

Received: 2018.10.13

Accepted: 2019.01.22

Published: 2019.05.19

Use of Genome-Scale Integrated Analysis to Identify Key Genes and Potential Molecular Mechanisms in Recurrence of Lower-Grade Brain Glioma

Authors' Contribution:

Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

AEG 1 **Teng Deng***
BCDEG 2 **Yi-Zhen Gong***
BC 3 **Xiang-Kun Wang**
C 3 **Xi-Wen Liao**
BD 3 **Ke-Tuan Huang**
BC 3 **Guang-Zhi Zhu**
DF 1 **Hai-Nan Chen**
BF 1 **Fang-Zhou Guo**
A 1 **Li-Gen Mo**
AG 4 **Le-Qun Li**

1 Department of Neurosurgery, Affiliated Tumor Hospital of Guangxi Medical University, Nanning, Guangxi, P.R. China
2 Department of Evidence-Based Medicine, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, P.R. China
3 Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, P.R. China
4 Department of Hepatobiliary Surgery, Affiliated Tumor Hospital of Guangxi Medical University, Nanning, Guangxi, P.R. China

* Teng Deng and Yi-Zhen Gong are equal contributors

Corresponding Author: Le-Qun Li, e-mail: lilequnxmu@163.com

Source of support: This research was supported by the Guangxi Natural Science Foundation under Grant No. 2015GXNSFAA139160. This work was also supported in part by the Innovation Project of Guangxi Graduate Education (JGY2018037) and the 2018 Innovation Project of Guangxi Graduate Education (YCBZ2018036)

Background: The aim of this study was to identify gene signals for lower-grade glioma (LGG) and to assess their potential as recurrence biomarkers.

Material/Methods: An LGG-related mRNA sequencing dataset was downloaded from The Cancer Genome Atlas (TCGA) Informix. Multiple bioinformatics analysis methods were used to identify key genes and potential molecular mechanisms in recurrence of LGG.

Results: A total of 326 differentially-expressed genes (DEGs), were identified from 511 primary LGG tumor and 18 recurrent samples. Gene ontology (GO) analysis revealed that the DEGs were implicated in cell differentiation, neuron differentiation, negative regulation of neuron differentiation, and cell proliferation in the forebrain. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database suggests that DEGs are associated with proteoglycans in cancer, the Wnt signaling pathway, ECM-receptor interaction, the PI3K-Akt signaling pathway, transcriptional deregulation in cancer, and the Hippo signaling pathway. The hub DEGs in the protein-protein interaction network are apolipoprotein A2 (APOA2), collagen type III alpha 1 chain (COL3A1), collagen type I alpha 1 chain (COL1A1), tyrosinase (TYR), collagen type I alpha 2 chain (COL1A2), neurotensin (NTS), collagen type V alpha 1 chain (COL5A1), poly(A) polymerase beta (PAPOLB), insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1), and anomalous homeobox (ANHX). GSEA revealed that the following biological processes may associated with LGG recurrence: cell cycle, DNA replication and repair, regulation of apoptosis, neuronal differentiation, and Wnt signaling pathway.

Conclusions: Our study demonstrated that hub DEGs may assist in the molecular understanding of LGG recurrence. These findings still need further molecular studies to identify the assignment of DEGs in LGG.

MeSH Keywords: **Biological Markers • Brain Neoplasms • Glioma**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/913602>

 3095

 2

 7

 59



Background

Glioma is the most frequent primary brain tumor caused by the carcinogenesis of glial cells in the brain. According to the World Health Organization (WHO), gliomas can be differentiated into 4 grades (I–IV). Lower-grade gliomas (LGG) ([WHO] grades II and III) are aggressive tumors that occur most commonly in the hemi-cerebrum of adults and include astrocytomas and oligodendrogliomas [1]. Due to the aggressive nature of glioma, it is difficult to remove the neoplasm completely by neurosurgical surgery, and the residual tumor is the leading cause of recurrence and disease progression [2]. Currently, the main treatments of gliomas are surgery, radiotherapy, chemotherapy, and targeted therapy [3,4]. Currently, new findings are being made in both basic and translational research, including disordered tumor proliferation signaling pathways, anti-angiogenic therapy, classifying tumor grade, and predicting overall survival [2,5–8]. To gain further understanding of this heterogeneous disease, we conducted an integrative genome-wide analysis of 511 LGGs from adults, using bioinformatics analysis. By comparing the differentially expressed genes (DEGs) between the primary and recurrent LGG tumor samples, we used bioinformatics methods to screen the hub DEGs and explore the biological processes of DEGs. It is helpful to understand and develop targeted therapeutic drugs for LGG recurrence.

The Cancer Genome Atlas (TCGA) is a catalogue of genetic mutations responsible for cancer (including LGG and glioblastoma), using genome sequencing and bioinformatics [9,10]. High-throughput genome analysis techniques (including a minimum of 6000 candidate genes and microRNA sequences) from TCGA lead to a superior perception of the genetic basis of this disease. There were 511 LGG patients providing 529 samples (282 males, 228 females, and 1 patient with sex unknown), including 511 primary tumors and 18 recurrent tumors to facilitate information sharing across cancer genomic research. The TCGA database collected mRNA gene expression profiles in LGG and the profiles of LGG; first occurrence and recurrence were compared to identify differentially expressed genes (DEGs) and DEGs were further sorted by gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) analysis. Next, gene co-expression and protein–protein interaction (PPI) networks were investigated by the Search Tool for Recurring Instances of Neighboring Genes (STRING), and module analysis of the DEGs was proceeded by Molecular Complex Detection (MCODE) in Cytoscape software. By studying the biological functions and pathways of these DEGs, we can learn more about the pathogenesis, occurrence, and development of LGG and improve the diagnosis and treatment of this disease in the immediate future.

Material and Methods

Data source

This was a study of publicly available data from The Cancer Genome Atlas (TCGA). RNA sequencing information of LGG was downloaded from TCGA (<https://cancergenome.nih.gov>; accessed Nov 4, 2017) [2]. A total of 529 LGG samples, including 511 primary tumor samples and 18 recurrent samples, and RNA-sequencing (RNA-seq) datasets were accessed. All these LGG biospecimens were collected from patients diagnosed with diffuse glioma undergoing surgical resection and who had received no prior treatment for their disease (chemotherapy or radiotherapy). All the data were sourced from the TCGA website and used in accordance with the TCGA data access policies and publication guidelines (<https://cancergenome.nih.gov/publications/publicationguidelines>). Therefore, additional ethics committee approval was not necessary for the present study.

DEGs

We initially made a comparison of the expression of mRNAs between LGG primary tumors and recurrent LGG by differential expression analysis. Then, we screened the differentially expressed mRNAs by edgeR [11,12] in the R platform. According to DEG scanning, the genes that met a criterion with false discovery rate (FDR) <0.05 and fold control (FC) ≥ 2.0 [13–15] were identified as DEGs. Next, we plotted the heat map of these DEGs by the *gplots* package in the R platform.

GO enrichment analysis and KEGG pathway analysis of DEGs

For the purpose of analyzing the DEGs at the biological function level, GO enrichment analysis and KEGG pathway analysis were performed by using the Database for Annotation, Visualization, and Integration Discovery (DAVID, <https://david.ncifcrf.gov>, version 6.8) online tool. Results of GO and KEGG pathway analysis fulfilling a nominal P-value <0.05 were considered statistically significant [16,17]. The P-value of the GO enrichment analysis of the DEGs was used for locating the most likely related GO term and confirming gene function in biological models [18,19]. The KEGG pathway analysis of DEGs can facilitate our understanding of the metabolic pathways altered under experimental conditions, particularly in mechanistic studies.

