	3.31	3.22	COL3A1
4.98	0.00	4.97	0.82	4.36	5.08	0.00	IGFBP2
3.67	0.00	2.89	2.66	5.04	3.05	4.36	TGFBI
4.19	2.56	6.05	0.14	5.22	5.59	0.92	TOP2A
3.79	0.00	2.98	4.01	2.23	3.52	3.98	TNC
4.11	1.64	1.95	1.04	1.70	1.88	1.56	MYBL2
2.99	2.07	2.07	0.00	1.59	1.72	2.00	CDK4
3.53	0.82	2.41	1.85	3.05	2.40	4.14	ANXA2
3.40	0.00	3.13	3.61	1.92	3.14	3.82	COL1A1
3.04	2.53	1.61	2.47	1.76	1.65	3.28	SPARC
3.55	2.37	1.84	0.42	3.18	1.58	2.26	VIM
2.76	1.50	1.78	0.00	1.61	1.62	1.65	ANXA5
2.48	1.98	2.16	1.75	2.24	1.86	2.83	SOX4
3.16	2.85	2.49	4.04	3.24	2.45	0.00	TIMP1
3.77	0.93	4.98	0.28	3.59	4.92	0.00	RRM2
3.04	1.27	2.30	1.85	2.66	2.25	2.96	S100A10
2.98	2.05	3.71	1.16	1.91	3.75	0.00	FOXM1
3.40	0.00	2.85	0.00	3.16	3.05	1.64	SERPINE1
2.97	2.49	2.49	1.44	2.52	2.34	0.00	COL1A2
−5.55	−1.39	−2.79	−0.04	−2.66	−2.78	−3.83	SNAP25
−5.77	0.00	−3.73	0.00	0.00	−3.85	−3.61	SYT1
−2.87	−0.96	−1.48	−1.82	−1.03	−1.51	−2.05	VAMP2
−2.42	−1.77	−1.41	−1.77	−0.78	−1.10	−1.62	ENO2
−4.27	−2.09	−2.69	0.00	0.00	−2.67	−2.46	INA
−4.17	−1.94	−2.78	−1.98	0.00	−2.73	−2.44	KIAA0513
−3.92	−0.83	−2.43	−2.70	−1.90	−2.46	−2.53	STXBP1
−4.38	−1.81	−3.14	0.00	0.00	−3.23	−3.12	SNCA
−4.93	0.00	−3.23	−3.06	0.00	−3.22	−3.41	NPTX1
−3.79	−1.09	−2.72	−2.64	0.00	−2.75	−2.09	STX1A
−2.75	0.00	−1.88	−1.58	−0.23	−2.04	−1.84	YWHAH
−3.46	−1.03	−3.63	−1.41	−2.17	−3.05	0.00	PRKCZ
−4.95	0.00	−3.65	0.00	0.00	−3.39	−1.99	CHGA
−3.56	−2.22	−2.18	−2.57	0.00	−2.23	−2.38	RTN1
−2.35	−1.43	−1.37	−1.17	−1.05	−1.11	−1.57	APBB1
−2.18	−2.34	−1.55	−0.15	−0.55	−1.77	−1.39	PPP3CA
−4.39	0.00	−4.30	−2.45	0.00	−4.06	−2.88	STMN2
−3.65	−1.01	−2.53	−2.44	0.00	−2.42	−1.49	KCNAB2
0.00	−2.42	−1.46	−1.60	−1.11	−1.54	−1.82	ALDOC
−3.21	0.00	−2.23	0.00	−2.46	−1.89	−1.79	SPOCK1

FIGURE 2 Heat map for the 20 top up-regulated and 20 top down-regulated DEGs in GBM using the RobustRankAggreg method with *q*-value < .05 and fold change > 2

tumour (PNET)) related to nervous system diseases. In primary malignant neoplasm of brain which including three DEGs (CDK4, EZH2, MYC). We could also find 4 DEGs (CDK4, EZH2, FOXM1 and TOP2A) in the type of malignant peripheral nerve sheath tumour. In addition, CDK4 and EZH2 were determined to be centralized in all four types (Figure 7).

As shown in Figure 8, based on GlioVis database (TCGA-GBM and CGGA-GBM), CDC20, CCNB2 and MYBL2 (*P*-value < 0.01) (Figure 8)

