Identification of key modules and hub genes in glioblastoma multiforme based on co-expression network analysis

Chun Li[1], Bangming Pu[2], Long Gu[3], Mingwei Zhang[4], Hongping Shen[1], Yuan Yuan[1] and Lishang Liao[4] (iD)

1 GCP Center, the Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou, China
2 Department of Hepatobiliary Surgery, the Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou, China
3 Department of Emergency Medicine, the Affiliated Hospital of Southwest Medical University, Luzhou, China
4 Department of Neurosurgery, the Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou, China

Glioblastoma multiforme (GBM) is the most malignant primary tumour in the central nervous system, but the molecular mechanisms underlying its pathogenesis remain unclear. In this study, data set GSE50161 was used to construct a co-expression network for weighted gene co-expression network analysis. Two modules (dubbed brown and turquoise) were found to have the strongest correlation with GBM. Functional enrichment analysis indicated that the brown module was involved in the cell cycle, DNA replication, and pyrimidine metabolism. The turquoise module was primarily related to circadian rhythm entrainment, glutamatergic synapses, and axonal guidance. Hub genes were screened by survival analysis using The Cancer Genome Atlas and Human Protein Atlas databases and further tested using the GSE4290 and Gene Expression Profiling Interactive Analysis databases. The eight hub genes (*NUSAP1*, *SHCBP1*, *KNL1*, *SULT4A1*, *SLC12A5*, *NUF2*, *NAPB*, and *GARNL3*) were verified at both the transcriptional and translational levels, and these gene expression levels were significant based on the World Health Organization classification system. These hub genes may be potential biomarkers and therapeutic targets for the accurate diagnosis and management of GBM.

Glioblastoma multiforme (GBM), which accounts for 47.1% of primary malignant brain tumours [1], is classified as a grade IV glioma by the World Health Organization (WHO) and a highly lethal tumour [2]. GBM is considered incurable, with a 5-year survival rate of only 5.5% [1]. The poor prognosis is also related to the high recurrence rate of tumours despite aggressive multimodality treatments, including maximal surgical resection, radiotherapy, and chemotherapy [3].

Molecular data on brain tumours have a significant impact on prognosis and clinical management [4]. In 2016, the WHO classification of tumours of the central nervous system improved the traditional diagnostic criteria—which was based only on histological properties

Abbreviations
AGGF1, angiogenic factors with G-patch and FHA domain1; BPs, biological processes; CGGA, Chinese Glioma Genome Atlas; CIN, chromosomal instability; DEGs, differentially expressed genes; GBM, glioblastoma multiforme; GO, Gene Ontology; HPA, the Human Protein Atlas; IHC, immunohistochemically; KEGG, Kyoto Encyclopedia of Genes and Genomes; NDC80, the nuclear division cycle 80; SAC, the spindle assembly checkpoint; TCGA, The Cancer Genome Atlas; TFs, transcription factors; WGCNA, weighted gene co-expression network analysis; WHO, the World Health Organization.

—and incorporated molecular markers, by establishing the 'ISN-Haarlem' consensus [5]. Specifically, an integrated diagnostic approach based on clinical, histopathological, and molecular data is required to more accurately distinguish glioma tumour subtypes and define prognosis in response to specific treatments. According to a recent study, glioblastoma patients may benefit from molecular targeted therapies [6]. Here, we report two modules using weighted gene co-expression network analysis (WGCNA) [7], which might provide potential biomarkers and therapeutic targets for more accurate diagnosis and treatment of GBM.

WGCNA, using soft-thresholding techniques to convert a gene co-expression similarity measure into a network connection strength, can reconstruct gene co-expression modules and summarise modules using module eigengene [8]. WGCNA can yield more robust results as compared to unweighted networks [7]. WGCNA is one of the most extensive methods of genomic analysis, and it has a high degree of superiority by focussing on a group of genes, rather than a single gene, to minimise bias. In addition, WGCNA does not require cut-off criteria and can retrieve momentous information, which is otherwise omitted with other methods. WGCNA has been applied in various cancers, such as pancreatic cancer, breast cancer, and osteosarcoma [9–11]. To identify key biomarkers and to further understand the molecular mechanisms of GBM, we used WGCNA to analyse GBM from a new perspective in this study.

Materials and methods

Data information

Two microarray profiles, including GBM samples (GSE50161 and GSE4290), were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) [12]. As for the GSE50161 data set, 34 paediatric GBM samples and 13 normal samples were selected for WGCNA [13]. As for the GSE4290 data set, 81 malignant glioma cell samples and 23 nontumour samples from epilepsy patients were selected for validation [14]. The two data sets above were both based on the platform of the Affymetrix Human Genome U133 Plus 2.0 Array.

Construction of WGCNA

GBM and normal samples were ranked by the median absolute deviation from large to small, and we precalculated the power parameter of the top 5000 genes using the

pickSoftThreshold function of WGCNA [8]. This function provided the appropriate soft-thresholding power for network construction by calculating the scale-free topology fit index for several powers. Adjacency was turned into topological overlap, which could measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for the network generation. Modules were grouped with tightly connected genes, which had similar expression profiles, and then identified on the dendrogram using the Dynamic Tree Cut algorithm. The first principal component of a given module is defined as the module eigengene, which can be considered as a representative of the gene expression profile in a module. The dissimilarity of module eigengenes was calculated to choose a cutline to merge some modules.

Identification of significant GBM modules

To find significant gene modules, we created module–trait relationships to detect the correlation between module eigengenes and the GBM trait. The $\log_{10}$ transformation of the *P*-value was defined as the gene significance. The higher the absolute value of gene significance, the more biologically significant the gene [8]. Module significance was determined as the average absolute gene significance measure for all genes in a given module. The module membership measure is highly related to the intramodular connectivity. Highly connected intramodular hub genes tend to have high module membership values to the respective module. Generally, the module with the absolute module significance ranked first among all the selected modules was considered as the one most related to GBM traits and was called the key module.

Functional enrichment analysis

In order to gain further insight into the function of genes in key modules, eigengenes in the key module were subjected to Gene Ontology (GO) enrichment analysis to identify the enrichment results of biological processes (BPs), cell components, and molecular functions using the CLUS-TERPROFILER R package [15]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of key modules were also performed [16]. Fisher's exact test based on hypergeometric distribution was used for statistical analysis. $P < 0.05$ was considered to have statistical significance and to achieve significant enrichment.

Identification of hub genes in key modules

R function exportNetworkToCytoscape was used to export the key modules as networks in text format for CYTOSCAPE (ver3.6.1, https://cytoscape.org/) [17]. After importing the texts in CYTOSCAPE, we sorted the node degree of genes in

the key modules and screened the top 30 genes in each module. The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) is a landmark cancer genomics programme that contains over 20 000 molecularly characterised primary cancer and matched normal samples spanning 33 cancer types [18]. To further screen hub genes among the key modules, we downloaded the GBM gene expression data from TCGA database for survival analysis (grouped by median expression level) of the top 30 genes in each module above. If there were no significant results of survival analysis by TCGA in a given module, we used the Human Protein Atlas (HPA, https://www.proteinatlas.org/) to screen candidate genes associated with survival in the module by searching for the best cut-off expression value in TCGA. By limiting the sample size in the group, the grouping was not less than 20% of the total sample in the HPA, and genes with the lowest log-rank *P*-values (log-rank $P < 0.05$) of survival analysis were selected [19]. The R package limma was applied to screen differentially expressed genes (DEGs) between GBM and nontumour samples in GSE50161 [20]. The corresponding *P*-values of less than 0.01 and log |fold change (FC)| larger than 1.5 were selected as the cut-off criteria for DEGs. DEGs associated with survival were considered to be hub genes in GBM tumorigenesis.

Validation of hub genes

We performed hub genetic verification in two separate data sets from different sources. For data set GSE4290, the screening methods and cut-off criteria for DEGs were the same as in GSE50161. Volcano plots and hierarchical clustering analyses were carried out using R packages ggplot2 and pheatmap, respectively. GO function enrichment and KEGG pathway analysis were also performed on DEGs. The differential expression levels of these hub genes were also verified in GEPIA (http://gepia.cancer-pku.cn/) [21]. The GBM and normal data sources were from TCGA and the Genotype-Tissue Expression project [22]. In addition, the HPA database provided immunohistochemically (IHC) stained specimens of the proteins of genes. It was used for verifying the translational levels of hub genes in normal and tumour tissues (including three normal samples and at least nine GBM samples).

Hub gene expression distribution (WHO classification)

To further investigate the distribution of these eight key genes in the WHO classification, we used R to analyse the mRNAseq_693 data set in the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn/) database. The mRNAseq_693 data set contained 188 GBM samples in WHO grade II, 255 samples in grade III and 249 samples in grade IV. Gene expression data were subjected to a *t*-test

after log2 (expression + 0.001) conversion. A *P*-value of less than 0.05 was considered statistically significant.

Prediction of transcription factors of hub genes

Identification of the transcription factors (TFs) that operate a perturbed gene network and detection of their target genes are instrumental steps in uncovering key insights into oncogenic programmes, including the discovery of therapeutic targets [23–25]. The iRegulon in CYTOSCAPE is a plugin to reverse engineer the transcriptional regulatory network underlying a co-expressed gene set using cis-regulatory sequence analysis [26]. iRegulon relies on the analysis of the regulatory sequences around each gene in the gene set to detect enriched TF motifs or ChIP-seq peaks, using databases of nearly 10 000 TF motifs and 1000 ChIP-seq data sets or 'tracks'. Next, it associates enriched motifs and tracks with candidate TFs and determines the optimal subset of direct target genes. In order to explore the TFs of hub genes in key modules, we used the iRegulon plugin to predict the TFs of hub genes in the two modules. The screening conditions were normalised enrichment scores > 3.