Protein–protein interaction network and module analysis

The online STRING database (<https://string-db.org>; accessed September 18, 2017), which was designed to construct a crucial PPI network, evaluates the interactive relationships among DEGs and examines the physical and functional association

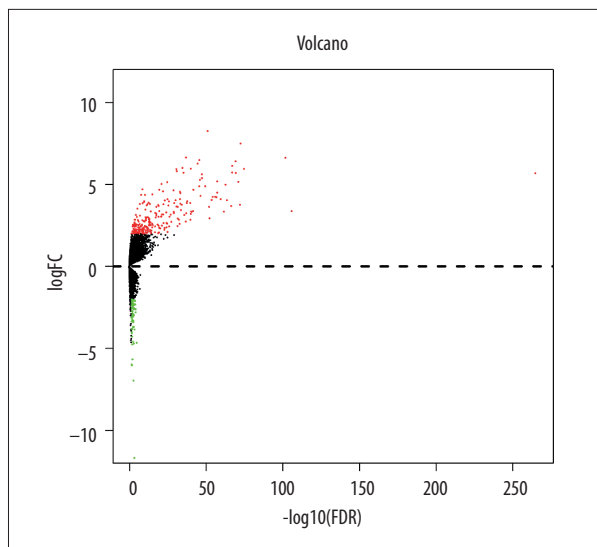


Figure 1. Volcano plot of the DEGs between LGG primary and recurrence tumor tissues. Red: upregulation; green: downregulation; black: non-differentially-expressed genes. DEGs – differentially expressed genes; LGG – lower-grade glioma

between oncogenesis-related DEGs [20]. Therefore, we assessed the PPI network of the DEGs with STRING. Only co-expressional validated interactions with a combined score >0.15 were considered statistically significant. Subsequently, we mapped PPI networks of DEGs with Cytoscape software, version 3.5.1, a bioinformatics integration platform, and used MCODE to construct modules of protein–protein interactions with the annotations as follows: scores >6 and nodes >7 in Cytoscape. Moreover, GO enrichment and KEGG pathway analysis were performed in the modules to assess for potential oncogenesis-related DEGs, with a P-value <0.05 considered statistically significant.

Gene set enrichment analysis (GSEA)

To further explore the potential molecular mechanisms of LGG recurrence, we also performed a bioinformatics analysis, GSEA, to investigate the differences in pathways and functions between LGG primary and recurrence tumor tissues. The parameter of permutations was set at 1000. The nominal P-value of enrichment results cut-off was 0.05, while the false discovery rate (FDR) cut-off was 0.25 [21–23].

Statistical analysis

In order to combat error in multiple comparisons, we conducted multiple testing with the Benjamini-Hochberg procedure [13–15] in edgeR to control FDR. A value of $P < 0.05$ was considered statistically significant. All statistical analyses were

carried out with SPSS version 20.0 (IBM Corporation, Armonk, NY, USA) and R3.3.1.

Results

Identification of DEGs

In total, 511 LGG tumor and 18 recurrent specimens were screened for DEGs. The *edgeR* package identified 326 differentially expressed mRNAs, $|\log_2 FC| \geq 2$, and P-value <0.05 of which 248 were upregulated and 78 were downregulated and showed a consistent direction of differential expression (Figure 1). A heat map of 326 DEGs is shown in Supplementary Figure 1.

GO enrichment analysis and KEGG analysis of DEGs

DAVID was utilized to determine the specific GO terms and KEGG pathways of the DEGs. According to the GO analysis, DEGs were generally enriched in biological process (BP), molecular function (MF), and GO cell components (CC), including neuron differentiation, cell differentiation, negative regulation of neuron differentiation, and cell proliferation in the forebrain (Figure 2A). In the KEGG pathway analysis, DEGs were significantly enriched in the wnt signaling pathway, ECM-receptor interaction, signaling pathways regulating pluripotency of stem cells, PI3K-Akt signaling pathway, proteoglycans in cancer, transcriptional misregulation in cancer and Hippo signaling pathway (Figure 2B).

Identification of hub genes and MCODE analysis from the PPI network

STRING analysis was performed to identify the hub genes and hub nodes with the highest degrees of interaction. These included apolipoprotein A2 (APOA2), collagen type III alpha 1 chain (COL3A1), collagen type I alpha 1 chain (COL1A1), tyrosinase (TYR), collagen type I alpha 2 chain (COL1A2), neurotensin (NTS), collagen type V alpha 1 chain (COL5A1), poly(A) polymerase beta (PAPOLB), insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1), and anomalous homeobox (ANHX). The Column chart of the mRNA expression level between LGG primary tumor and those recurrent of the top 10 rank hub genes are shown in Figure 3 and Table 1. Among them, APOA2 gave rise to 14, the highest node degree. Figure 4 shows the PPI network of the DEGs. Moreover, the overall number of 326 DEGs were analyzed by the MCODE plugin. We selected the top 3 significant modules and analyzed their functional annotation (Figure 5A–5C). GO term enrichment and KEGG pathway analysis revealed that only the genes of module II were mainly associated with ion binding, homeostasis, and regulation of the lipid metabolic process (Table 2). There were no statistically significant differences between module I and III.