FIGURE 3 PPI network of the DEGs in GBM. The top one biological processes (BP), molecular functions (MF), cellular components (CC), KEGG pathway and Reactome pathway were marked as red, purple, green, yellow and blue, respectively
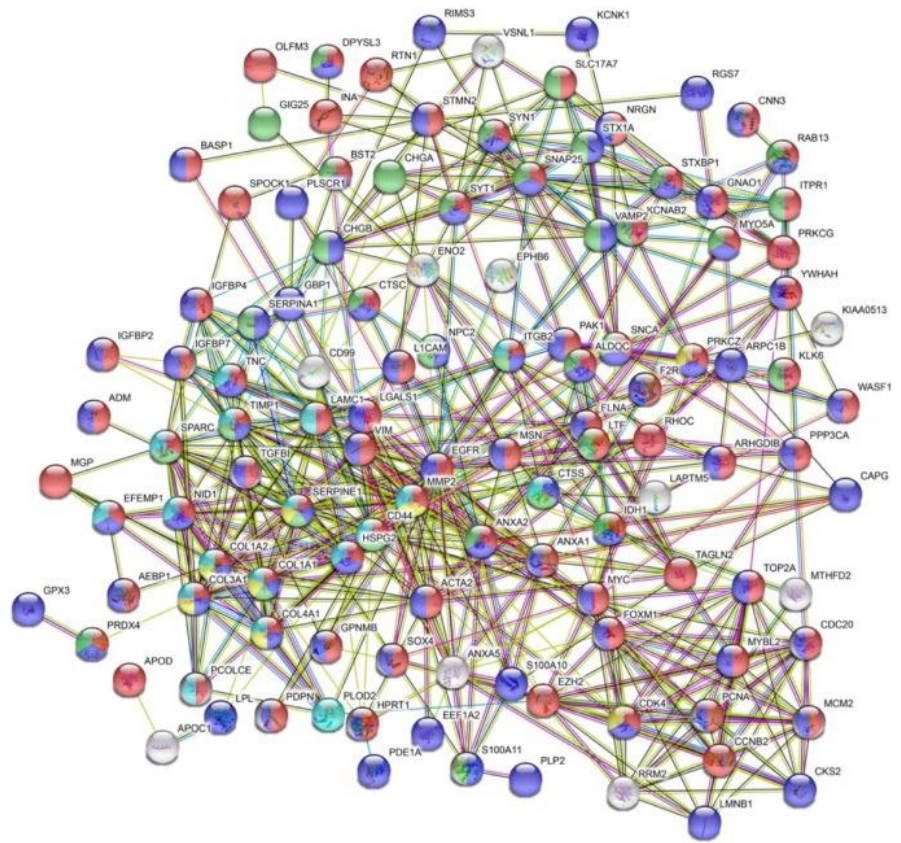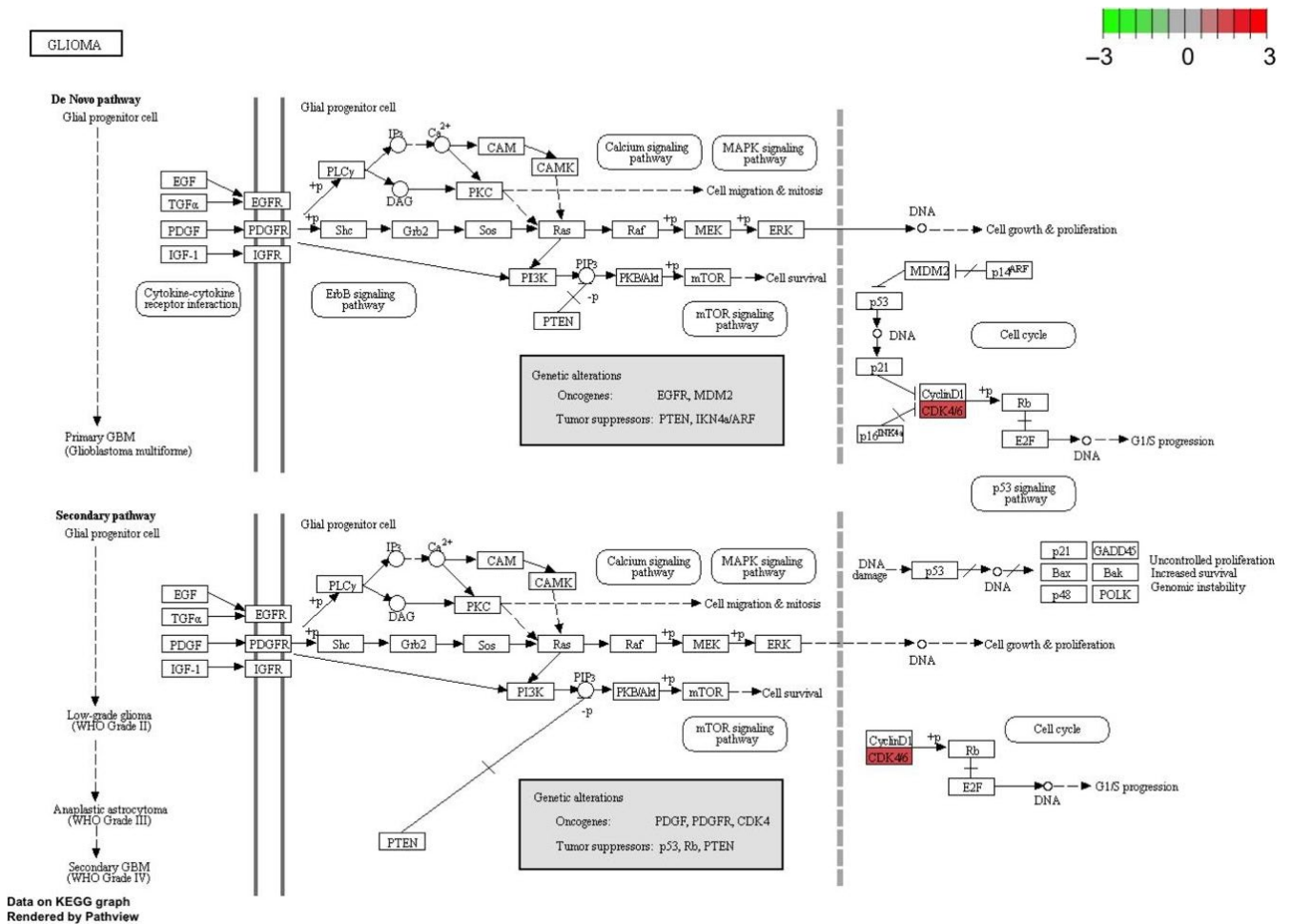

FIGURE 4 Hub genes screening from MCODE and cytoHubba plugin in cytoscape v3.7.2. A-E, MCODE results; F, cytoHubba result
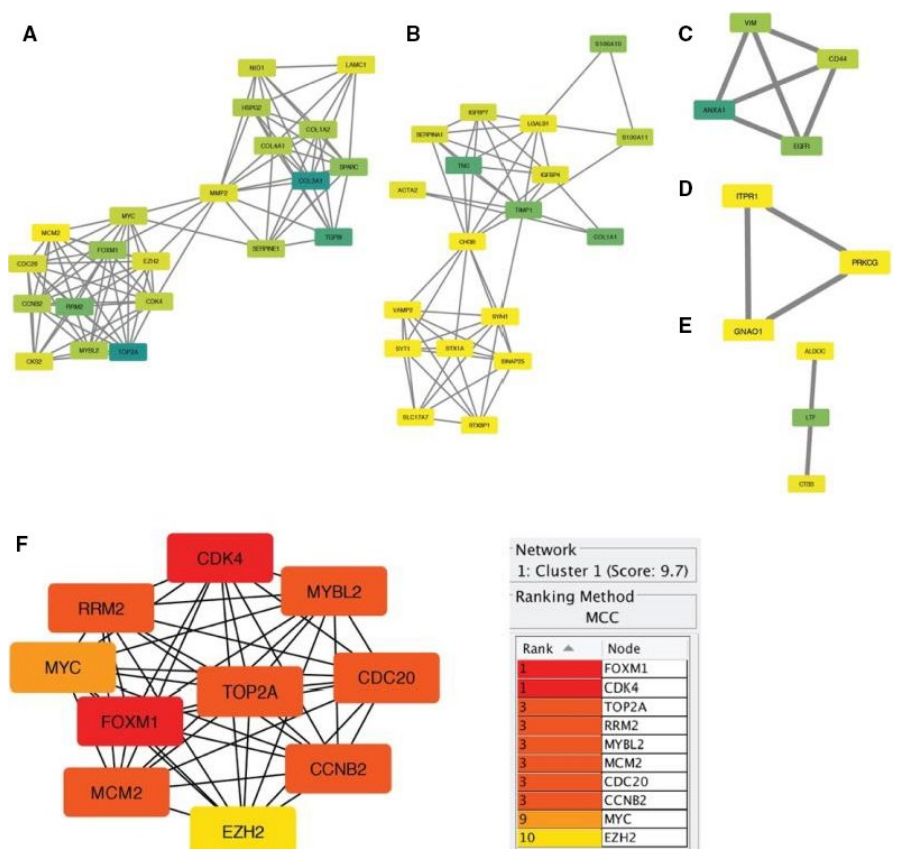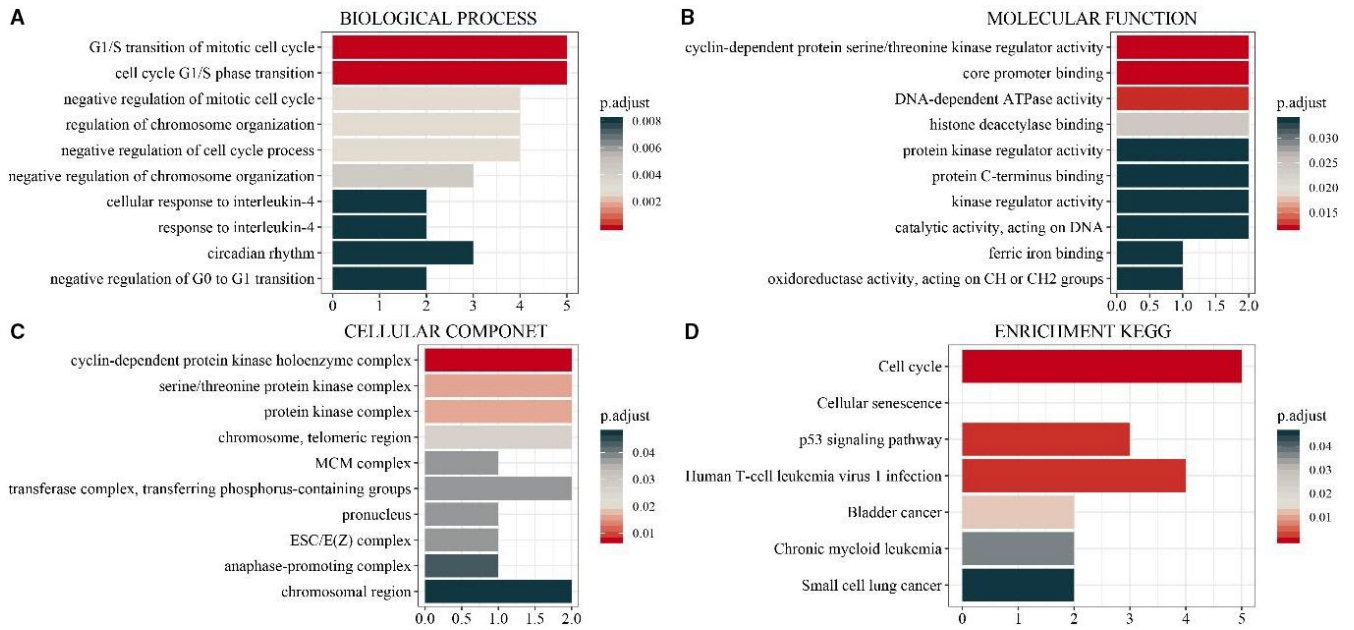
FIGURE 5 GO enrichment annotations and KEGG pathways of the DEGs in GBM. A, Top 10 biological process terms. B, Top 10 molecular function terms. C, Cellular component terms. D, significantly enriched KEGG pathways (adjusted *P*-value < .05)