Validates the role of high expression of Hub gene by siRNA transfection knockdown method

Hub gene identification has confirmed the high expression of four Hub genes of *NUSAP1, SHCBP1, NUF2* and *KNL1* in tumours. In order to further study the role of these four genes in GBM, we used siRNA transfection knockdown method to knock down the mRNA expression of four Hub genes of *NUSAP1, SHCBP1, NUF2 and KNL1* in human astroblastoma cell line U87 (it was purchased from the Committee on Type Culture Collection of the Chinese Academy of Sciences. Shanghai, China) and then determined the level of mRNA of four Hub genes by RT-PCR method (All primers are listed in Table 1); next, we analysed the variation of proliferation and cloning ability of U87 cells by CCK8 method and clone formation

Table 1. The sequences of siRNA and RT-qPCR primers.

Gene	Primer	Tm
SAP1-F	TGGACCTCTAATGATGGGCAG	60
SAP1-R	AGGCTTGTTCTTGCGAATCCC	60
SHCBP1-F	GGAAGTGTATCCTGTTGAGGGA	60
SHCBP1-R	ACCAGGTATTGTTCCATCCTGT	60
NUF2-F	GGAAGGCTTCTTACCATTCAGC	60
NUF2-R	GACTTGTCCGTTTTGCTTTTGG	60
KNL1-F	CTTCACACCGAGGACTCAAGA	60
KNL1-R	TTTGATGTGTAGAAGAGGCACTG	60
GAPDH-F	GGAGCGAGATCCCTCCAAAAT	60
GAPDH-R	GGCTGTTGTCATACTTCTCATGG	60

experiment, and finally, the role of four Hub genes was validated.

Statistical analyses

Data were analysed by using IBM SPSS 21.0 software (IBM, Chicago, IL, USA) and are presented as the mean ± SD. Quantitative PCR results and clone formation were analysed by Student's *t*-test, and CCK8 cell proliferation was analysed by one-way analysis.

Results

WGCN construction and module selection

We analysed 5000 genes, which were divided into two clusters (Fig. 1A). Our fitting degree of the scale-free topological model was 0.85. Thus, this network conformed to the power-law distribution and was close to the real biological network state. Network topology for thresholding powers from 1 to 20 was performed, and the relatively balanced scale independence (Fig. 1B) and mean connectivity (Fig. 1C) of the WGCNA were identified subsequently. It showed that 8 had the best power value. Thus, β = 8 was selected to produce a hierarchical clustering tree (dendrogram) of 5000 genes. As a result, nine modules were identified (Fig. 2A). Different colours represented different modules, and genes that could not be classified in any module were put into the grey module, which was removed in the subsequent analysis (Fig. 2B).

Correlation between modules and identification of key modules

The analytic results of the interaction relationships of nine modules showed that genes within modules displayed more topological overlap than the genes across modules, according to the topological overlap heatmap in the gene network (Fig. 3A). This revealed that
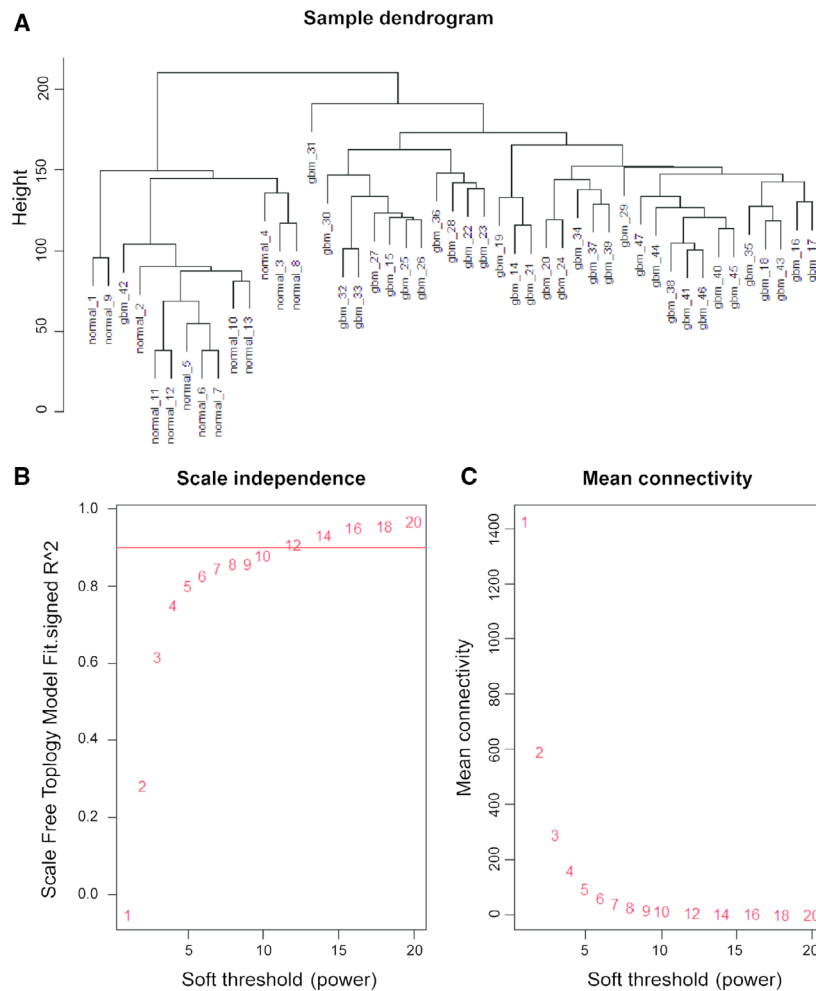
Fig. 1. Clustering of samples and determination of soft-thresholding power. (A) Clustering of samples. (B) The scale-free topology fitting index (R^2, y-axis) as a function of the soft-thresholding power (x-axis). The red line indicates $R^2 = 0.9$. (C) The mean connectivity (degree, y-axis) is displayed as a function of the soft-thresholding power (x-axis). Red Arabic numbers in the panels denote different soft-thresholds. β = 8, there was a trade-off between maximising the scale-free topology model fitting index (R^2) and maintaining a high mean number of connections.
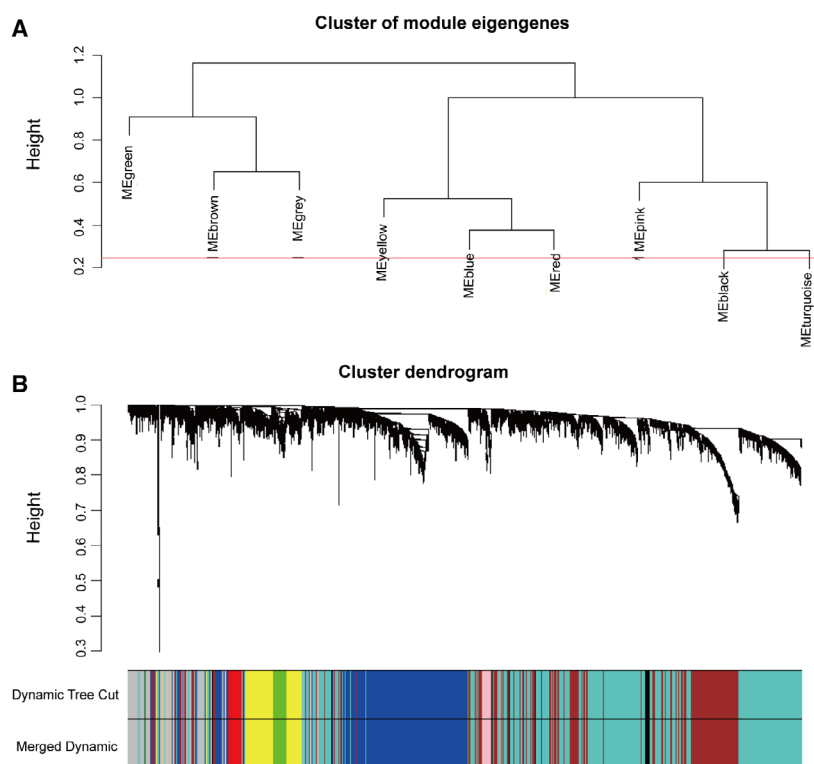
Fig. 2. Cluster of module eigengenes and cluster dendrogram. (A) Cluster dendrogram of module eigengenes. (B) The cluster dendrogram of genes in GSE50161. Each branch of the dendrogram represents a gene. Each designated colour represents a co-expression module.

each module was independent of each other. Similar results were demonstrated by the heatmap plotted according to adjacencies (Fig. 3C). The module eigengene of the turquoise and brown modules revealed a high correlation with disease status compared with other modules (Fig. 3B). The turquoise module (correlation index: -0.85, $P = 8\mathrm{e}^{-14}$) was negatively correlated with the disease, while the brown module (correlation index: 0.88, $P = 8\mathrm{e}^{-15}$) was positively correlated. Thus, we identified turquoise and brown modules as the key modules. Figure 4A,B illustrate the strong correlation between module membership and gene significance in brown and turquoise modules, respectively.

## Functional enrichment analysis of the two key modules

For the GO enrichment analysis of key modules, the detailed information of the brown module is shown in Fig. 5A. They were mainly enriched in cell division, including chromosome segregation, organelle fission, and nuclear division, all of which were positively correlated with tumorigenesis. Detailed information of the turquoise module is shown in Fig. 5B. They were mainly enriched in the regulation of the nervous system, such as positive regulation of neuronic

development and differentiation, which were negatively correlated with tumorigenesis. Detailed information of the KEGG pathway analysis for key modules is shown in Fig. 6A,B. The brown and turquoise modules were mainly focussed on the cell cycle and neural signalling pathway, respectively. To provide solid insights, data set GSE4290 was analysed, and the results of DEGs in the data set are shown in Fig. S1A,B. KEGG pathway analysis and GO enrichment analysis of the DEGs were carried out (Figs S1C and S2). As we have seen, the KEGG pathway is enriched in pathways such as the cell cycle and GABAergic synapse. In the GO functional analysis of DEGs, the BP components of upregulated genes were mainly enriched in chromosome segregation and DNA replication, and the BP components of down-regulated genes were mainly enriched in neuronic development and synaptic function. This finding is consistent with the functional analysis of our key modules.