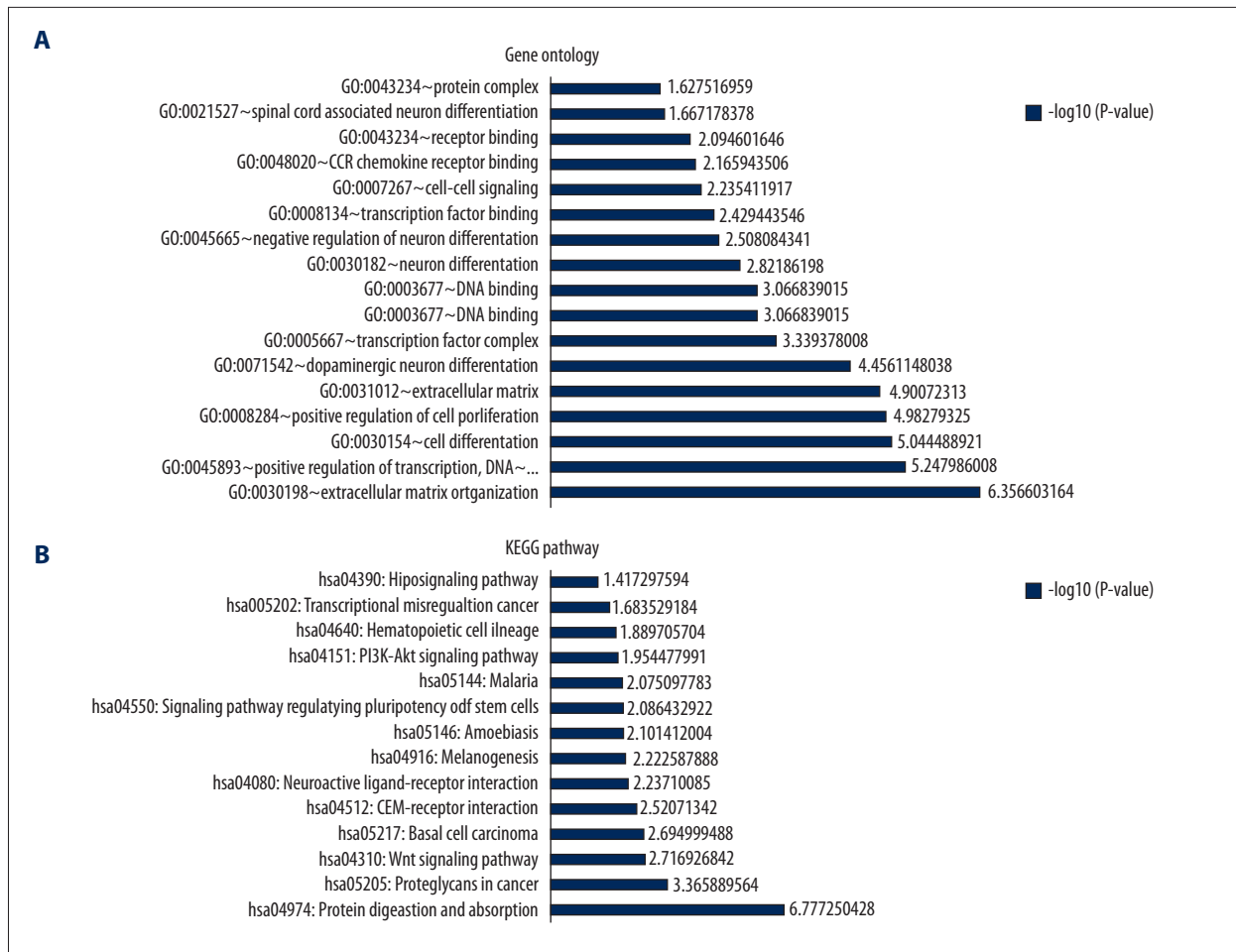


Figure 2. Functional assessment for DEGs between LGG primary and recurrence tumor tissues. (A) GO term enrichment results; (B) KEGG enrichment results. DEGs – differentially-expressed genes; LGG – lower-grade glioma; GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes

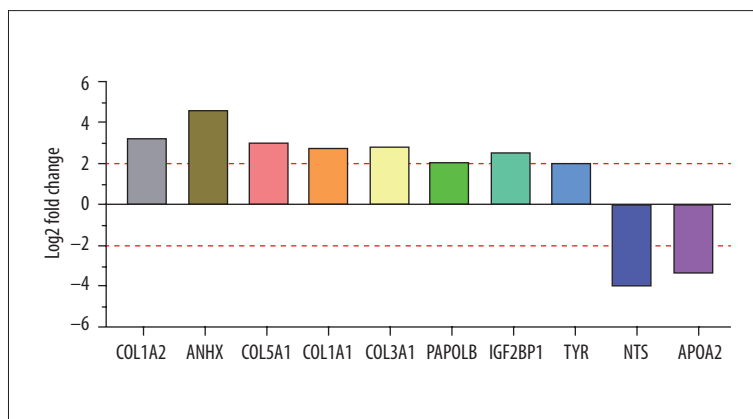


Figure 3. Column chart of the top 10 rank hub genes mRNA expression level between LGG primary and recurrence tumor tissues. APOA2 – apolipoprotein A2; COL3A1 – collagen type III alpha 1 chain; COL1A1 – collagen type I alpha 1 chain; TYR – tyrosinase; COL1A2 – collagen type I alpha 2 chain; NTS – neurotensin; COL5A1 – collagen type V alpha 1 chain; PAPOLB – poly(A) polymerase beta; IGF2BP1 – insulin-like growth factor 2 mRNA-binding protein 1; ANHX – anomalous homeobox; LGG – lower-grade glioma

GSEA

GSEA analysis between the primary and recurrent tumor tissues using the c2 reference gene set suggest that LGG recurrence is significantly associated with changes in cell cycle, DNA

replication, apoptosis by CDKN1A via TP53, DNA repair, regulation of apoptosis, neuronal differentiation, and Wnt signaling pathway genes (Figure 6A–6P). However, no results reached the statistical significance in the GSEA analysis using the c5 reference gene set.

Table 1. Fold change list of the top 10 hub differentially expressed genes between LGG primary and recurrence tumor tissues.

ID	Log2FC	P value	FDR	Change
NTS	-3.998653652	0.004369524	0.020863775	Down-regulation
APOA2	-3.357158871	0.009228557	0.036982495	Down-regulation
PAPOLB	2.01120252	1.43E-12	9.93E-11	Up-regulation
TYR	2.058023707	0.000339247	0.002609601	Up-regulation
IGF2BP1	2.558024781	1.10E-10	5.33E-09	Up-regulation
COL3A1	2.81332624	4.01E-13	3.07E-11	Up-regulation
COL1A1	2.816660818	1.61E-13	1.31E-11	Up-regulation
COL5A1	3.047914457	8.18E-24	1.51E-21	Up-regulation
COL1A2	3.249927208	3.21E-32	8.57E-30	Up-regulation
ANHX	4.581419911	5.44E-25	1.04E-22	Up-regulation

FC – fold change; APOA2 – apolipoprotein A2; COL3A1 – collagen type III alpha 1 chain; COL1A1 – collagen type I alpha 1 chain; TYR – tyrosinase; COL1A2 – collagen type I alpha 2 chain; NTS – neurotensin; COL5A1 – collagen type V alpha 1 chain; PAPOLB – poly(A) polymerase beta; IGF2BP1 – insulin like growth factor 2 mRNA binding protein 1; ANHX – anomalous homeobox; LGG – lower grade glioma; FDR – false discovery rate.

Discussion

Numerous studies have shown that genetic status is more likely to reflect disease subtypes than is histology [2,5,6]. An integrated, multiplatform genomics system was utilized to investigate the biological basics of adult LGG. The present study compared LGG tumor and 18 recurrent specimens to identify DEGs and then used gene set enrichment analysis (GSEA) to investigate and interpret differences in gene expression. The *edgeR* package identified 326 differentially expressed mRNAs, of which 248 were upregulated and 78 were downregulated, showing a consistent direction of differential expression. We then performed GO enrichment and KEGG analysis to consider the interactions of these DEGs. We identified a number of key genes that may provide new insight into LGG therapies.

GO enrichment analysis revealed that the DEGs were mostly involved in cell differentiation, positive regulation of cell proliferation, neuron differentiation, positive regulation of epithelial cell proliferation, and cell-cell signaling. CD109 is a critical regulator of the progression of LGG. It has been reported that CD109-positive brain tumor stem cells exert a proliferative effect on cell differentiation of glioma to promote their survival and resistance to conventional therapies [24]. Additionally, ciliary neurotrophic factor receptor alpha subunit (CNTFRalpha) and CNTF have considerable roles in neuronal survival, neuroglial differentiation and glioma growth. Hypomethylation leading to CNTFRalpha upregulation has been proposed to be involved in glioma growth regulation [25].