FIGURE 6 Glioma pathway from KEGG analysis which was performed by pathview package. One hub gene CDK4 in glioma pathway marked as red colour
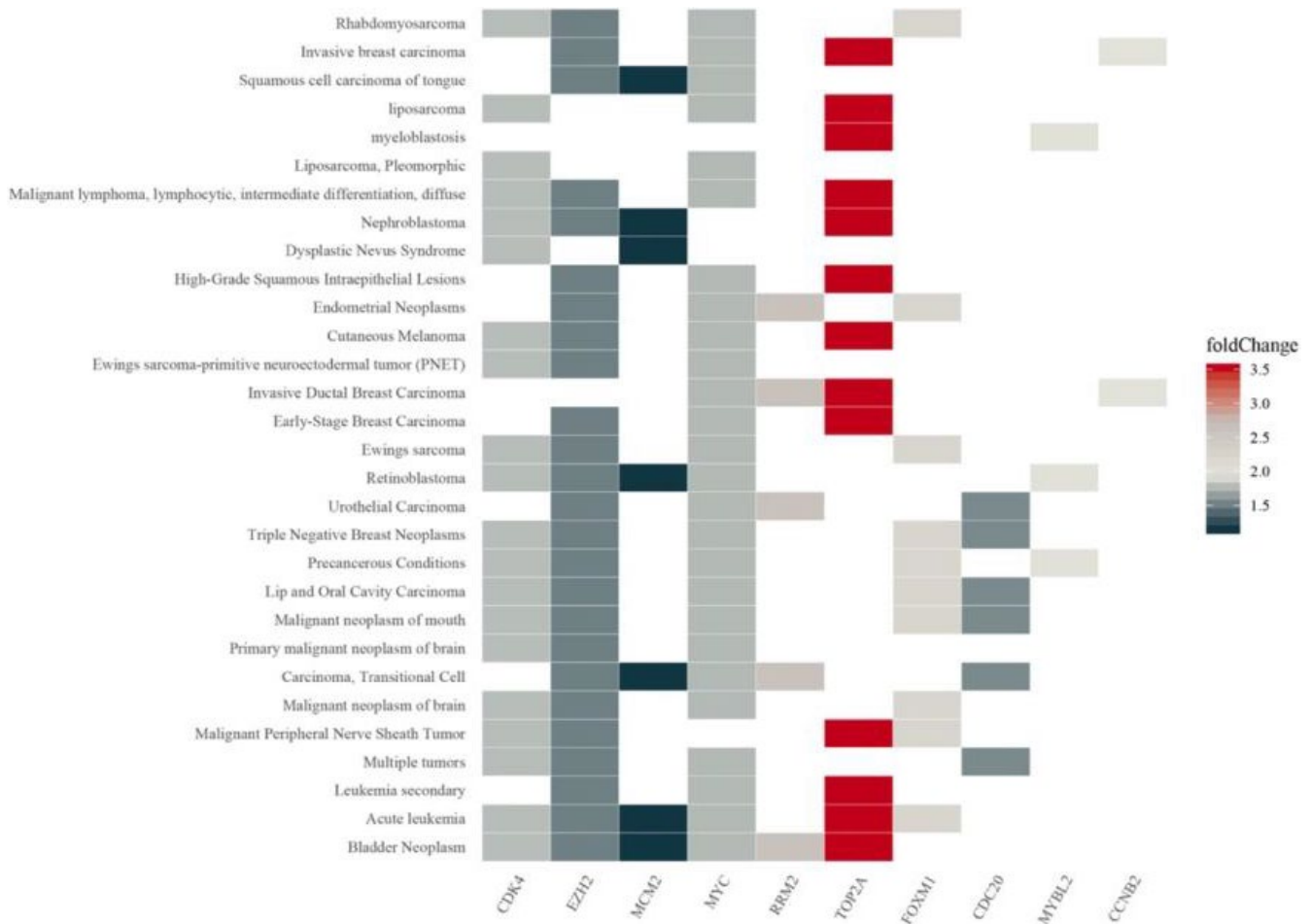
FIGURE 7 GSEA heatplot of the hub genes expression profiles using clusterProfiler

was correlated with the survival of determine optimal cutoff for Kaplan-Meier survival analysis (Figure 8). GlioVis used the maximally selected statistics to determine the optimal cutoff for continuous variables, which provided in the 'survminer' package. CDC20, CCNB2 and MYBL2 were highly expressed in GBM samples. Thus, above three predicted hub genes might play important roles in 'Cellular senescence' and 'Cell cycle' and were considered potential prognostic biomarkers in GBM and were the subject of further study.