Identification of hub genes in key modules

The top 30 genes screened in each module using CY-TOSCAPE are shown in Fig. 6C,D. Figure 7 shows the results of the survival analysis. Figure 8A is the result of the differential expression level analysis. *NUSAP1*, *SHCBP1*, *NUF2* and *KNL1* were the hub genes in the
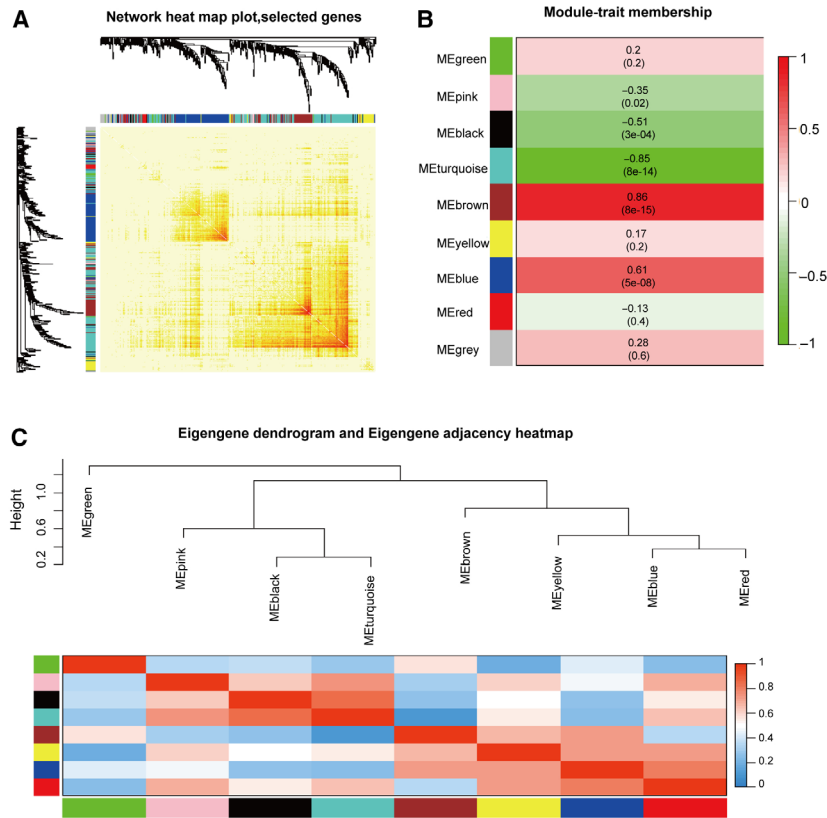
Fig. 3. Identification of key modules. (A) Network heatmap plot of co-expression genes. Different colours of the horizontal axis and the vertical axis represent different modules. The progressively saturated yellow and red colours indicate the high co-expressed interrelation in the heatmap. (B) Heatmap of the correlation between module eigengenes and GBM. Each row represents a module eigengene and each column represents trait. The table is coloured by correlation according to the colour legend. The turquoise module is the most negatively correlated with GBM, and the brown module is the most positively correlated with GBM. (C) The hierarchical clustering dendrogram and the adjacency heatmap of each module. The top is the hierarchical clustering of the module crucial genes (labelled by colours). The bottom is the adjacency heatmap, where each column and row correspond to one module crucial gene (labelled by colour) or trait labelled by *y*. Red represents high adjacency (positive correlation), while blue colour represents low adjacency (negative correlation). Squares of red colour along the diagonal are the meta-module.
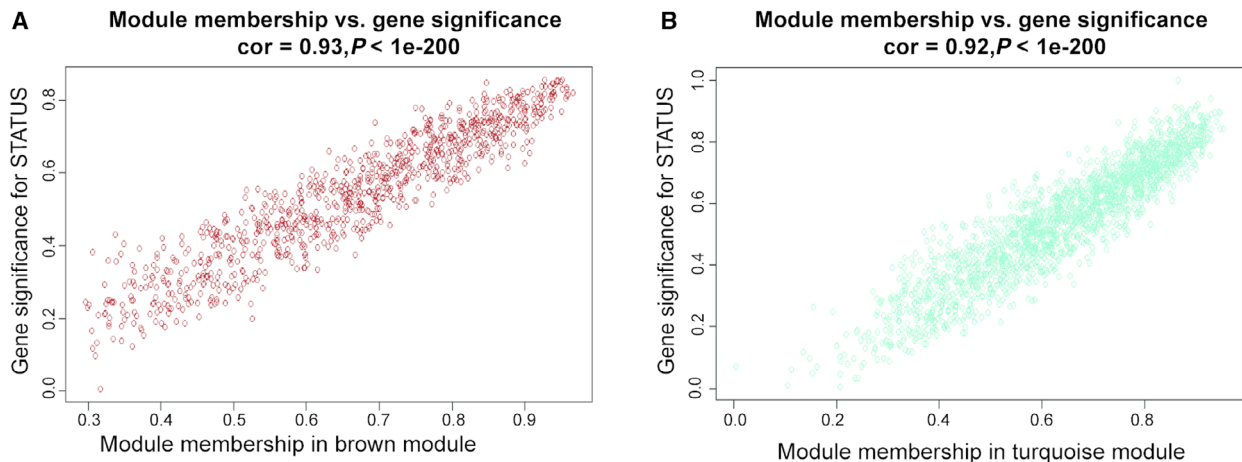
Fig. 4. Module eigengenes in key modules. (A) Scatter plot of module eigengenes in the brown module. (B) Scatter plot of module eigengenes in the turquoise module.
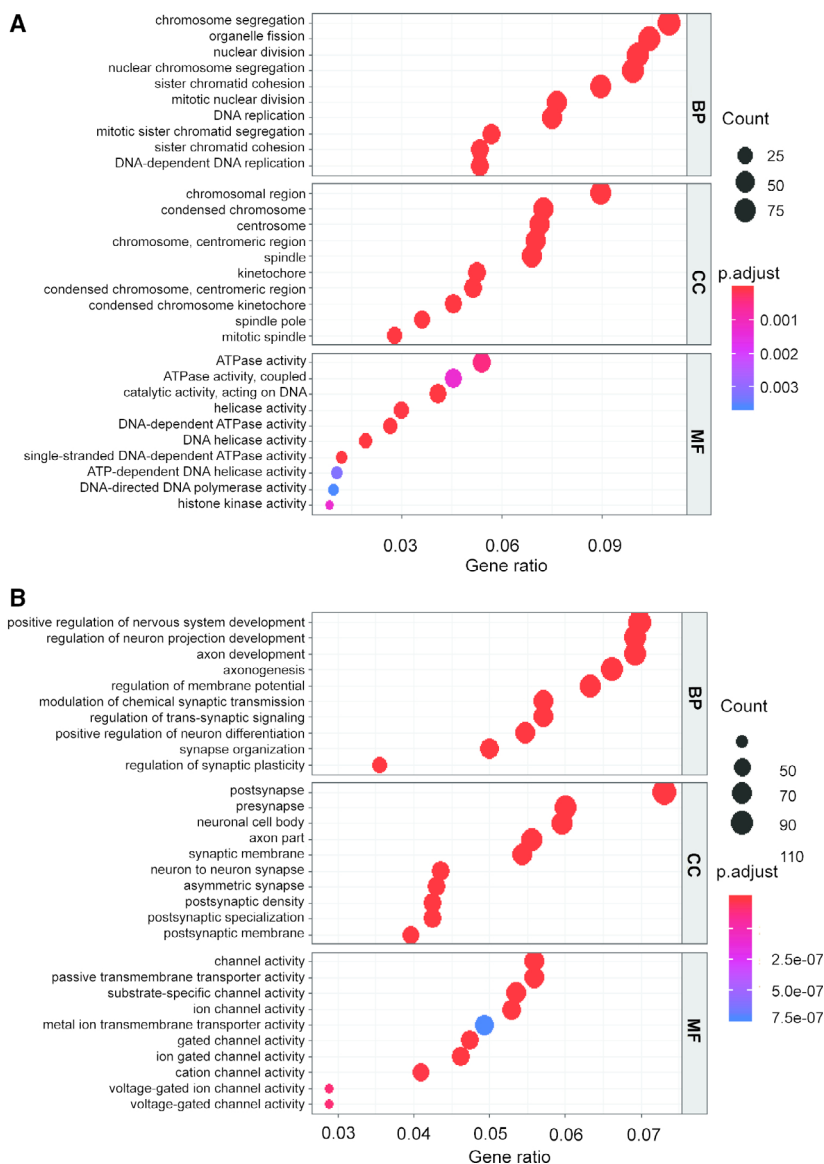
Fig. 5. GO enrichment analysis of key modules. (A) GO enrichment analysis of genes in the brown module (top 10 in BP, CC, and MF are listed). The *y*-axis depicts names of terms in BP, CC, and MF, respectively, and the *x*-axis depicts gene ratio in the module. The circle size represents the count and colours represent the *P*-value; (B) GO enrichment analysis of turquoise module genes (top 10 in BP, CC and MF are listed). CC, cellular component; MF, molecular function.

brown module and *SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3* in the turquoise module.

Validation of hub genes

The expression status of eight hub genes in normal and GBM samples of the other two data sets are shown in Fig. 8B,C, respectively. The results were similar to those of the previous data set. We found differences in IHC staining between the tumour samples and the normal cerebral cortex in the HPA database. It showed the translation expression levels of *NUSAP1*, *SHCBP1* and *KNL1*, which were positively correlated with disease status as they were upregulated in GBM samples. It also showed the translational

expression levels of *SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3*, which were negatively correlated with disease status as they were downregulated in GBM samples (Fig. 9). Unfortunately, there were no related IHC samples of *NUF2* in the database. Overall, these results showed that protein levels were consistent with previously described transcription levels.