In addition, the results of the KEGG analysis mainly included DEGs involved in the Wnt signaling pathway, PI3K-Akt signaling

pathway, ECM-receptor interaction, and neuroactive ligand-receptor interactions. Wnt signaling pathways manage proliferation, motility, and survival in many human cell types. Indeed, the dickkopf 1 (DKK1) gene codes for a secreted Wnt inhibitory factor and functions as a pro-apoptotic factor in glioma cells [26]. By means of variable activation of the Wnt/beta-catenin signaling pathway, the (pro)renin receptor (PRR) plays a crucial role in development of glioma [27], with poor prognostic implications for the final outcome of brain glioma via SFRP1 inhibition and Wnt activation reported to contribute to the infiltrative glioma phenotype at early stages of disease progression [28]. Radio-proteomics analysis has shown that the phenotypic consequences of molecular aberrations in LGGs are correlated with expression of PI3K/Akt signaling pathways [29]. Moreover, the tumorigenesis of pleomorphic glioblastoma may increase with the enhancement aggregation of affected genes on specific molecular pathways (e.g., focal adhesion or ECM-receptor interaction) [30]. An integrated analysis of mutated genes from India identified GBMs with defective neuroactive ligand-receptor interaction pathways were associated with a significantly worse prognosis [31]. In addition, the findings from GSEA also verify the KEGG enrichment results showing that, compared to LGG primary, recurrence tumor tissues were significantly enriched in the Wnt signaling pathway. Other enrichment results of GSEA also suggested that LGG recurrence may involve the basic status of cells, such as cell cycle, differentiation, DNA repair, and regulation of apoptosis. Therefore, we conclude that the recurrent mechanism of LGG involves the basic state of neurons, such as neuronal differentiation, cell cycle, DNA repair, and apoptosis.

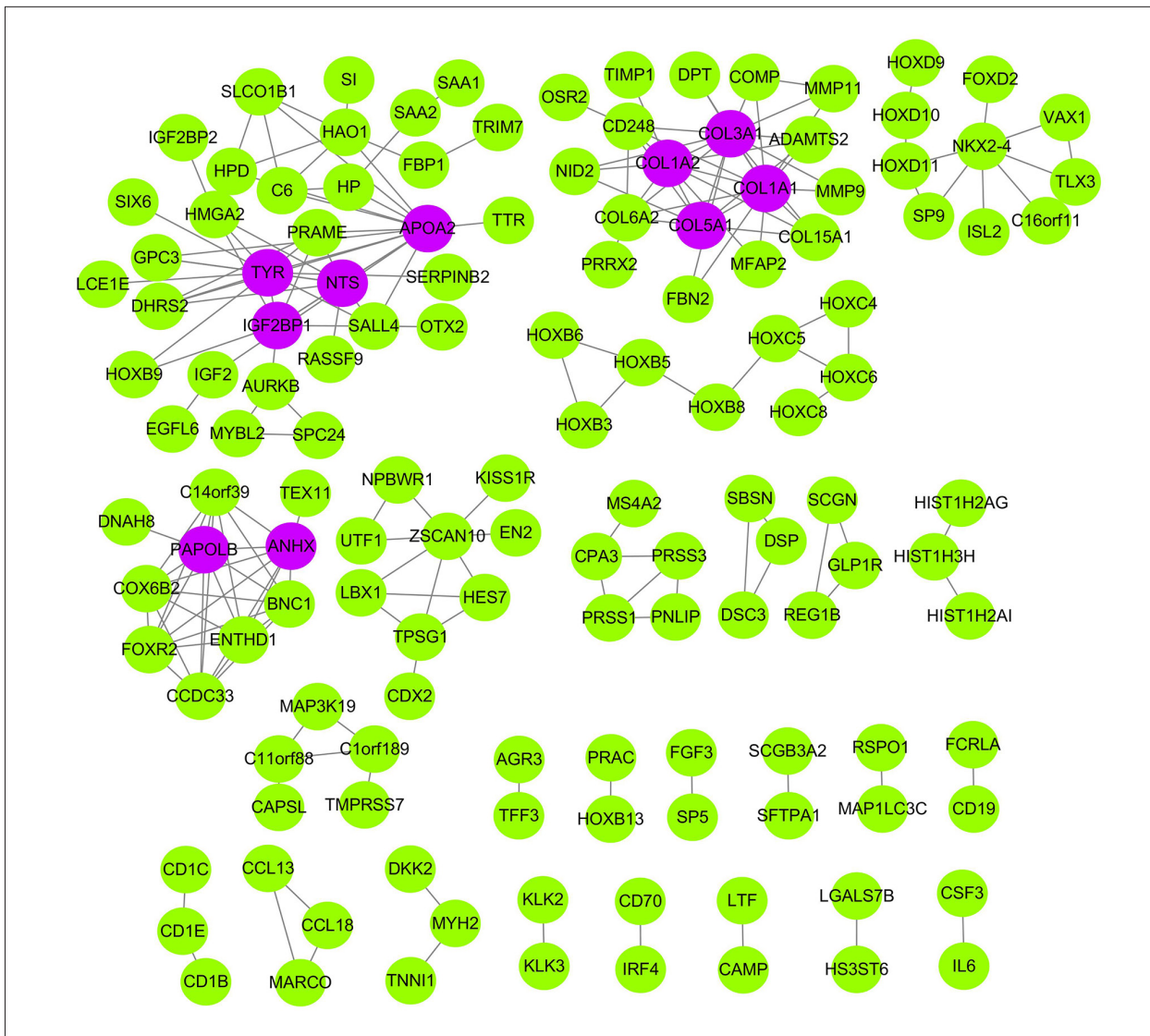


Figure 4. Protein-protein interaction networks of DEGs between LGG primary and recurrence tumor tissues. Green text: DEGs; Purple text: 10 hub genes of DEGs in the PPI networks. DEGs – differentially-expressed genes; LGG – lower-grade glioma; PPI – protein-protein interaction.

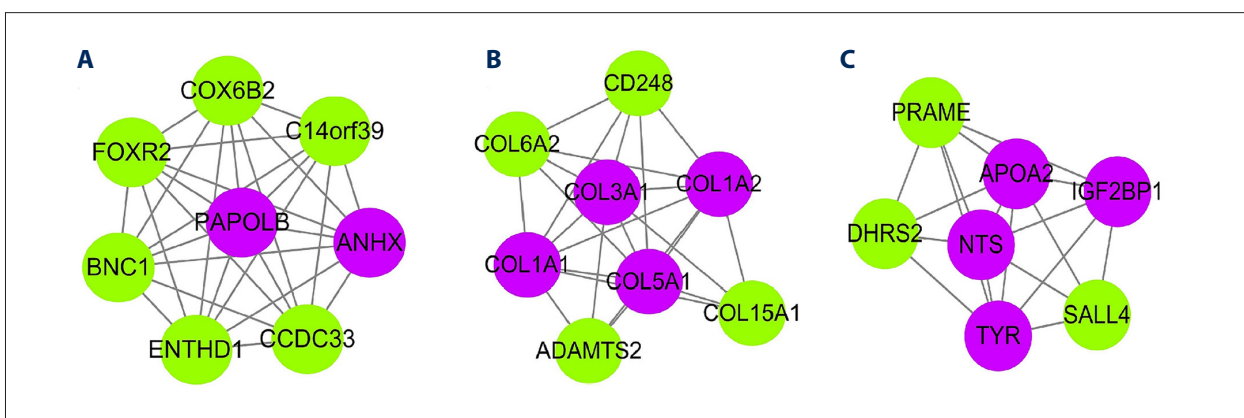


Figure 5. Top 3 modules from the PPI networks. (A) Module 1; (B) module 2; (C) module 3. PPI – protein-protein interaction

Table 2. The result of the top module II using GO term enrichment and KEGG pathway analysis.