3.4 | Tumour-infiltrating immune cells analyses and target interactive analyses

We used hub genes to predict immune-cell profiling in GBM by TIMER, a web tool to analysze infiltrated immune cells in the TCGA dataset. And we conducted the abundance of six types of tumour-infiltrating immune cells (B cells, CD4 + T cells, CD8 + T cells, neutrophils, macrophages and dendritic cells) and purity (Figure S3). The expression levels of RRM2, MYC and MCM2 were significantly correlated with dendritic cell (DC) infiltration. We detected that MCM2 had a positive correlation with the purity of GBM and was highly correlated to DCs. The corrections among the 10 hub genes in GBM were shown in Figure S4.

Subsequently, we predicted the target of ten hub genes and predicted their network interaction with lncRNA, miRNA and transcription factors (TF) (Figure 9 and Table S3). A total of 88 lncRNAs could target the ten hub genes (Table S3), such as HNF1A-AS1 could target EZH2 and MYC. In Figure 9A, CDK4 was targeted by hsa-miR-21-5p and hsa-miR-193b-3p. We also found TOP2A could regulate RRM2, TOP2A, CDC20 and EZH2 expression (Figure 9B). FOXM1 could interact with PAX2 (Paired Box 2) and E2F4 (E2F Transcription Factor 4). Combined the results of survival analysis and lncRNA-, miRNA- and TF network analysis, we were interested to further illustrate the molecular mechanism that were regulated these ten hub genes.

3.5 | Validation in the CGGA, HPA and CCLE database: CDC20, CCNB2, MYBL2

We further validated these ten central genes in the CGGA database and detected that CCNB2, CDC20 and MYBL2 expression was regular in different tissues (Figure 10, Figures S5 and S6). Subsequently, we evaluated the prognostic accuracy of CCNB2, CDC20 and MYBL2 by calculating the time-dependent ROC, AUC (area under the ROC curve) for one-, three- and five-year survival in glioma patients. CCNB2, CDC20 and MYBL2 showed good prognostic
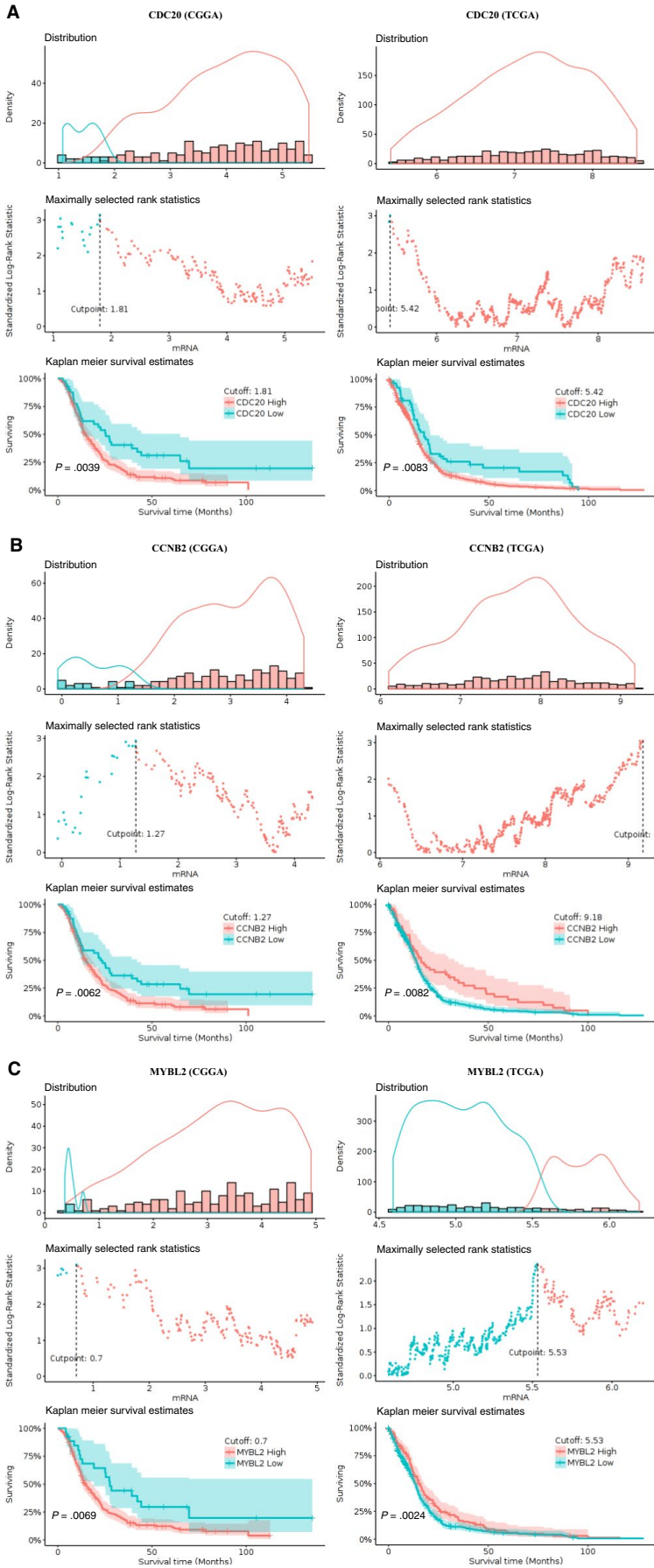
FIGURE 8 Kapla-Meier plots for DEGs to visualize the survival differences using GlioVis database
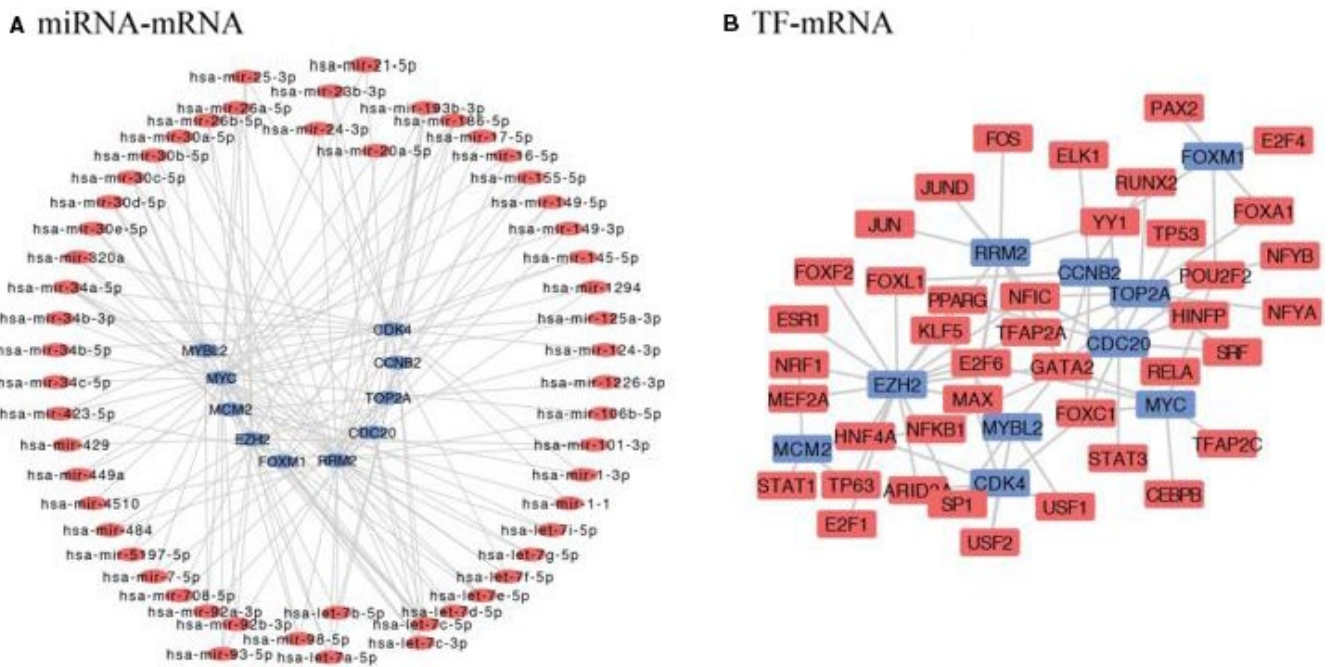
FIGURE 9 TF and miRNA with seven hub gene and their regulatory network. Blue shows hub genes, circle red shows miRNAs, square red means TF