Hub gene expression distribution (WHO classification)

The expression levels of these eight hub genes, *NUSAP1, SHCBP1, NUF2, KNL1, SULT4A1, SLC12A5, NAPB* and *GARNL3*, showed significant differences in the WHO classification (Fig. 10). A
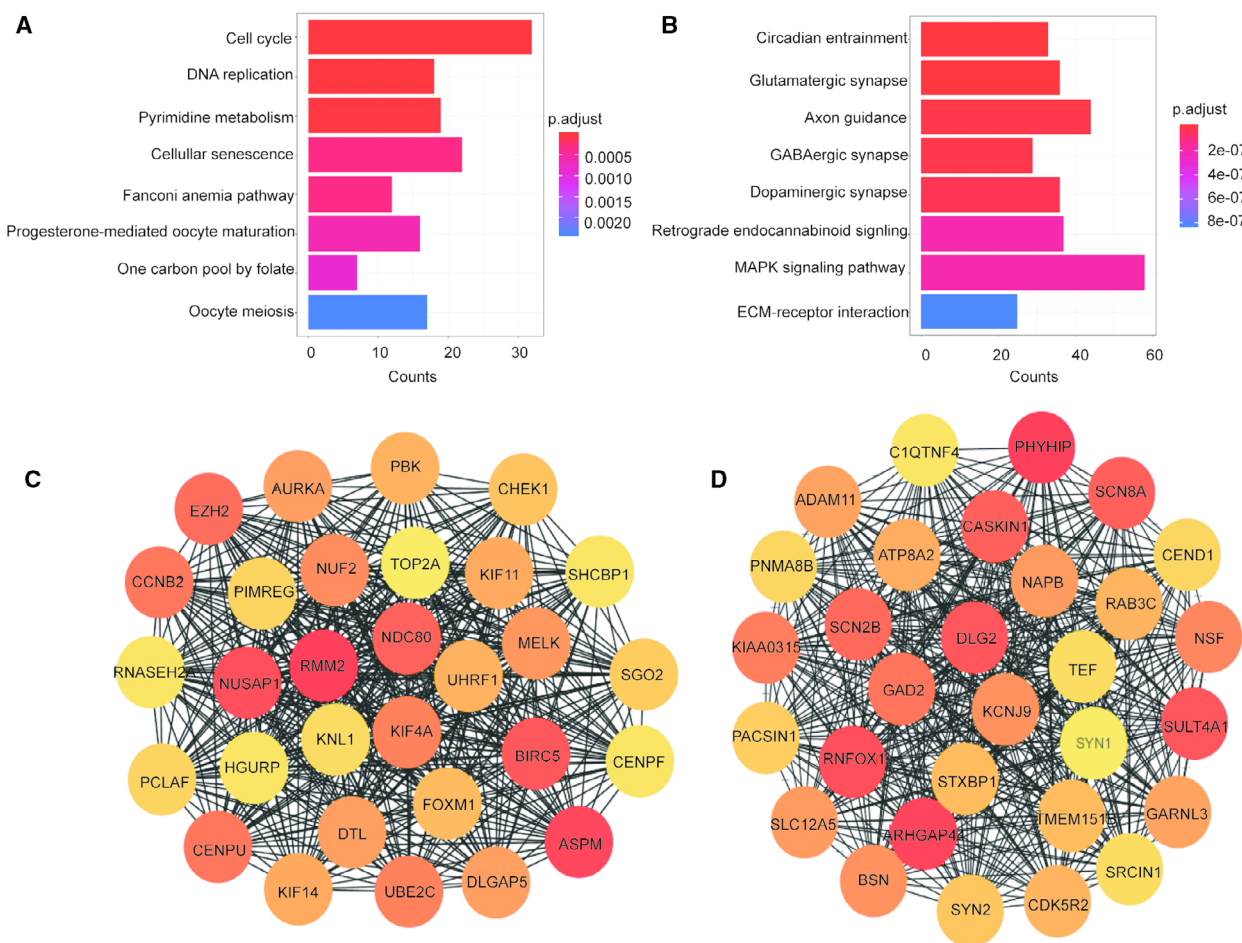
Fig. 6. KEGG pathway enrichment and top 30 genes of key modules. (A) The KEGG pathways of the brown module. The *y*-axis depicts the names of the terms of the pathways, and the *x*-axis represents the count. Colours represent *P*-values. (B) The KEGG pathways of the turquoise module. (C) In the brown module, genes with node degrees in the top 30 are displayed. (D) In the turquoise module, genes with node degrees in the top 30 are displayed. The higher the rank of the genes, the deeper the colour of the genes.

higher WHO classification indicated higher expression levels of *NUSAP1, NUF2, SHCBP1* and *KNL1* genes in the brown module, and lower expression levels of *SULT4A1, SLC12A5, NAPB* and *GARNL3* in the turquoise module; the difference was statistically significant.
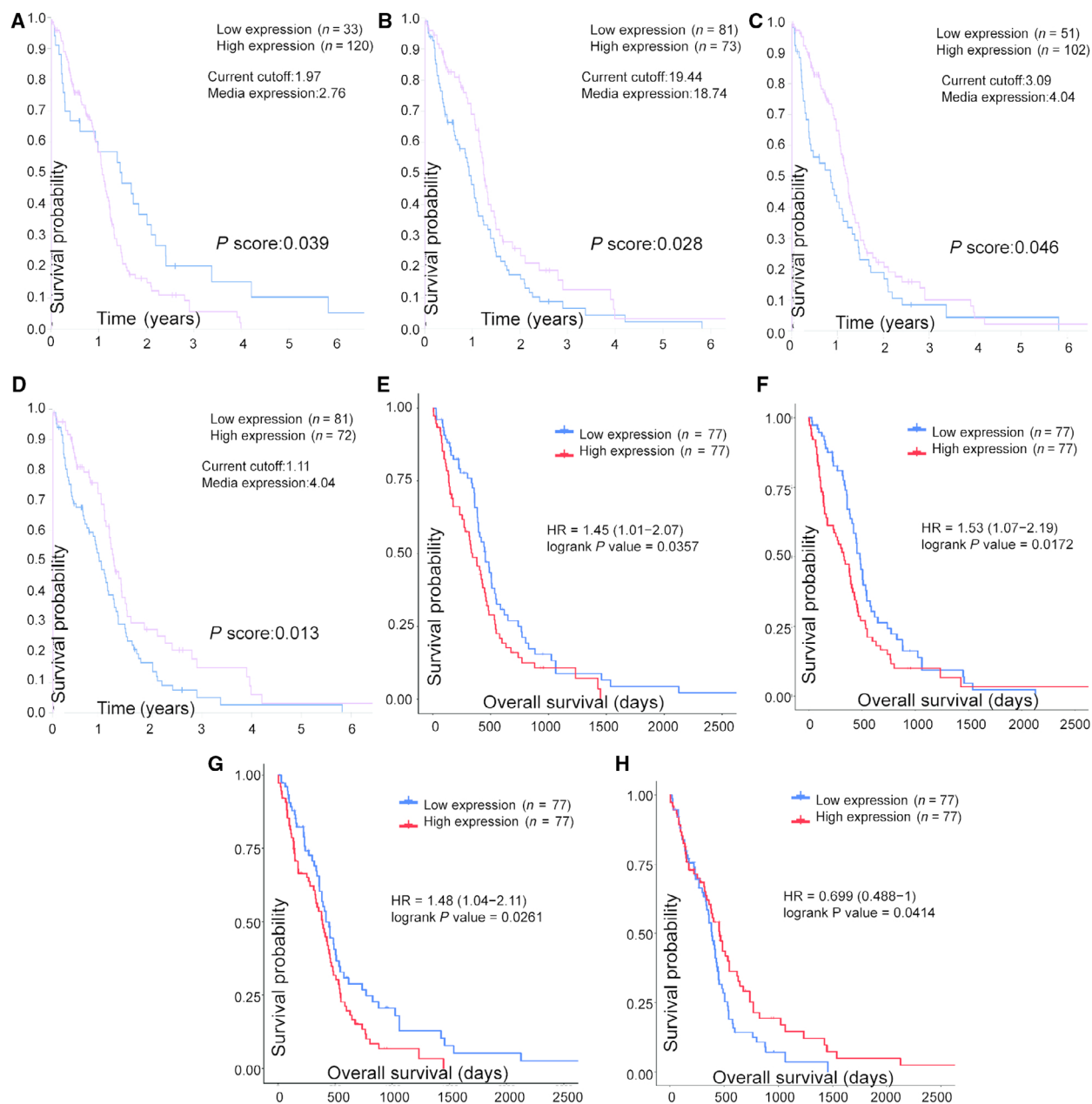
TFs of hub genes

Table 2 shows the prediction of the top 20 TFs of hub genes in the brown and turquoise modules. The number of targets (hub genes) and the number of motifs/tracks for each TF are also listed in the table. E2F4 is the most important TF of *NUSAP1, SHCBP1, NUF2* and *KNL1* in the brown module, and ATF2 is the most important TF of *SULT4A1, SLC12A5, NAPB* and *GARNL3* in the turquoise module.

The role of high expression of Hub gene about the proliferation and clone formation of GBM cells

Cell proliferation and clone formation may be key events in promoting the development of cancer. Therefore, we can analyse the function of these four genes in GBM by analysing the effects of these four genes on the key events of these cancer cells. As shown in Fig. 11A, the results showed that the expression of target genes in siRNA group decreased significantly compared with siNC. We can continue to use these siRNA for functional experimental research. First, we detected the effect of knockdown target gene in the U87 cells by CCK8 method, the results as shown in Fig. 11B that the cell proliferation rate of siRNA group (siNU-SAP1, siSHCBP1, siNUF2 and siKNL1) decreased

Fig. 7. Survival analysis of hub genes. (A–D) In the brown module, genes with top 30 node degrees and significant results of survival analysis in HPA are *SHCBP1*, *NUSAP1*, *NUF2* and *KNL1* ($P < 0.05$ was regarded as significant). (E–H) In the turquoise module, genes with top 30 node degrees and significant results of survival analysis in TCGA are *SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3* ($P < 0.05$ was regarded as significant).

significantly compared with siNC; in other words, the genes of *NUSAP1, SHCBP1, NUF2* and *KNL1* could promote cell proliferation. We further used clone formation experiment to assay the variation of proliferation and cloning ability of U87 cells. As shown in Fig. 11C, the number of clones in the siRNA group (siNUSAP1, siSHCBP1, siNUF2 and siKNL1) decreased significantly. The above results indicated that *NUSAP1,*

SHCBP1, NUF2 and *KNL1* play a key role in promoting the proliferation of GBM, which further confirms the reliability of our previous analysis data.