ID	Term	P value	Genes
GO			
GO:0030574	Collagen catabolic process	1.68E-14	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, ADAMTS2, COL5A1
GO:0005581	Collagen trimer	6.12E-11	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, COL5A1
GO:0030199	Collagen fibril organization	8.65E-10	COL3A1, COL1A2, COL1A1, ADAMTS2, COL5A1
GO:0005788	Endoplasmic reticulum lumen	2.54E-09	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, COL5A1
GO:0048407	Platelet-derived growth factor binding	7.19E-09	COL3A1, COL1A2, COL1A1, COL5A1
GO:0005201	Extracellular matrix structural constituent	7.86E-09	COL3A1, COL1A2, COL15A1, COL1A1, COL5A1
GO:0005578	Proteinaceous extracellular matrix	1.36E-08	CD248, COL1A2, COL6A2, COL15A1, ADAMTS2, COL5A1
GO:0031012	Extracellular matrix	2.24E-08	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, COL5A1
GO:0030198	Extracellular matrix organization	6.13E-07	COL3A1, COL1A2, COL6A2, COL1A1, COL5A1
GO:0005576	Extracellular region	3.05E-06	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, ADAMTS2, COL5A1
GO:0070208	Protein heterotrimerization	1.35E-05	COL1A2, COL6A2, COL1A1
GO:0043588	Skin development	8.30E-05	COL3A1, ADAMTS2, COL5A1
GO:0001568	Blood vessel development	1.04E-04	COL1A2, COL1A1, COL5A1
GO:0071230	Cellular response to amino acid stimulus	1.60E-04	COL3A1, COL1A2, COL1A1
GO:0007155	Cell adhesion	6.54E-04	COL6A2, COL15A1, COL1A1, COL5A1
GO:0005584	Collagen type I trimer	7.68E-04	COL1A2, COL1A1
GO:0005615	Extracellular space	8.67E-04	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1
GO:0030168	Platelet activation	9.55E-04	COL3A1, COL1A2, COL1A1
GO:0001501	Skeletal system development	0.001350963	COL3A1, COL1A2, COL1A1
GO:0050776	Regulation of immune response	0.002265849	COL3A1, COL1A2, COL1A1
GO:0032964	Collagen biosynthetic process	0.002498958	COL1A1, COL5A1
GO:0043589	Skin morphogenesis	0.003746428	COL1A2, COL1A1
GO:0043206	Extracellular fibril organization	0.00499256	COL3A1, COL5A1
GO:0070062	Extracellular exosome	0.013361493	CD248, COL1A2, COL6A2, COL15A1, COL5A1
GO:0046332	SMAD binding	0.017698138	COL3A1, COL1A2
GO:0007179	Transforming growth factor beta receptor signaling pathway	0.03773359	COL3A1, COL1A2
GO:0005178	Integrin binding	0.042743441	COL3A1, COL5A1
GO:0046872	Metal ion binding	0.043964665	COL3A1, COL1A2, COL1A1, COL5A1
GO:0050900	Leukocyte migration	0.04977108	COL1A2, COL1A1

Table 2 continued. The result of the top module II using GO term enrichment and KEGG pathway analysis.

ID	Term	P value	Genes
KEGG			
hsa04974	Protein digestion and absorption	2.99E-10	COL3A1, COL1A2, COL6A2, COL15A1, COL1A1, COL5A1
hsa04512	ECM-receptor interaction	1.16E-07	COL3A1, COL1A2, COL6A2, COL1A1, COL5A1
hsa04510	Focal adhesion	3.75E-06	COL3A1, COL1A2, COL6A2, COL1A1, COL5A1
hsa04151	PI3K-Akt signaling pathway	2.94E-05	COL3A1, COL1A2, COL6A2, COL1A1, COL5A1
hsa05146	Amoebiasis	3.43E-05	COL3A1, COL1A2, COL1A1, COL5A1
hsa04611	Platelet activation	6.33E-05	COL3A1, COL1A2, COL1A1, COL5A1

GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

We used the STRING online tool and Cytoscape software to construct the PPI network of DEGs. The highest-grade interaction genes were APOA2, COL3A1, COL1A1, TYR, COL1A2, NTS, COL5A1, PAPOLB, IGF2BP1, and ANHX. APOA2 encodes apolipoprotein (apo-) A-II, which is the second most abundant protein of high-density lipoprotein particles. Apolipoprotein A-II deficiency or hypercholesterolemia is caused by this gene defect. Moreover, disruption to the blood-brain barrier in pediatric brain tumor patients is likely to be the main reason for high overexpression of apolipoprotein A-II in the cerebrospinal fluid (CSF) [32]. Furthermore, some studies suggest that ApoA-II may have drug delivery potential and utility as a marker in identifying indolent disease, such as the pathogenesis of liver cancer [33–35].

The second hub gene, COL3A1, encodes collagen alpha-1 (III) chain, a protein which is a precursor to collagen III, continually combined with type I collagen. As a member of the collagen family, collagen alpha-1 (III) (COL3A1) is an important protein in the development and progression of several tumors, as it induces profound angiogenic responses in the host brain microvasculature that promote tumor growth [36,37]. Moreover, some studies determined that COL3A1 expression was upregulated in glioma and was directly correlated with the tumor grade [36,38]. As COL3A1 localizes to the extracellular matrix and was selectively expressed by the microvasculature, it may be a novel regulator of glioblastoma cell behavior and may be a suitable biomarker for diagnostic use or as a novel target for gene therapies against glioma [37,38].

COL1A1 encodes the major component of type I collagen, the fibrillar collagen that exists in masses of connective tissues. In both LGG and glioblastoma, overexpression of COL1A1 was found to be directly correlated with the tumor grade [39]. The upregulated expression of collagen 2 in glioma is accompanied by upregulation of COL1A1. The effect of mesenchymal-associated genes on cell invasion and gliomasphere initiation

were identified *in vitro*, by silencing of COL1A1, in which the gene has the potential for stratifying patients with glioma into subgroups for diagnosis of recurrence risk, as well as prognostic and therapeutic evolution [37,39,40].

Human tyrosinase is a single-membrane-spanning transmembrane protein, encoded by the TYR gene. Complete tumor resection can improve progression-free survival and overall survival. The ability to differentiate tumor cells during surgery can increase the success rate of tumor resection, thereby increasing the survival time of patients. Fluorescence-labeled tumor-specific antibodies (second tyrosinase-related protein) can accurately and specifically bind and identify tumor cells *in vivo*. The development of such brain tumor markers will contribute to the application of intraoperative fluorescence microscopy to increase the resection range of various intracranial tumors [41]. Some studies [42] have found that the tyrosinase promoter was significantly overexpressed in primary glioma samples, and tyrosinase expression correlated with the grade of the tumor. The role of the intracranial microenvironment in sculpting a location-specific profile of tumor antigen expression is very important, and clinical application of targeting novel anti-glioma therapies with checkpoint inhibition for aggressive brain tumors may be feasible [43,44].

The COL1A2 gene encodes one of the chains for type I collagen. COL1A2 plays an important role in medulloblastoma tumorigenesis [45]. Additionally, COL1A2 expression was differentiated in infant medulloblastomas of the desmoplastic histopathological subtype, suggesting that this distinct molecular pathogenesis may be the basis for these tumors and their relatively good prognosis. It is a powerful demonstration that the methylation status of specific genes may be relevant to the biological subclassification of medulloblastoma [45–47].