accuracy for the CGGA dataset (Figure 10A, Figures S5A and S6A). After univariate and multivariate cox analysis of key clinical and molecular factors, we found that CCNB2 expression (*P*-value < 0.001, HR = 1.598/1.256), and CDC20 expression (*P*-value < 0.001, HR = 1.606/1.246), and MYBL2 expression (*P*-value < 0.001, HR = 1.570/1.258) were independent prognostic factors for gliomas in the CGGA dataset (Figure 10B-C, and Figures S5B,C and S6B,C). Because of the IDH1 mutation status and WHO classification are important for the prognosis of GBM, so it is necessary to determine whether our risk score is an independent prognostic factor for overall survival. The prognostic significance of IDH1 mutation status and 1p19q codeletion status in GBM in the CGGA database was shown to be highly significant. When GBM patients in the CGGA dataset were classified into two groups according to CDC20, CCNB2 and MYBL2 expression for survival analysis. For core hub genes, the expression levels of CCNB2, CDC20 and MYBL2 were also significantly higher in GBM with IDH1 mutations than in those of IDH1 wild-type (Figure 10H and Figures S5H and S6H). CCNB2, CDC20 and MYBL2 can classify patients into high-risk and low-risk groups according to different 1p19q_codeletion states (Figure 10D and Figures S5D and S6D), age (Figure 10E and Figures S5E and S6E) and chemical status (Figure 10F and Figures S5F and S6F), WHO ranks (Figure 10G and Figures S5G and S6G), IDH1 states (Figure 10H and Figures S5H and S6H) and PRS (main, recurrent, second category) Type (Figure 10I and Figures S5I and S6I), which p-value < 0.05. After univariate and multivariate cox analysis of core hub genes (CDC20, CCNB2 and MYBL2), we found these three genes independently indicated unfavourable prognosis in CGGA database, the expression of CCNB2, CDC20 and MYBL2 in different types could be independent prognostic marker in GBM.

We selected CCLE and HPA database to further validate the core hub genes. In glioma cells from CCLE, we could see the expression of core gene both in RNA-seq expression and Achilles shRNA knockdown (Figure 11). In GBM U251-MG cells from HPA, immunofluorescent staining of human cell line U251-MG showed us the gene location, green represents antibody, red means microtubules. CDC20 detected in the nucleoplasm and cytosol (Figure 11A), compared with other RNA cell lines, although the gene expression is not cell-dependent, the expression of CDC20 in U251-MG is the highest (Figure S7). CCNB2 localized to the cytosol and the Golgi apparatus, cell cycle dependent gene expression according to correlation analysis (Figure 11B). The localization of MYBL2 is the nucleoplasm (Figure 11C).

4 | DISCUSSION

Recent bioinformatic studies aimed to analyse differentially expressed genes, miRNAs and lncRNAs demonstrated the robustness of the results obtained through the integration of different GEO and TCGA datasets,[37,38] encouraged researchers to carry out the real-time analysis of in-motion big data, while protecting privacy and security.[39] With the development of sequencing and bioinformatics technology, more and more gene datasets are released on the public platform, we need to sort them out, and analysed them from more angles.[27,40,41] Compared with low-grade disease, hsa-miR-506-514 cluster, hsa-miR-592, hsa-miR-199a-5p were related to the overall survival of the uveal melanoma patients.[42] Up-regulated hsa-miR-183-5p and down-regulated hsa-miR-195-5p were directly related to colorectal cancer in the cancer development.[43] Increased C-MYC
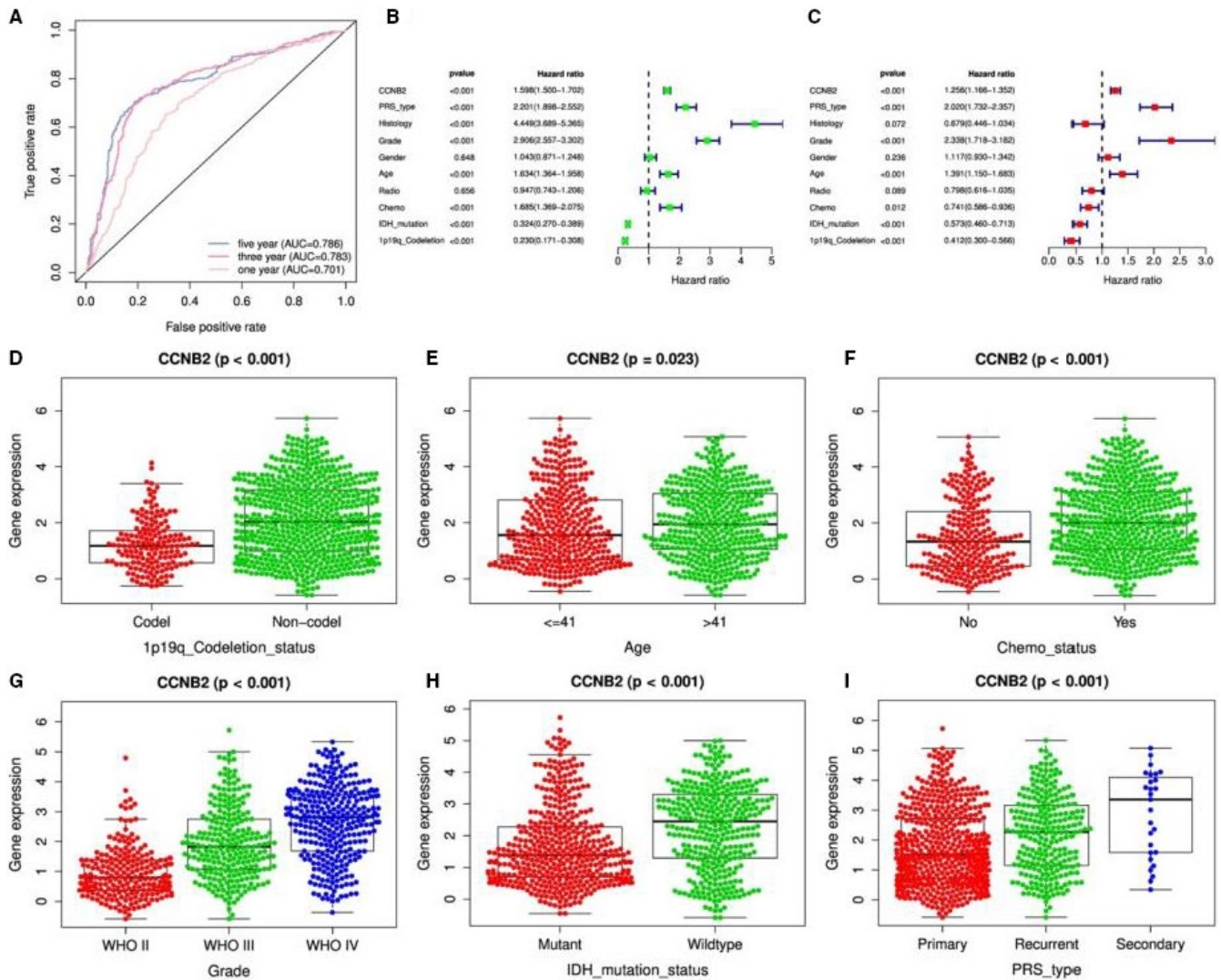
FIGURE 10 CCNB2 validated in CGGA database. A, Prognostic accuracy of the CCNB2 evaluated by the AUC of the time-dependent ROC with respect to 1, 3 and 5 year survival of glioma patients in the CGGA dataset; B and C, Univariate and multivariate cox analysis of CCNB2 in CGGA; D. to J. The expression of CCNB2 in different types