Discussion

GBM is the most malignant glioma [27]. An in-depth study on the molecular level and mechanism level of
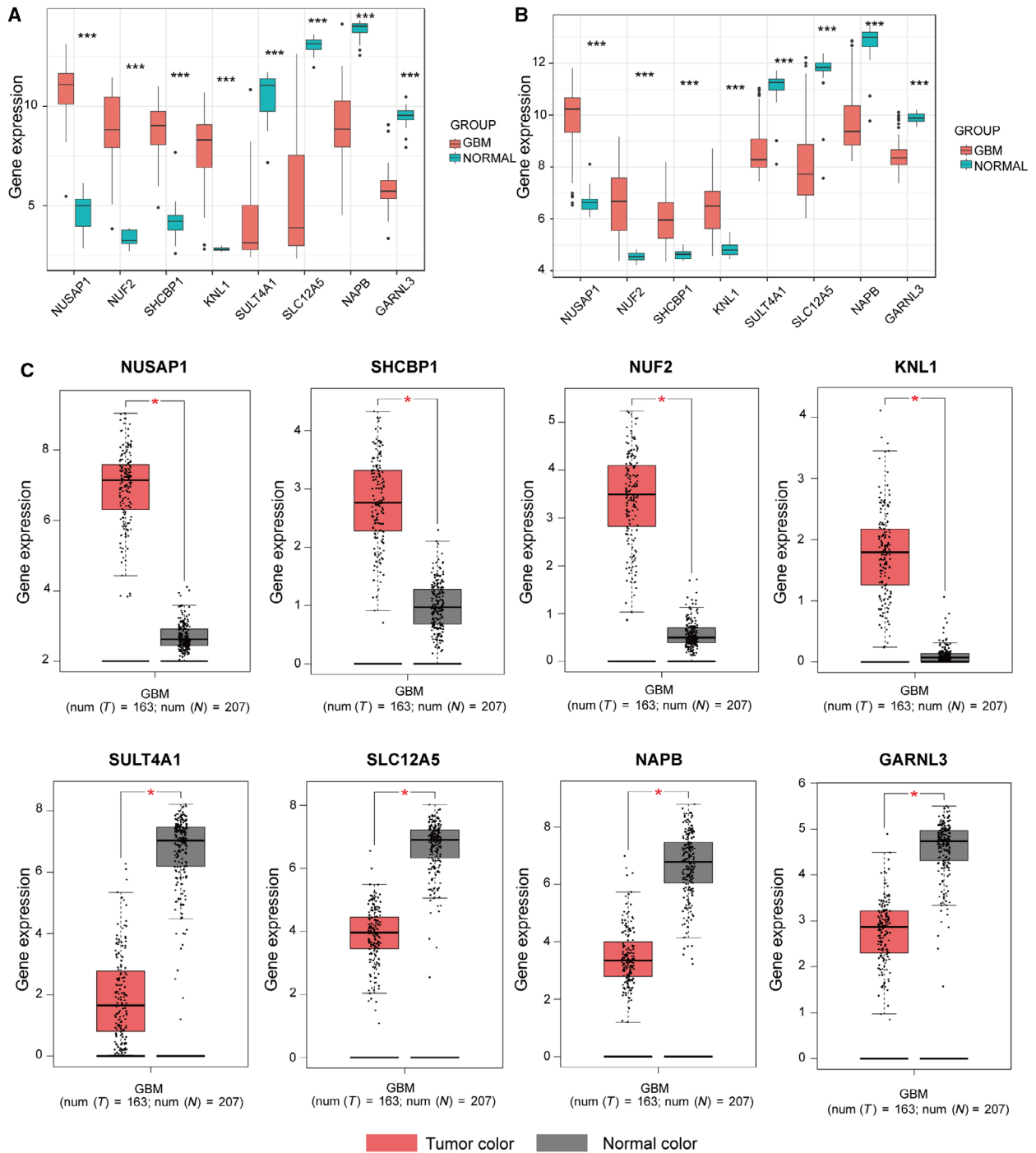
Fig. 8. Screening and validation of hub genes at the transcriptional level. (A) Screening hub genes in GSE50161 (including 34 paediatric GBM samples and 13 normal samples). The expression status of four hub genes (*NUSAP1*, *NUF2*, *SHCBP1* and *KNL1*) was positively correlated with disease status. This was also consistent with the results for the brown module of WGCNA. The expression status of four hub genes (*SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3*) was negatively correlated with disease status. This was consistent with the results in the turquoise module of WGCNA. (B) Validation of eight hub genes in GSE4290 (including 81 paediatric GBM samples and 23 normal samples), and the results were the same as earlier. (C) The differential expression levels of hub genes were demonstrated in TCGA by GEPIA. These results were consistent with the above results. These results fully demonstrated the reliability of our findings. (*t*-test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).
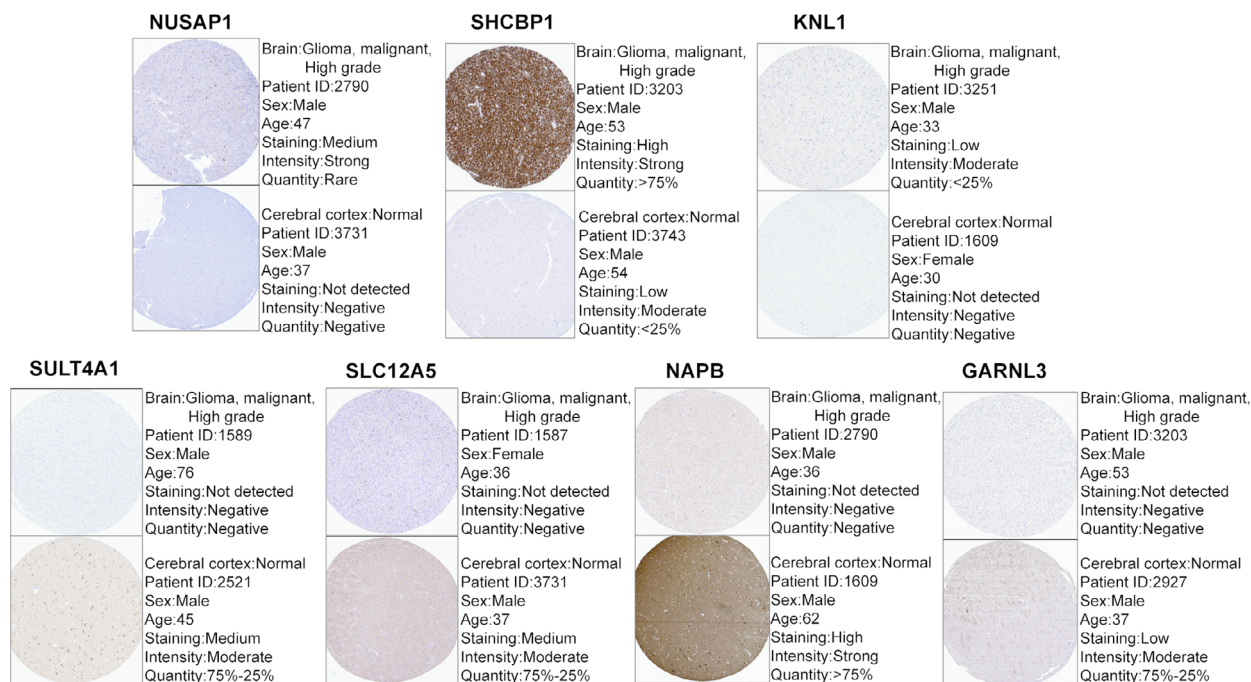
Fig. 9. Validation of hub genes at the translational level by The HPA database. The translational expression levels of *NUSAP1*, *SHCBP1* and *KNL1* in GBM were higher than those in normal tissue. The translational expression levels of *SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3* in GBM were lower than those in normal tissue.

GBM is helpful to find new prevention and treatment targets for GBM. We used WGCNA to explore the tumorigenic factors of GBM and validated the signalling pathways and differential gene levels in two other independent data sets. Finally, two modules (brown module and turquoise module) and eight hub genes (*NUSAP1, NUF2, SHCBP1, KNL1, SULT4A1, SLC12A5, NAPB* and *GARNL3*) in the occurrence of GBM were obtained.

It has been reported that chromosomal structural variation and gene mutations are key factors influencing the occurrence and development of GBM [28,29]. Chromosomal instability (CIN) is one of the characteristics of tumours, and nonmultiple chromosomes cause genomic alterations in cells, resulting in the acquisition of tumour characteristics in normal cells [30]. Interestingly, we found that the BP component of the GO enrichment analysis in the brown module was primarily enriched in chromosome segregation, while the cell component was mainly enriched in the spindle and chromosome, suggesting that chromosomal instability plays an active role in the occurrence of GBM. Consistent with our findings, previous studies demonstrated that the cell cycle process and chromosome instability play an important role in the tumorigenesis of GBM [31,32]. Simultaneously, the KEGG pathway analysis

of the genes in the brown module showed a consistent result. As a malignant tumour in the central nervous system, GBM is characterised by cell cycle disruption and malignant cell proliferation [33], which is in accordance with our results.