The NTS gene encodes a common precursor for 2 peptides: neuromedin N and neurotensin. Neurotensin is a secreted

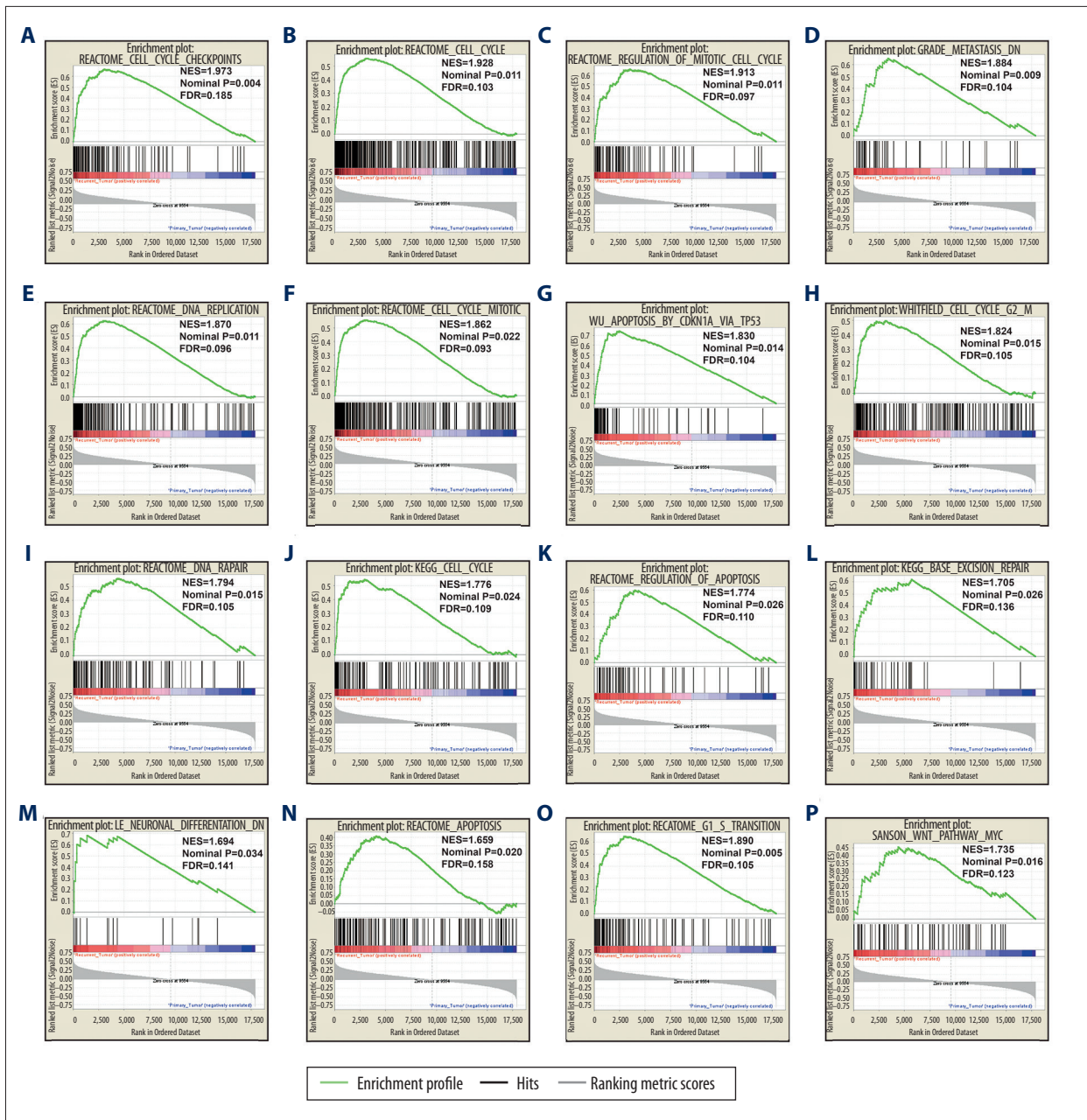
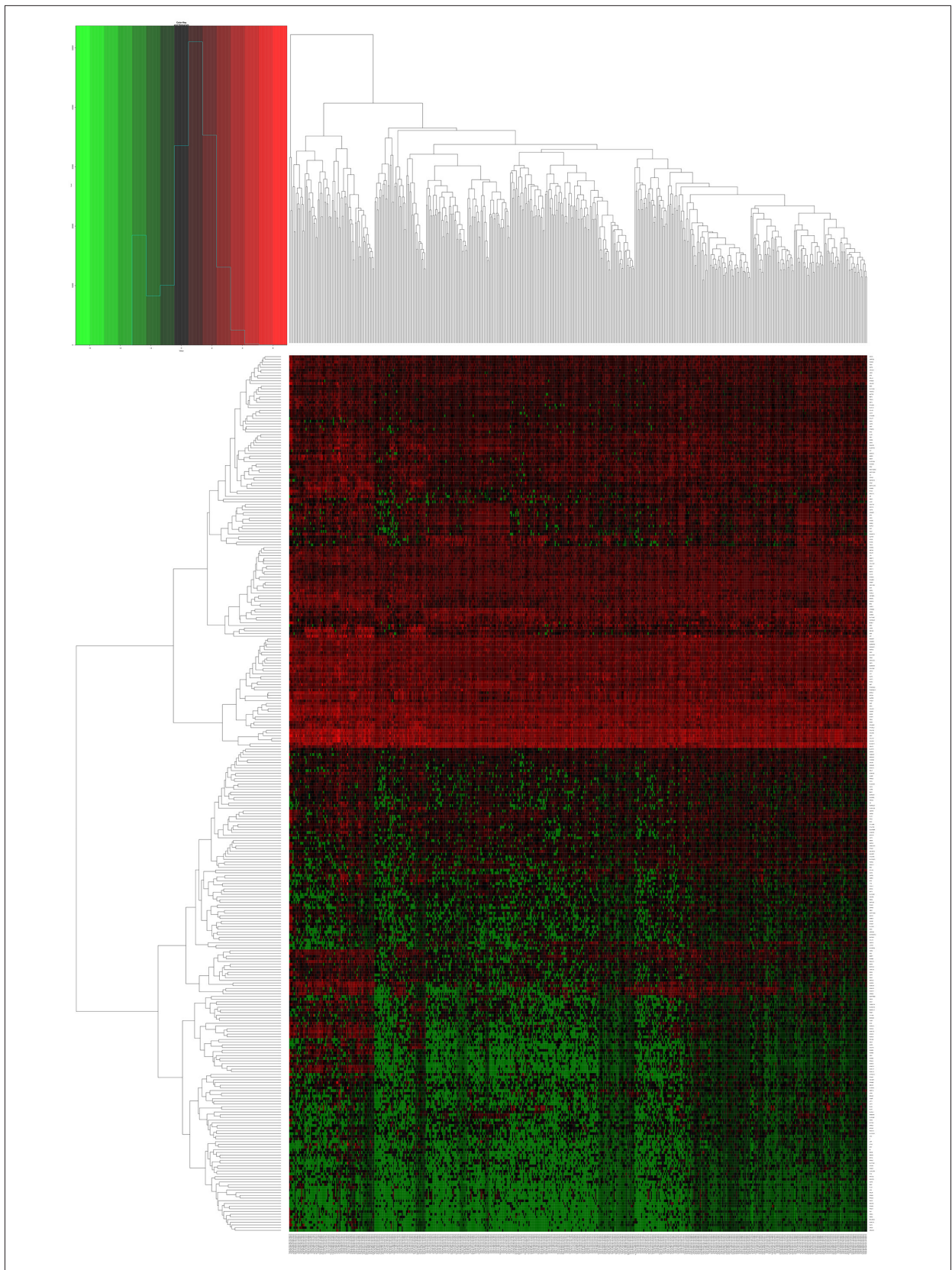


Figure 6. GSEA comparison of c2 reference gene set (A–P) between LGG primary and recurrence tumor tissues. ES – enrichment score; FDR – false discovery rate; GSEA – gene set enrichment analysis.

tridecapeptide, diffusely distributed throughout the central nervous system, and may serve as a neurotransmitter or a neuromodulator. Moreover, neurotensin (NTS) and its primary receptor, NTSR1, are implicated in cancer progression. Abnormal gene expression of NTS/NTSR1 results in the proliferation of glioblastoma cells. The expression level of NTS and NTSR1 was positively correlated with pathological grading of glioma. In keeping with this, unfavorable prognosis in glioma patients was associated with high expression levels of NTS and NTSR1 [48,49]. Furthermore, the suppression of the

NTSR1 function or the upregulation of miR-29b-1 and miR-129-3p expression decreased glioma cell proliferation, which suggests the NTS/NTSR1/c-Myc/miRNA axis may be a feasible therapeutic target for glioblastoma therapy [50].