associated with glucocorticoid and was resistant in acute lymphoblastic leukaemia.[44] In this study, we selected six GEO datasets (389 GBM tumour and 67 normal tissues) and TCGA-GBM dataset (169 GBM and 5 normal samples), which integrated them by the RRA method, with the criteria of log2 (fold change) > 1 and adjusted p-value < 0.05, we filtered 133 DEGs (82 up-regulated and 51 down-regulated) in totally. After PPI network and functional analysis, we identified ten DEGs (FOXM1, CDK4, TOP2A, RRM2, MYBL2, MCM2, CDC20, CCNB2, MYC and EZH2) as core genes in GBM. GO, KEGG and GSEA enrichment analyses were used to further explore pathways for the development and progression of GBM. TIMER immunity infiltration and survival analysis were validated in conjunction with the TCGA database. At the same time, we further verified the core genes in the CGGA, HPA and CCLE database.

FOXM1 (Forkhead Box M1) is a member of FOX family and located on the chr12p13.33, which emerged as a key molecule implicated in initiation and progression of cancer.[45] FOXM1 is a high-risk myeloma gene with poor prognosis, FOXM1 was up-regulated between GBM and normal tissues, and enriched in the GO term: cell cycle arrest, G2/M transition of mitotic cell cycle, negative regulation of stress-activated MAPK cascade, suggesting that FOXM1 might be a potential gene in gliomas development. MYBL2 (MYB Proto-Oncogene Like 2) is a member of the MYB family of TF genes, is a key downstream factor of AKT/FOXM1 signalling to promote progression of human glioma.[46] The expression of COL1A2 (Collagen Type I Alpha 2 Chain) in GBM could significantly improve survival benefit after aggressive treatment compared with the proneural patients.[47] COL1A2 was associated with poor outcomes in GBM and validated to be significantly linked to poor prognosis in both TCGA and CGGA database,[48] as well as our study, so the over expression of COL1A2 might be important to the development of GBM. CDC20 (Cell Division Cycle 20) was reported as a target for overcoming TMZ-resistance (Temozolomide resistance) in GBM.[49] CDC20 detected in the nucleoplasm and cytosol, compared with other RNA cell lines, although the gene expression is not cell-dependent, the expression of CDC20 in U251-MG GBM cell line is the highest. We
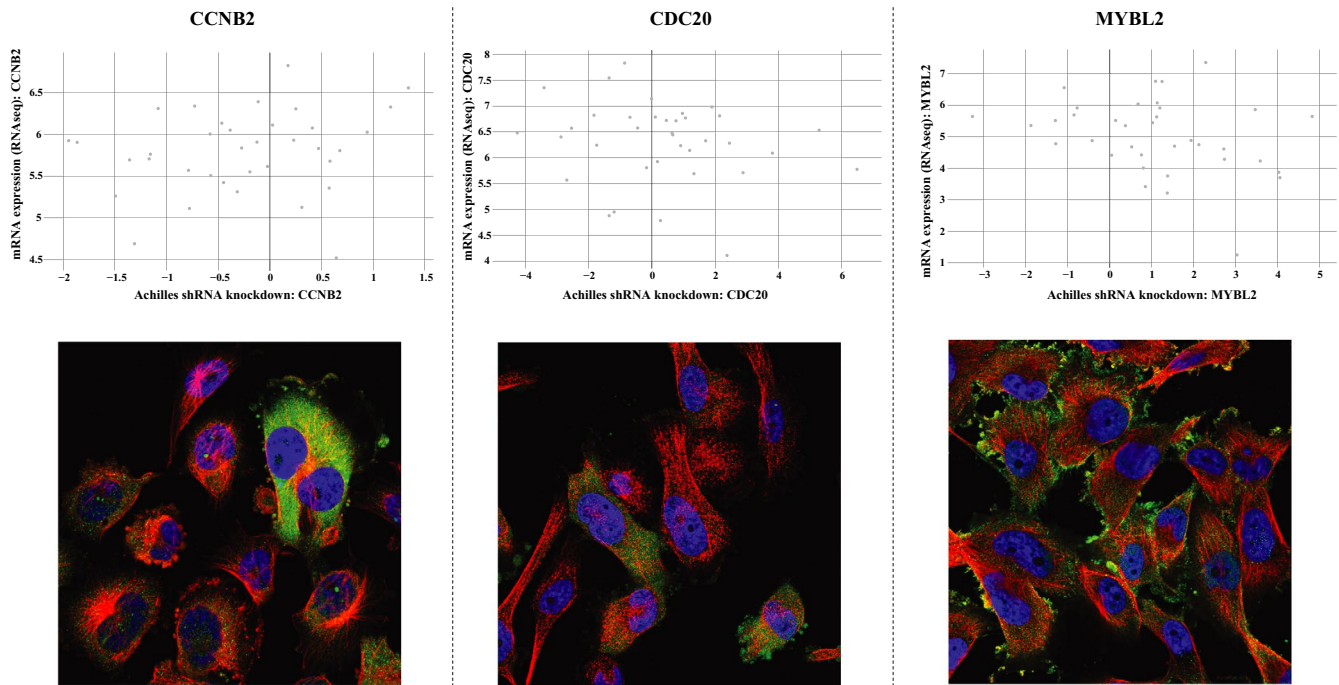
FIGURE 11 Validation of core hub genes in CCLE and HPA database. A, CCNB2; B, CDC20; C, MYBL2. The upper part of figure means the comparison of the RNA-seq expression in the glioma cells and the Achilles shRNA knockdown gene. The X axis is achilles shRNA knockdown: core gene, Y axis means mRNA expression (RNA-seq): core gene. The lower part of figure shows immunofluorescent staining of human cell line U251-MG, green represents antibody, red means microtubules, blue means nucleus. The scale bar was 20 μm

also found CDC20 is overexpressed in GBM with a poor prognosis in GBM patients.