In the turquoise module, genes are mainly enriched in positive neuron development and synaptic function in the GO function annotation. The genes are mainly negatively correlated with the status of GBM, which is consistent with the progressive regulation of neurons and synapses to inhibit tumour development. Correspondingly, the main KEGG pathways in the turquoise module-demonstrated enrichment included circadian rhythm entrainment, glutamatergic synapses, axonal guidance, GABAergic synapses, dopaminergic synapses, retrograde endogenous cannabinoid signalling, MAPK signalling pathway and extracellular matrix (ECM)–receptor interaction. Changes in circadian rhythm parameters in the mouse model were found to be associated with glioma diagnosis [34]. Netrin-1, an axonal guidance molecule, has been reported to be associated with invasive and angiogenic phenotypes [35]. The MAPK signalling pathway regulates the normal cell cycle [36], and the ECM–receptor interaction is essential for regulating cell adhesion and cell differentiation [37]. Thus, an abnormal MAPK
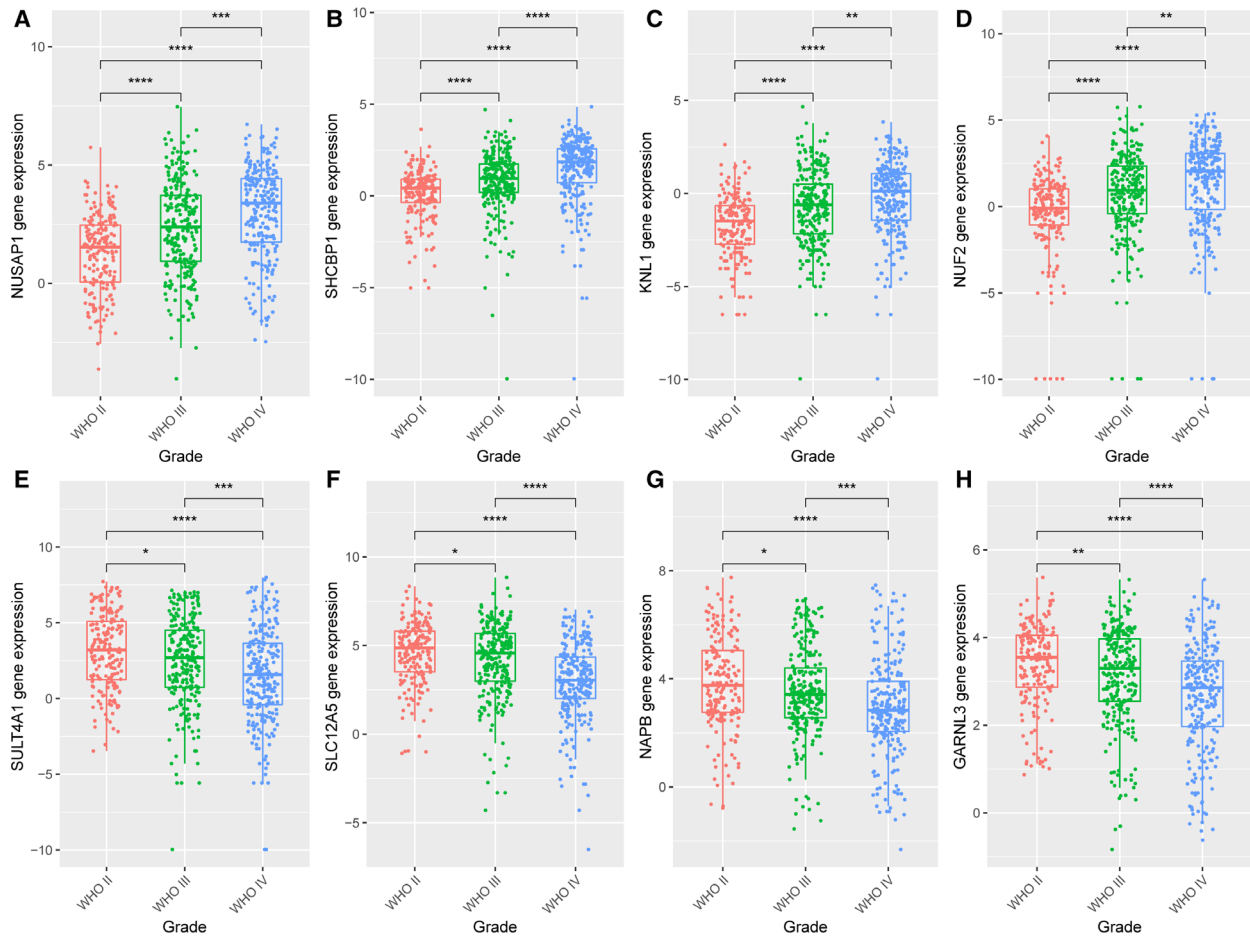
Fig. 10. Hub gene expression distribution in the World Health Organization classification (188 GBM samples in WHO grade II, 255 samples in grade III and 249 samples in grade IV). The expression levels of *NUSAP1, SHCBP1, KNL1* and *NUF2* genes of the brown module and *SULT4A1, SLC12A5, NAPB* and *GARNL3* genes of the turquoise module were significantly different in the different World Health Organization classification. (*t*-test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

signalling pathway and an aberrant ECM–receptor interaction may promote GBM cell proliferation, prevent cell differentiation and increase its invasiveness.

As a pivotal stage of the cell cycle, DNA replication and chromosome segregation are thought to play critical roles in tumorigenesis. The dynamics of microtubule defects and spindle anomalies lead to CIN, which produces a multilayer genomic instability that is common in human cancers [38]. During cell division, the spindle assembly checkpoint (SAC) prevents the separation of repeated chromosomes until each chromosome is properly attached to the spindle device [39]. Error-free chromosome separation relies on a stable connection between the kinetochore and spindle microtubules, ensuring the correct separation of chromosomes during cell division [40]. Interestingly, the four hub genes (*NUSAP1, NUF2, SHCBP1* and *KNL1*) in the brown module are associated with microtubules

and spindles. *NUSAP1* encodes nucleolar and spindle-associated protein 1, a nucleolar spindle-associated protein that plays a role in spindle microtubule organisation [41]. *NUF2* encodes cell division associated 1, responsible for kinetochore–microtubule attachment under normal physiological conditions, and is, therefore, an essential protein for the isolation of sister chromatids during mitosis [42]. *SHCBP1* encodes Src homologous and collagen (SHC) SH2-binding protein 1, a protein essential for midbody organisation and cytokinesis completion [43]. *KNL1* encodes kinetochore scaffold 1, which is involved in microtubule attachment to chromosome centromeres and the activation of spindle checkpoints during mitosis [44]. Many studies have reported that some of the four hub genes in the brown module are cancer-associated genes that play a positive role in tumorigenesis and malignant phenotype in glioma. *NUSAP1* is a prognostic

Table 2. The top 20 TFs of hub genes in key modules. NES, normalised enrichment score.

Hub genes source	TFs	NES	Targets	Motifs/Tracks
Brown module	E2F4	10.219	4	7
	SIN3A	8.997	4	7
	CRX	7.359	2	17
	PAX4	6.075	3	8
	SOX14	5.985	3	3
	ZFY	5.955	3	6
	ALX4	5.107	2	5
	FOXM1	5.032	3	4
	FOXJ2	4.927	2	9
	GTF3C2	4.672	1	7
	KLF5	4.619	2	15
	ZBED1	4.619	2	3
	ZNF423	4.589	1	1
	RXRA	4.574	2	10
	ARID3C	4.567	2	6
	ARID3A	4.544	1	1
	YY1	4.529	1	15
	NF1	4.522	1	3
	E2F3	4.507	1	3
	E1V4	4.499	1	3
Turquoise module	ATF2	6.734	4	44
	REST	6.460	3	11
	ESR1	5.375	4	4
	EBF1	4.895	3	4
	CREB1	4.885	3	5
	HOXB4	4.838	3	3
	OSR1	4.806	2	3
	TEAD4	4.668	2	1
	CLK1	4.574	2	1
	RFX2	4.369	2	30
	GCM1	4.058	3	4
	HOXA5	3.990	3	25
	PRRX1	3.916	2	4
	DBP	3.911	2	2
	ZBTB18	3.742	2	2
	NFKB1	3.516	2	1
	FOXN4	3.155	1	6
	CRX	3.110	1	1
	GATA5	3.089	1	4
	TEAD1	3.073	3	4

factor for gliomas [45], and silencing *NUSAP1* can inhibit GBM cell proliferation both *in vivo* and *in vitro* [46]. Knockdown of *NUF2* by small interfering RNA can inhibit tumour growth and induce apoptosis in human glioma cells [47]. *SHCBP1* is highly expressed in gliomas and promotes proliferation and invasion of glioma cells by activating the NF-κB signalling pathway [48]. *KNL1* is a type of cancer/testis antigen [49] and has been confirmed to be highly expressed in high-grade glioma cell lines and glioma patients, but its detailed function in glioma has not been explored [50]. We predicted the TFs of four hub genes, and some of

them are correlated with gliomas in previous reports. For example, E2F4 is potentially a key transcriptional regulator in GBM that regulates the transcription of multiple genes [51]. KLF5 is involved in GBM angiogenesis by regulating angiogenic factors with G-patch and FHA domain1 (*AGGF1*) expression [52]. FOXM1 activates *MMP2* to enhance the invasiveness of gliomas [53]. In summary, all four hub genes in the brown module are associated with microtubules and spindles, and chromosome separation error is vital in tumorigenesis. It should be noted that *NUF2* participates in the normal function of SAC as part of the nuclear division cycle 80 (NDC80) complex, while *KNL1* is directly involved in the activation of SAC. Meanwhile, *NUSAP1* and *KNL1* directly or indirectly influence the repair of DNA damage [54,55]. Based on our results, we speculate that the abnormally high expression of the four hub genes leads to CIN or DNA repair disorders during the cell division process and normal cells acquire cancer cell characteristics.