An alpha chain for one of the low-abundance fibrillar collagens was encoded by COL5A1 gene. It has been proven that fibrin collagen exists widely in GBMs and mediates the invasion of tumor cells through a collagen-containing matrix [51]. Collagen scaffolds have been shown to be a major factor



Supplementary Figure 1. Heat map of the 326 DEGs. Red: upregulation; green: downregulation. DEGs – differentially-expressed genes.

associated with angiogenesis, metastasis, and rapid tumor growth in solid gliomas [52].

PAPOLB gene is a protein-coding gene for poly(A) polymerase beta. Some studies [53,54] suggest that PAPOLB regulates germ cell morphogenesis by modulating specific transcription factors at posttranscriptional and posttranslational levels. Our study found that this gene is associated with the development of glioma, but this requires more research to clarify.

IGF2BP1 gene encodes a member of the insulin-like growth factor 2 mRNA-binding protein family. It regulates some genes (including insulin-like growth factor 2, -actin and-transduction proteins) translation by binding to their mRNA. Some studies [55,56] found that IGF2BP1 subtype has become a biomarker for diagnosis and prognosis of astrocytomas. IGF2BP3 was identified as a GBM-specific proliferative and pre-invasive marker that activates the oncogene PI3K and MAPK pathways through the action of IGF-2 [57].

The ANHX gene is a protein-coding gene for anomalous homeobox. Through the identification of homologous domains, homologous proteins regulate the expression of target genes and guide the formation of many body structures during early embryonic development [58]. Over the past 2 decades, many studies have been published on how homologous proteins bind to DNA and increase their specificity by interacting with other proteins to regulate cell and tissue-specific gene expression [59]. The present study has demonstrated that the ANHX gene potentially acts as novel biomarker and therapeutic target for clinical assay development.

In the present study, a major limitation is that the results generated were obtained from a unitary cohort from TCGA, and the samples size of LGG recurrent tumor tissues is small. Since this

is a bioinformatics study of data mining, the number of samples, differential expression genes, and hub genes screening methods can affect the reproducibility of the results. Therefore, our findings still need to be verified in another cohort in future studies. However, despite this limitation, we identified a large number of DEGs between LGG primary and recurrent tumor tissues, and investigated the key genes and potential molecular mechanism by using multiple bioinformatics analysis methods, including GSEA. These findings may help to advance the understanding of LGG recurrence, and may have potential applications of value in LGG recurrence detection or targeted therapy.

Conclusions

In this study, 326 differentially-expressed genes associated with LGG recurrence were integrated analyzed and hub 10 genes were identified by bioinformatics. We also explored the potential mechanisms related to LGG relapse through functional enrichment assessment and GSEA methods. The study revealed that LGG relapse influences the basic functional status of cells, such as cell cycle, differentiation, DNA repair, and regulation of apoptosis. However, the findings of our study still require further verification in other cohorts.

Acknowledgments

The authors thank the contributors of The Cancer Genome Atlas (<https://portal.gdc.cancer.gov/>) for sharing the LGG dataset on open access.

Conflicts of interest

None.

References:

- 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: 803–20
- Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG et al: Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med*, 2015; 372: 2481–98
- Stupp R, Brada M, Tonn J et al: High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 2014; 25: 93–101
- Buckner JC, Shaw EG, Pugh SL et al: Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. *N Engl J Med*, 2016; 374: 1344–55
- Eckel-Passow JE, Lachance DH, Molinaro AM et al: Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med*, 2015; 372: 2499–508
- Chen B, Liang T, Yang P et al: Classifying lower grade glioma cases according to whole genome gene expression. *Oncotarget*, 2016; 7: 74031–42
- Mehta S, Lo Cascio C: Developmentally regulated signaling pathways in glioma invasion. *Cell Mol Life Sci*, 2018; 75(3): 385–402
- Vuong HG, Altibi AMA, Duong UNP et al: TERT promoter mutation and its interaction with IDH mutations in glioma: Combined TERT promoter and IDH mutations stratifies lower-grade glioma into distinct survival subgroups-A meta-analysis of aggregate data. *Crit Rev Oncol Hematol*, 2017; 120: 1–9
- Wang Z, Jensen MA, Zenklusen JC: A practical guide to The Cancer Genome Atlas (TCGA). *Methods Mol Biol*, 2016; 1418: 111–41
- Lee JS: Exploring cancer genomic data from the cancer genome atlas project. *BMB Rep*, 2016; 49: 607–11
- Shen L, Zhao L, Tang J et al: Key genes in stomach adenocarcinoma identified via network analysis of RNA-seq data. *Pathol Oncol Res*, 2017; 23(4): 745–52
- Robinson MD, McCarthy DJ, Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010; 26: 139–40
- Reiner A, Yekutieli D, Benjamini Y: Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 2003; 19: 368–75
- Benjamini Y, Drai D, Elmer G et al: Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, 2001; 125: 279–84

15. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995; 57: 289–300
16. Huang DW, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 2009; 37: 1–13
17. Huangda W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009; 4: 44–57
18. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res*, 2006; 34: D322–26
19. Ashburner M, Ball CA, Blake JA et al: Gene ontology: Tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet*, 2000; 25: 25–59
20. Szklarczyk D, Franceschini A, Wyder S et al: STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 2015; 43: D447–52
21. Liao X, Liu X, Yang C et al: Distinct diagnostic and prognostic values of microchromosome maintenance gene expression in patients with hepatocellular carcinoma. *J Cancer*, 2018; 9: 2357–73
22. Huang R, Liao X, Li Q: Identification of key pathways and genes in TP53 mutation acute myeloid leukemia: Evidence from bioinformatics analysis. *Onco Targets Ther*, 2018; 11: 163–73
23. Liao X, Huang R, Liu X et al: Distinct prognostic values of alcohol dehydrogenase mRNA expression in pancreatic adenocarcinoma. *Onco Targets Ther*, 2017; 10: 3719–32
24. Shiraki Y, Mii S, Enomoto A et al: Significance of perivascular tumour cells defined by CD109 expression in progression of glioma. *J Pathol*, 2017; 243: 468–80
25. Fan K, Wang X, Zhang J et al: Hypomethylation of CNTFRalpha is associated with proliferation and poor prognosis in lower grade gliomas. *Sci Rep*, 2017; 7: 7079
26. Guo KT, Fu P, Juerchott K et al: The expression of Wnt-inhibitor DKK1 (Dickkopf 1) is determined by intercellular crosstalk and hypoxia in human malignant gliomas. *J Cancer Res Clin Oncol*, 2014; 140: 1261–70
27. Kouchi M, Shibayama Y, Ogawa D et al: (Pro)renin receptor is crucial for glioma development via the Wnt/beta-catenin signaling pathway. *J Neurosurg*, 2017; 127: 819–28
28. Delic S, Lottmann N, Stelzl A et al: MiR-328 promotes glioma cell invasion via SFRP1-dependent Wnt-signaling activation. *Neuro Oncol*, 2014; 16: 179–90
29. Lehrer M, Bhadra A, Ravikumar V et al: Multiple-response regression analysis links magnetic resonance imaging features to de-regulated protein expression and pathway activity in lower grade glioma. *Oncoscience*, 2017; 4: 57–66
30. Backes C, Harz C, Fischer U et al: New insights into the genetics of glioblastoma multiforme by familial exome sequencing. *Oncotarget*, 2015; 6: 5918–31
31. Pal J, Patil V, Kumar A et al: Loss-of-function mutations in calcitonin receptor (CALCR) identify highly aggressive glioblastoma with poor outcome. *Clin Cancer Res*, 2018; 24: 1448–58
32. de Bont JM, den Boer ML, Reddingius RE et al: Identification of apolipoprotein A-II in cerebrospinal fluid of pediatric brain tumor patients by protein expression profiling. *Clin Chem*, 2006; 52: 1501–9
33. Malik G, Ward MD, Gupta SK et al: Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer. *Clin Cancer Res*, 2005; 11: 1073–85
34. Julovi SM, Xue A, Thanh LT et al: Apolipoprotein A-II plus lipid emulsion enhance cell growth via SR-B1 and target pancreatic cancer *in vitro* and *in vivo*. *PLoS One*, 2016; 11: e0151475
35. Jiang ZH, Zhang ZY, He M et al: [The screening and identification of Apolipoprotein A-II from serum differential proteins in hepatocellular carcinoma patients]. *Zhonghua Gan Zang Bing Za Zhi*, 2010; 18: 445–49 [in Chinese]
36. Gao YF, Zhu T, Chen J et al: Knockdown of collagen alpha-1(III) inhibits glioma cell proliferation and migration and is regulated by miR128-3p. *Oncol Lett*, 2018; 16: 1917–23
37. Liu Y, Carson-Walter EB, Cooper A et al: Vascular gene expression patterns are conserved in primary and metastatic brain tumors. *J Neurooncol*, 2010; 99: 13–24
38. Gao YF, Mao XY, Zhu T et al: COL3A1 and SNAP91: Novel glioblastoma markers with diagnostic and prognostic value. *Oncotarget*, 2016; 7: 70494–503
39. Mustafa DA, Sieuwerts AM, Zheng PP et al: Overexpression of colligin 2 in glioma vasculature is associated with overexpression of heat shock factor 2. *Gene Regul Syst Bio*, 2010; 4: 103–7
40. Balbous A, Cortes U, Guilloteau K et al: A mesenchymal glioma stem cell profile is related to clinical outcome. *Oncogenesis*, 2014; 3: e91
41. Fenton KE, Martirosyan NL, Abdelwahab MG et al: *In vivo* visualization of GL261-luc2 mouse glioma cells by use of Alexa Fluor-labeled TRP-2 antibodies. *Neurosurg Focus*, 2014; 36: E12
42. Ulasov IV, Rivera AA, Nettelbeck DM et al: An oncolytic adenoviral vector carrying the tyrosinase promoter for glioma gene therapy. *Int J Oncol*, 2007; 31: 1177–85
43. Cockle JV, Rajani K, Zaidi S et al: Combination viroimmunotherapy with checkpoint inhibition to treat glioma, based on location-specific tumor profiling. *Neuro Oncol*, 2016; 18: 518–27
44. Zhang JG, Eguchi J, Kruse CA et al: Antigenic profiling of glioma cells to generate allogeneic vaccines or dendritic cell-based therapeutics. *Clin Cancer Res*, 2007; 13: 566–75
45. Schwalbe EC, Lindsey JC, Straughton D et al: Rapid diagnosis of medulloblastoma molecular subgroups. *Clin Cancer Res*, 2011; 17: 1883–94
46. Liang Y, Diehn M, Bollen AW et al: Type I collagen is overexpressed in medulloblastoma as a component of tumor microenvironment. *J Neurooncol*, 2008; 86: 133–41
47. Anderton JA, Lindsey JC, Lusher ME et al: Global analysis of the medulloblastoma epigenome identifies disease-subgroup-specific inactivation of COL1A2. *Neuro Oncol*, 2008; 10: 981–94
48. Ouyang Q, Gong X, Xiao H et al: Neurotensin promotes the progression of malignant glioma through NTSR1 and impacts the prognosis of glioma patients. *Mol Cancer*, 2015; 14: 21
49. Ayala-Sarmiento AE, Martinez-Fong D, Segovia J: The Internalization of neurotensin by the low-affinity neurotensin receptors (NTSR2 and vNTSR2) activates ERK 1/2 in glioma cells and allows neurotensin-polyplex transfection of tGAS1. *Cell Mol Neurobiol*, 2015; 35: 785–95
50. Ouyang Q, Chen G, Zhou J et al: Neurotensin signaling stimulates glioblastoma cell proliferation by upregulating c-Myc and inhibiting miR-29b-1 and miR-129-3p. *Neuro Oncol*, 2016; 18: 216–26
51. Huijbers IJ, Iravani M, Popov S et al: A role for fibrillar collagen deposition and the collagen internalization receptor endo180 in glioma invasion. *PLoS One*, 2010; 5: e9808
52. Noreen R, Chien CC, Chen HH et al: FTIR spectro-imaging of collagen scaffold formation during glioma tumor development. *Anal Bioanal Chem*, 2013; 405: 8729–36
53. Kashiwabara SI, Tsuruta S, Yamaoka Y et al: PAPOLB/TPAP regulates spermiogenesis independently of chromatoid body-associated factors. *J Reprod Dev*, 2018; 64: 25–31
54. Kashiwabara SI, Tsuruta S, Okada K et al: Adenylation by testis-specific cytoplasmic poly(A) polymerase, PAPOLB/TPAP, is essential for spermatogenesis. *J Reprod Dev*, 2016; 62: 607–14
55. Tian X, Zhang L, Sun L et al: Low expression of insulin-like growth factor binding protein 7 associated with poor prognosis in human glioma. *J Int Med Res*, 2014; 42: 651–58
56. Santosh V, Arivazhagan A, Sreekanthreddy P et al: Grade-specific expression of insulin-like growth factor-binding proteins-2, -3, and -5 in astrocytomas: IGFBP-3 emerges as a strong predictor of survival in patients with newly diagnosed glioblastoma. *Cancer Epidemiol Biomarkers Prev*, 2010; 19: 1399–408
57. Suvasini R, Shruti B, Thota B et al: Insulin growth factor-2 binding protein 3 (IGF2BP3) is a glioblastoma-specific marker that activates phosphatidylinositol 3-kinase/mitogen-activated protein kinase (PI3K/MAPK) pathways by modulating IGF-2. *J Biol Chem*, 2011; 286: 25882–90
58. Corsetti MT, Briata P, Sanseverino L et al: Differential DNA binding properties of three human homeodomain proteins. *Nucleic Acids Res*, 1992; 20: 4465–72
59. Burglin TR, Affolter M: Homeodomain proteins: An update. *Chromosoma*, 2016; 125: 497–521