DC are the antigen-presenting cells, pathways of DCs are important to control the development of immune response and decisions on vaccination.[50] MYC (MYC Proto-Oncogene, BHLH Transcription Factor) in GBM was highly correlated with DC (partial cor = 0.27, P-value = 1.95e-08), could control the immune response. Activation of the MYC signalling pathway in normal astrocytes exposed to GBM-EV may be the mechanism by which GBM acquires a phenotype that promotes tumour progression,[51] MYC was enriched in developmental process and protein binding in this study, was consistent with the predictions of this study.

The miRNAs found altered in GBM have been widely reported, hsa-miR-21 was overexpressed in the development and progression of GBM.[52] Our previous study has established ceRNA network and identified miRNA, lncRNA and TF were glioma-related molecules in GBM.[6] In current study, hsa-miR-21-5p was involved in the regulation of CDK4 and MYC, which known to be detected in the GBM development. Six lncRNA-related ceRNA combined with four small molecule compounds were considered to help identify the regulatory functions of lncRNAs in the pathogenesis of GBM.[53] Via recruiting EZH2, LncRNA HOTAIR modulated the chromatin architecture.[54] We also found that lncRNA APTR and lncRNA H19 were up-regulated in GBM which interacted with EZH2, played a role in inhibiting the cell proliferation and promoting the tumorigenesis. LncRNA HOTAIR could also down-regulate EZH2 promoting cell cycle, cell

proliferation and cell invasion. E2F4 was increased in GBM, which stimulated the GBM growth.[55] FOXM1 could also interact with E2F4 in this study. It was confirmed that the miRNAs here identified are able to target the hub genes here identified (specially CDK and cycline families).

Ten hub genes were all overexpressed in GBM, functional enrichment results focused on these up-regulated genes, there were many significant enrichment results. CDK4, MCM2 and MYC were response to G1/S transition of mitotic cell cycle (GO:0 000 082) in biological process and enriched in the cell cycle (hsa04110) in KEGG pathway. CDK4, RRM2 and CCNB2 were related to p53 signalling pathway (hsa04115), which might activate the p53 signalling pathway via CDK4, RRM2 and CCNB2, might also have regulatory effects in glioma cells. EZH2 (Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit) was up-regulated in GBM, lncRNA HOTAIR, APTR and H19 could interact with EZH2, which promoted cell cycle, cell proliferation, cell invasion and tumorigenesis in GBM.

In our study, CCNB2, CDC20 and MYBL2 were up-regulated in GBM compared with control brain tissues, with poor prognosis in GBM patients. We found that mRNA expression of CCNB2, CDC20 and MYBL2 was significantly different in primary, recurrent and secondary GBM (primary > recurrent>secondary), suggesting that CCNB2, CDC20 and MYBL2 might be involved in the origin of GBM. The results described above indicated that CCNB2, CDC20 or MYBL2 was independent prognostic marker for overall survival and might play a significant role in determining glioma prognosis, and the

association between CCNB2, CDC20, MYBL2 and GBM should be investigated further.

5 | CONCLUSIONS

In summary, bioinformatics analysis in the context of big data identified key roles for ten hub genes FOXM1, CDK4, TOP2A, RRM2, MYBL2, MCM2, CDC20, CCNB2, MYC and EZH2 in the development, progression, diagnosis, treatment and prognosis of GBM. These results provide candidate gene for molecular targeting therapy and biomarker for radiotherapy of GBM cancer, and further in vivo and in vitro experiments are needed to validate the role of these screened genes and pathways. At the same time, further research is needed for the synergistic interaction between CDC20, CCNB2 and MYBL2.

CONFLICTS OF INTEREST

The authors confirm that there are no conflicts of interest.

AUTHOR CONTRIBUTION

Lan Jiang: Data curation (equal); Formal analysis (equal); Validation (equal); Writing-original draft (equal); Writing-review & editing (equal). **Min Zhong:** Formal analysis. **Tianbing Chen:** Formal analysis (equal); Funding acquisition (equal). **Xiaolong Zhu:** Formal analysis (equal). **Hui Yang:** Formal analysis (equal). **Kun Lv:** Data curation (equal); Funding acquisition; Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

The microarray data of human tissue from GSE2223, GSE4058, GSE4290, GSE13276, GSE68848 and GSE70231 were downloaded from the Gene Expression Omnibus (GEO) database.

TCGA-GBM database.