We also screened *SULT4A1*, *SLC12A5*, *NAPB* and *GARNL3* in the turquoise module that was downregulated in data sets and TCGA. *SULT4A1* encodes sulfotransferase family 4A member 1, a brain-specific sulfotransferase involved in the metabolism of neurotransmitters [56]. *SLC12A5* encodes a potassium and chloride transporter of the SLC12 family, which is exclusively expressed in the central nervous system and retina [57]. *NAPB* encodes the NSF attachment protein beta, which is preferentially expressed in brain tissues [58]. *GARNL3* encodes a GTPase-activating Rap/RanGAP domain like 3, associated with the positive regulation of GTPase activity. These four hub genes are preferentially expressed in brain tissues, and their dysfunction can lead to neuropsychiatric diseases. For example, *SULT4A1* and *SLC12A5* are associated with schizophrenia [59,60]. Alternatively, spliced isoforms of *NAPB* are associated with autism, and *GARNL3* is linked with intellectual disability [61,62]. There are few studies on the three genes (*SULT4A1*, *NAPB*, *GARNL3*) in GBM, and a few studies have reported the possibility of *SLC12A5* as a biomarker for GBM [63]. However, due to the significant differences in their expression in GBM, further research is needed. For the TF prediction of these genes, we found that some TFs demonstrated a positive regulation of nervous system development and function in previous reports. For example, defects in ATF-2 cause dysplasia and neurological abnormalities in mice [64]. REST is involved in coordinating the neural induction and differentiation processes [65]. ESR1 polymorphisms are associated with the risk of developing dementia [66]. In summary, all four hub genes are associated with the
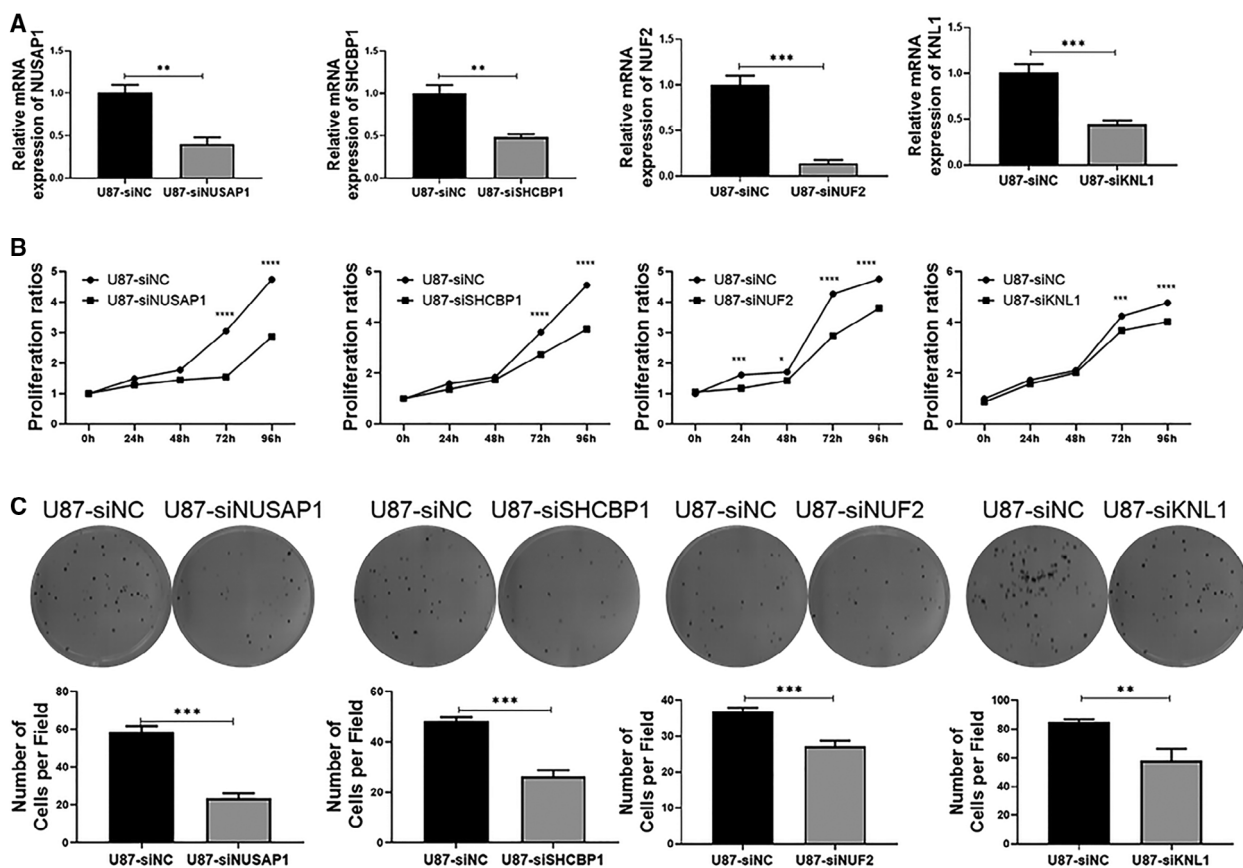
Fig. 11. Knockout *NUSAP1, SHCBP1, NUF2* and *KNL1* gene can decrease the proliferation and cloning of U87 cells. (A) The level of *NUSAP1, SHCBP1, NUF2* and *KNL1* mRNA was determined by RT-PCR after siRNA transfection. (B) The proliferation ability of U87 cells was evaluated by CCK8 method. (C) Clone formation analysis was used to evaluate the cloning ability of U87 cells (unpaired *t*-test, $N = 3$, mean $\pm$ SD, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$).

normal development of the nervous system and can be considered negatively associated with tumorigenesis. The proteins encoded by the above four hub genes play a key role in energy metabolism, which is vital for normal cellular processes, especially in the central nervous system. Therefore, further research into the role and mechanism of these genes in the development of GBM is necessary.

In addition, this study also found that these eight hub genes all had significant differences in the WHO grading of glioma. The expression levels of these genes in the WHO classification were consistent with the results of our study, further revealing the potential of these genes as biomarkers. In general, the above results provide further insights that enhance our understanding of the pathogenesis of GBM at the molecular level.

This study has several limitations. First, it focussed on bioinformatics data mining and analysis. Further mechanistic studies need to be performed to

understand the detailed role of these genes in GBM fully. Second, due to the limitations in data availability, this study did not perform a comprehensive analysis of GBM subtypes based on clinical data. Finally, we validated our findings only in a single Gene Expression Omnibus data set and TCGA database, and use of more data sources would be required for further verification.

Conclusions

This study explored the tumorigenic factors of GBM using WGCNA. We identified two modules (brown and turquoise) and eight hub genes (*NUSAP1, NUF2, SHCBP1, KNL1, SULT4A1, SLC12A5, NAPB* and *GARNL3*) in the occurrence of GBM. The brown module plays a positive role in GBM tumorigenesis, primarily in the cell cycle, chromosome separation and DNA replication. Simultaneously, the turquoise module plays a negative role in tumorigenesis,

primarily in the positive regulation of nervous system development and cell differentiation. These two key modules enhance our understanding of tumorigenic mechanisms in patients with glioblastoma. In addition, these eight hub genes and corresponding TFs may act as prognostic biomarkers and therapeutic targets for GBM.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

The following information was supplied regarding data availability: All the original data in this study were downloaded from the public databases including GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA (https://portal.gdc.cancer.gov/). For the hub genes identification, we used GSE50161 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE50161), GSE4290 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE4290) and GEPIA (http://gepia.cancer-pku.cn/). The translational expression level data sets were from the HPA database (https://www.proteinatlas.org/). As the prediction of TFs, we used the iRegulon plugin (http://iregulon.aertslab.org/) in CY-TOSCAPE. The resource of gene expression distribution level was from the CGGA database (http://www.cgga.org.cn/).

Author contributions

CL conceived and designed the experiments, analysed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft. BP conceived and designed the experiments, authored or reviewed drafts of the paper. LG analysed the data and contributed materials /analysis tools. MZ conceived and designed the experiments, approved the final draft. HS performed the experiments. YY performed the experiments. LL conceived and designed the experiments, and approved the final draft.

References

1 Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C and Barnholtz-Sloan JS (2017) CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro Oncol* **19**, V1–V88.

2 Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P and Ellison DW (2016) The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* **131**, 803–820.

3 Hu B, Wang Q, Wang YA, Hua S, Sauve CG, Ong D, Lan ZD, Chang Q, Ho YW, Monasterio MM *et al.* (2016) Epigenetic activation of WNT5A drives glioblastoma stem cell differentiation and invasive growth. *Cell* **167**, 1281–1295 e18.

4 Haberler C and Wohrer A (2014) Clinical neuropathology practice news 2–2014: ATRX, a new candidate biomarker in gliomas. *Clin Neuropathol* **33**, 108–111.

5 Masui K, Mischel PS and Reifenberger G (2016) Molecular classification of gliomas. *Handb Clin Neurol* **134**, 97–120.

6 Wick W, Dettmer S, Berberich A, Kessler T, Karapanagiotou-Schenkel I, Wick A, Winkler F, Pfaff E, Brors B, Debus J *et al.* (2019) N2M2 (NOA-20) phase I/II trial of molecularly matched targeted therapies plus radiotherapy in patients with newly diagnosed non-MGMT hypermethylated glioblastoma. *Neuro Oncol* **21**, 95–105.

7 Zhang B and Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17.

8 Langfelder P and Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.

9 Su Q, Zhu EC, Qu YL, Wang DY, Qu WW, Zhang CG, Wu T and Gao ZH (2018) Serum level of co-expressed hub miRNAs as diagnostic and prognostic biomarkers for pancreatic ductal adenocarcinoma. *J Cancer* **9**, 3991–3999.

10 Tang J, Kong D, Cui Q, Wang K, Zhang D, Gong Y and Wu G (2018) Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front Oncol* **8**, 374.

11 Liu X, Hu AX, Zhao JL and Chen FL (2017) Identification of key gene modules in human osteosarcoma by co-expression analysis weighted gene co-expression network analysis (WGCNA). *J Cell Biochem* **118**, 3953–3959.

12 Clough E and Barrett T (2016) The gene expression omnibus database. *Methods Mol Biol* **1418**, 93–110.

13 Griesinger AM, Birks DK, Donson AM, Amani V, Hoffman LM, Waziri A, Wang M, Handler MH and

Foreman NK (2013) Characterization of distinct immunophenotypes across pediatric brain tumor types. *J Immunol* **191**, 4880–4888.

14 Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, Passaniti A, Menon J, Walling J, Bailey R *et al.* (2006) Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* **9**, 287–300.

15 Yu G, Wang LG, Han Y and He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287.

16 Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.

17 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504.

18 Tomczak K, Czerwinska P and Wiznerowicz M (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19**, A68–A77.

19 Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419.

20 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47.

21 Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* **45**, W98–W102.

22 Consortium, GT (2013) The genotype-tissue expression (GTEx) project. *Nat Genet* **45**, 580–585.

23 Borggrefe T and Oswald F (2009) The Notch signaling pathway: transcriptional regulation at Notch target genes. *Cell Mol Life Sci* **66**, 1631–1646.

24 Riley T, Sontag E, Chen P and Levine A (2008) Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* **9**, 402–412.

25 Dang CV (2012) MYC on the path to cancer. *Cell* **149**, 22–35.

26 Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D *et al.* (2014) iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol* **10**, e1003731.

27 de Gooijer MC, Guillen Navarro M, Bernards R, Wurdinger T and van Tellingen O (2018) An experimenter's guide to glioblastoma invasion pathways. *Trends Mol Med* **24**, 763–780.

28 Dang L and Su SM (2017) Isocitrate dehydrogenase mutation and (R)-2-hydroxyglutarate: from basic discovery to therapeutics development. *Annu Rev Biochem* **86**, 305–331.

29 Sansal I and Sellers WR (2004) The biology and clinical relevance of the PTEN tumor suppressor pathway. *J Clin Oncol* **22**, 2954–2963.

30 Bakhoum SF and Landau DA (2017) Chromosomal instability as a driver of tumor heterogeneity and evolution. *Cold Spring Harb Perspect Med* **7**, a029611.

31 Vastrad C and Vastrad B (2018) Bioinformatics analysis of gene expression profiles to diagnose crucial and novel genes in glioblastoma multiform. *Pathol Res Pract* **214**, 1395–1461.

32 Waugh MG (2016) Chromosomal instability and phosphoinositide pathway gene signatures in glioblastoma multiforme. *Mol Neurobiol* **53**, 621–630.

33 Cloughesy TF, Cavenee WK and Mischel PS (2014) Glioblastoma: from molecular pathology to targeted treatment. *Annu Rev Pathol* **9**, 1–25.

34 Duhart JM, Brocardo L, Caldart CS, Marpegan L and Golombek DA (2017) Circadian alterations in a murine model of hypothalamic glioma. *Front Physiol* **8**, 864.

35 Sanvoranart T, Supokawej A, Kheolamai P, U-pratya Y, Poungvarin N, Sathornsumetee S and Issaragrisil S (2016) Targeting Netrin-1 in glioblastoma stem-like cells inhibits growth, invasion, and angiogenesis. *Tumour Biol* **37**, 14949–14960.

36 Sun Y, Liu WZ, Liu T, Feng X, Yang N and Zhou HF (2015) Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *J Recept Signal Transduct Res* **35**, 600–604.

37 Bosman FT and Stamenkovic I (2003) Functional structure and composition of the extracellular matrix. *J Pathol* **200**, 423–428.

38 McClelland SE (2017) Role of chromosomal instability in cancer progression. *Endocr Relat Cancer* **24**, T23–T31.

39 Santaguida S and Amon A (2015) Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol* **16**, 473–485.

40 Sacristan C and Kops GJ (2015) Joined at the hip: kinetochores, microtubules, and spindle assembly checkpoint signaling. *Trends Cell Biol* **25**, 21–28.

41 Mills CA, Suzuki A, Arceci A, Mo JY, Duncan A, Salmon ED and Emanuele MJ (2017) Nucleolar and spindle-associated protein 1 (NUSAP1) interacts with a SUMO E3 ligase complex during chromosome segregation. *J Biol Chem* **292**, 17178–17189.

42 Liu D, Ding X, Du J, Cai X, Huang Y, Ward T, Shaw A, Yang Y, Hu R, Jin C *et al.* (2007) Human NUF2 interacts with centromere-associated protein E and is essential for a stable spindle microtubule-kinetochore attachment. *J Biol Chem* **282**, 21415–21424.

43 Asano E, Hasegawa H, Hyodo T, Ito S, Maeda M, Chen D, Takahashi M, Hamaguchi M and Senga T

(2014) SHCBP1 is required for midbody organization and cytokinesis completion. *Cell Cycle* **13**, 2744–2751.

44 Ghongane P, Kapanidou M, Asghar A, Elowe S and Bolanos-Garcia VM (2014) The dynamic protein Knl1 – a kinetochore rendezvous. *J Cell Sci* **127**, 3415–3423.

45 Zhu T, Xie P, Gao YF, Huang MS, Li X, Zhang W, Zhou HH and Liu ZQ (2018) Nucleolar and spindle-associated protein 1 is a tumor grade correlated prognosis marker for glioma patients. *CNS Neurosci Ther* **24**, 178–186.

46 Qian Z, Li Y, Ma J, Xue Y, Xi Y, Hong L, Dai X, Zhang Y, Ji X, Chen Y *et al.* (2018) Prognostic value of NUSAP1 in progression and expansion of glioblastoma multiforme. *J Neurooncol* **140**, 199–208.

47 Huang SK, Qian JX, Yuan BQ, Lin YY, Ye ZX and Huang SS (2014) SiRNA-mediated knockdown against NUF2 suppresses tumor growth and induces cell apoptosis in human glioma cells. *Cell Mol Biol (Noisy-le-grand)* **60**, 30–36.

48 Zhou Y, Tan Z, Chen K, Wu W, Zhu J, Wu G, Cao L, Zhang X, Zeng X, Li J *et al.* (2018) Overexpression of SHCBP1 promotes migration and invasion in gliomas by activating the NF-kappaB signaling pathway. *Mol Carcinog* **57**, 1181–1190.

49 Takimoto M, Wei G, Dosaka-Akita H, Mao P, Kondo S, Sakuragi N, Chiba I, Miura T, Itoh N, Sasao T *et al.* (2002) Frequent expression of new cancer/testis gene D40/AF15q14 in lung cancers of smokers. *Br J Cancer* **86**, 1757–1762.

50 Akiyama Y, Komiyama M, Miyata H, Yagoto M, Ashizawa T, Iizuka A, Oshita C, Kume A, Nogami M, Ito I *et al.* (2014) Novel cancer-testis antigen expression on glioma cell lines derived from high-grade glioma patients. *Oncol Rep* **31**, 1683–1690.

51 Liu Y, Tennant DA, Zhu Z, Heath JK, Yao X and He S (2014) DiME: a scalable disease module identification algorithm with application to glioma progression. *PLoS One* **9**, e86693.

52 Yang C, Zheng J, Xue Y, Yu H, Liu X, Ma J, Liu L, Wang P, Li Z, Cai H *et al.* (2017) The effect of MCM3AP-AS1/miR-211/KLF5/AGGF1 axis regulating glioblastoma angiogenesis. *Front Mol Neurosci* **10**, 437.

53 Wang Z, Zhang S, Siu TL and Huang S (2015) Glioblastoma multiforme formation and EMT: role of FoxM1 transcription factor. *Curr Pharm Des* **21**, 1268–1271.

54 Kotian S, Banerjee T, Lockhart A, Huang K, Catalyurek UV and Parvin JD (2014) NUSAP1 influences the DNA damage response by controlling BRCA1 protein levels. *Cancer Biol Ther* **15**, 533–543.

55 Szczepanski S, Hussain MS, Sur I, Altmuller J, Thiele H, Abdullah U, Waseem SS, Moawia A, Nurnberg G, Noegel AA *et al.* (2016) A novel homozygous splicing mutation of CASC5 causes primary microcephaly in a large Pakistani family. *Hum Genet* **135**, 157–170.

56 Lewis AG and Minchin RF (2009) Lack of exonic sulfotransferase 4A1 mutations in controls and schizophrenia cases. *Psychiatr Genet* **19**, 53–55.

57 Hebert SC, Mount DB and Gamba G (2004) Molecular physiology of cation-coupled Cl- cotransport: the SLC12 family. *Pflugers Arch* **447**, 580–593.

58 Whiteheart SW, Griff IC, Brunner M, Clary DO, Mayer T, Buhrow SA and Rothman JE (1993) SNAP family of NSF attachment proteins includes a brain-specific isoform. *Nature* **362**, 353–355.

59 Meltzer HY, Brennan MD, Woodward ND and Jayathilake K (2008) Association of Sult4A1 SNPs with psychopathology and cognition in patients with schizophrenia or schizoaffective disorder. *Schizophr Res* **106**, 258–264.

60 Hyde TM, Lipska BK, Ali T, Mathew SV, Law AJ, Metitiri OE, Straub RE, Ye T, Colantuoni C, Herman MM *et al.* (2011) Expression of GABA signaling molecules KCC2, NKCC1, and GAD1 in cortical development and schizophrenia. *J Neurosci* **31**, 11088–11095.

61 Santos-Cortez RLP, Khan V, Khan FS, Mughal ZU, Chakchouk I, Lee K, Rasheed M, Hamza R, Acharya A, Ullah E *et al.* (2018) Novel candidate genes and variants underlying autosomal recessive neurodevelopmental disorders with intellectual disability. *Hum Genet* **137**, 735–752.

62 Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA *et al.* (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun* **5**, 3650.

63 Yang J, Wang L, Xu Z, Wu L, Liu B, Wang J, Tian D, Xiong X and Chen Q (2020) Integrated analysis to evaluate the prognostic value of signature mRNAs in glioblastoma multiforme. *Front Genet* **11**, 253.

64 Reimold AM, Grusby MJ, Kosaras B, Fries JW, Mori R, Maniwa S, Clauss IM, Collins T, Sidman RL, Glimcher MJ *et al.* (1996) Chondrodysplasia and neurological abnormalities in ATF-2-deficient mice. *Nature* **379**, 262–265.

65 Soldati C, Bithell A, Johnston C, Wong KY, Teng SW, Beglopoulos V, Stanton LW and Buckley NJ (2012) Repressor element 1 silencing transcription factor couples loss of pluripotency with neural induction and neural differentiation. *Stem Cells* **30**, 425–434.

66 Sundermann EE, Maki PM and Bishop JR (2010) A review of estrogen receptor alpha gene (ESR1) polymorphisms, mood, and cognition. *Menopause* **17**, 874–886.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Visualizing DEGs in GSE4290 and KEGG pathway enrichment. Note: (A) The number of genes that had an absolute fold change greater than 1.5 and a P-value less than 0.05. (B) Heat map hierarchical clustering showed that the top 100 DEGs in GBM groups compared with control groups. (C) KEGG pathway enrichment analysis of the DEGs in GSE4290 (P-value less than 0.001 are shown). Up-regulated pathways are labelled red, and down-regulated pathways are labelled blue. Abbreviations: DEGs, different expression genes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Fig. S2. GO enrichment analysis of DEGs in GSE4290. Note: GO enrichment analysis of up-regulated genes; (B) GO enrichment analysis of down-regulated genes.. Top 10 in BP, CC and MF are listed. Abbreviations: GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function; ATP, adenosine triphosphate.