ORCID

Lan Jiang (iD) https://orcid.org/0000-0003-2778-1081

REFERENCES

1. Neftel C, Laffy J, Filbin MG, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell*. 2019;178:835–849.e21.
2. Ito H, Nakashima H, Chiocca EA. Molecular responses to immune checkpoint blockade in glioblastoma. *Nat Med*. 2019;25:359–361.
3. Komori T. The 2016 WHO classification of tumours of the central nervous system: the major points of revision. *Neurol Med Chir*. 2017;57:301-311.
4. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131:803-820.
5. Sasmita AO, Wong YP, Ling APK. Biomarkers and therapeutic advances in glioblastoma multiforme. *Asia Pacific J Clin Oncol*. 2018;14:40-51.
6. Jiang L, Yang H, Chen T, Zhu X, Ye J, Lv K. Identification of HMG-box family establishes the significance of SOX6 in the malignant progression of glioblastoma. *Aging*. 2020;12:8084–8106..
7. Tuaeva NO, Falzone L, Porozov YB, et al. Translational application of circulating DNA in oncology: review of the last decades achievements. *Cells*. 2019;8:1251.
8. Silantyev AS, Falzone L, Libra M, et al. Current and future trends on diagnosis and prognosis of glioblastoma: from molecular biology to proteomics. *Cells*. 2019;8:863.
9. Li J-H, He Z-Q, Lin F-H, et al. Assessment of ctDNA in CSF may be a more rapid means of assessing surgical outcomes than plasma ctDNA in glioblastoma. *Molecul Cell Prob*. 2019;46:101411.
10. Giannini C, Sarkaria JN, Saito A, et al. Patient tumor EGFR and PDGFRA gene amplifications retained in an invasive intracranial xenograft model of glioblastoma multiforme. *Neuro-oncology*. 2005;7:164-176.
11. Jung C, Foerch C, Schänzer A, et al. Serum GFAP is a diagnostic marker for glioblastoma multiforme. *Brain*. 2007;130:3336-3341.
12. Yang Q, Wang R, Wei B, et al. Candidate biomarkers and molecular mechanism investigation for glioblastoma multiforme utilizing WGCNA. *BioMed Res Int*. 2018;2018:1–10.
13. Zou YF, Meng LB, He ZK, et al. Screening and authentication of molecular markers in malignant glioblastoma based on gene expression profiles. *Oncol Lett*. 2019;18:4593-4604.
14. Miyauchi E, Furuta T, Ohtsuki S, et al. Identification of blood biomarkers in glioblastoma by SWATH mass spectrometry and quantitative targeted absolute proteomics. *PLoS One*. 2018;13.
15. Bredel M, Bredel C, Juric D, et al. Tumor necrosis factor-α–induced protein 3 as a putative regulator of nuclear factor-κB–mediated resistance to O6-alkylating agents in human glioblastomas. *J Clin Oncol*. 2006;24:274-287.
16. Liang Y, Diehn M, Watson N, et al. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA*. 2005;102:5814-5819.
17. Sun L, Hui A-M, Su Q, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell*. 2006;9:287-300.
18. Mangiola A, Saulnier N, De Bonis P, et al. Gene expression profile of glioblastoma peritumoral tissue: an ex vivo study. *PLoS One*. 2013;8:e57145.
19. Madhavan S, Zenklusen J-C, Kotliarov Y, Sahni H, Fine HA, Buetow K. Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res*. 2009;7:157-167.
20. Rickman DS, Bobek MP, Misek DE, et al. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Can Res*. 2001;61:6885-6891.
21. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York, NY: Springer; 2005:397-420.
22. Xu J, Hou X, Pang L, et al. Identification of dysregulated competitive endogenous RNA networks driven by copy number variations in malignant gliomas. *Front Genet*. 2019;10:1055.
23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139-140.
24. Wickham H. ggplot2. *Wiley Interdisc Rev Comput Stat*. 2011;3:180-185.
25. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. 2015;19:A68.

26. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573-580.

27. Jiang L, Bi D, Ding H, et al. Systematic identification and evolution analysis of Sox genes in *Coturnix japonica* based on comparative genomics. *Genes*. 2019;10:314.

28. Kolde R, Kolde MR. Package 'pheatmap' [J]. *R Package*. 2015;1(7):790.

29. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2018;47:D607-D613.

30. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504.

31. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284-287.

32. Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Can Res*. 2017;77:e108-e110.

33. Zhou G, Soufan O, Ewald J, Hancock RE, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. 2019.

34. Heagerty PJ, Saha-Chaudhuri P, Saha-Chaudhuri MP. Package 'survivalROC' [J]. 2013.

35. Therneau TM, Lumley T. Package 'survival' [J]. *R Top Doc*. 2015;128:112.

36. Dillingham I, McCarthy A. Creating a Buzz around Corporate Reputation with Beeswarm Plots[J]. 2019.

37. Zhou RS, Zhang EX, Sun QF, et al. Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue. *BMC Cancer*. 2019;19:779.

38. You Z, Zhang Q, Liu C, Song J, Yang N, Lian L. Integrated analysis of lncRNA and mRNA repertoires in Marek's disease infected spleens identifies genes relevant to resistance. *BMC Genom*. 2019;20:245.

39. Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare Informat Res*. 2013;19:79-85.

40. Cao Y, Meng D, Han Y, et al. Comparative analysis of B-BOX genes and their expression pattern analysis under various treatments in *Dendrobium officinale*. *BMC Plant Biol*. 2019;19:245.

41. Jiang L, Bi D, Ding H, Ren Q, Wang P, Kan X. Identification and comparative profiling of gonadal microRNAs in the adult pigeon (*Columba livia*). *Br Poult Sci*. 2019;60:638–648.

42. Falzone L, Romano GL, Salemi R, et al. Prognostic significance of deregulated microRNAs in uveal melanomas. *Mol Med Rep*. 2019;19:2599-2610.

43. Falzone L, Scola L, Zanghì A, et al. Integrated analysis of colorectal cancer microRNA datasets: identification of microRNAs associated with tumor development. *Aging (Albany NY)*. 2018;10:1000-1014.

44. Chen Y, Jiang P, Wen J, et al. Integrated bioinformatics analysis of the crucial candidate genes and pathways associated with glucocorticoid resistance in acute lymphoblastic leukemia. *Cancer Med*. 2020;9:2918-2929.

45. Nandi D, Cheema PS, Jaiswal N, Nag A. FoxM1: repurposing an oncogene as a biomarker. *Semin Cancer Biol*. 2018;52:74–84.

46. Zhang X, Qiao-Li L, Huang Y-T, Zhang L-H, Zhou H-H. Akt/FoxM1 signaling pathway-mediated upregulation of MYBL2 promotes progression of human glioma. *J Exp Clin Cancer Res*. 2017;36:105.

47. Bernstock JD, Mooney JH, Ilyas A, et al. Molecular and cellular intratumoral heterogeneity in primary glioblastoma: clinical and translational implications. *J Neurosurg*. 2019;1:1-9.

48. Di Jia SL, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)*. 2018;10:592.

49. Wang J, Zhou F, Li Y, et al. Cdc20 overexpression is involved in temozolomide-resistant glioma cells with epithelial-mesenchymal transition. *Cell Cycle*. 2017;16:2355-2365.

50. Kawano M, Tanaka K, Itonaga I, et al. Dendritic cells combined with doxorubicin induces immunogenic cell death and exhibits antitumor effects for osteosarcoma. *Oncology Lett*. 2016;11:2169-2175.

51. Hallal S, Mallawaaratchy D, Wei H, et al. Extracellular vesicles released by glioblastoma cells stimulate normal astrocytes to acquire a tumor-supportive phenotype via p53 and MYC signaling pathways. *Mol Neurobiol*. 2019;56:4566-4581.

52. Candido S, Lupo G, Pennisi M, et al. The analysis of miRNA expression profiling datasets reveals inverse microRNA patterns in glioblastoma and Alzheimer's disease. *Oncol Rep*. 2019;42:911-922.

53. Liu Z, Wang X, Yang G, et al. Construction of lncRNA-associated ceRNA networks to identify prognostic lncRNA biomarkers for glioblastoma. *J Cell Biochem*. 2020;121:3502-3515.

54. Janaki Ramaiah M, Divyapriya K, Kartik Kumar S, Rajesh Y. Drug-induced modifications and modulations of microRNAs and long non-coding RNAs for future therapy against Glioblastoma Multiforme. *Gene*. 2020;723:144126.

55. Donaires FS, Godoy PR, Leandro GS, Puthier D, Sakamoto-Hojo ET. E2F transcription factors associated with up-regulated genes in glioblastoma. *Cancer Biomarker*. 2017;18:199-208.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.