

# SCIENTIFIC REPORTS



OPEN

## Gene expression meta-analysis in diffuse low-grade glioma and the corresponding histological subtypes

Siqi Wang<sup>1</sup>, Feng Jin<sup>1,3</sup>, Wenliang Fan<sup>1</sup>, Fang Liu<sup>1</sup>, Yan Zou<sup>1</sup>, Xuehan Hu<sup>1</sup>, Haibo Xu<sup>2</sup> & Ping Han<sup>1</sup>

Diffuse low-grade glioma (DLGG) is a well-differentiated, slow-growing tumour with an inherent tendency to progress to high-grade glioma. The potential roles of genetic alterations in DLGG development have not yet been fully delineated. Therefore, the current study performed an integrated gene expression meta-analysis of eight independent, publicly available microarray datasets including 291 DLGGs and 83 non-glioma (NG) samples to identify gene expression signatures associated with DLGG. Using INMEX, 708 differentially expressed genes (DEGs) (385 upregulated and 323 downregulated genes) were identified in DLGG compared to NG. Furthermore, 497 DEGs (222 upregulated and 275 downregulated genes) corresponding to two histological types were identified. Of these, high expression of *HIP1R* significantly correlated with increased overall survival, whereas high expression of *TBXAS1* significantly correlated with decreased overall survival. Additionally, network-based meta-analysis identified *FN1* and *APP* as the key hub genes in DLGG compared with NG. *PTPN6* and *CUL3* were the key hub genes identified in the astrocytoma relative to the oligodendroglioma. Further immunohistochemical validation revealed that *MTHFD2* and *SPARC* were positively expressed in DLGG, whereas *RBP4* was positively expressed in NG. These findings reveal potential molecular biomarkers for diagnosis and therapy in patients with DLGG and provide a rich and novel candidate reservoir for future studies.

According to histological characteristics, gliomas can be classified into grades I–IV based on World Health Organization (WHO) criteria published in 2007 and 2016<sup>1,2</sup>. Patients with low-grade glioma (grades I and II) have a median survival time of 4.7–9.8 years, with a range of up to 13 years for certain subtypes<sup>3,4</sup>. Grade I gliomas are often localized and are more likely to be cured after surgical resection. Grade II gliomas, also known as diffuse low-grade glioma (DLGG), account for approximately 15% of all gliomas<sup>1</sup> and have heterogeneous and complicated presentations that correspond to three histological types: astrocytoma (A), oligodendroglioma (OD), and oligoastrocytoma (OA). Okamoto *et al.*<sup>5</sup> demonstrated that histological type is a significant predictor of survival of patients with DLGG. Patients with OD survive longer (median survival time: 11.6 years) than patients with A (median survival time: 5.6 years)<sup>6</sup>. Although DLGGs have low-level proliferative activity, their natural course is to transform or dedifferentiate into high-grade glioma (WHO grade III–IV), and they often recur after surgical resection<sup>1,7</sup>. The surgical and medical management of DLGG remains one of the major controversies in current neurooncology. Thus, there is an urgent need to develop novel diagnostic and therapeutic targets for DLGG.

Hartmann *et al.* investigated the prognostic relevance of four prominent molecular markers in WHO grade II gliomas, including *TP53* mutation, 1p/19q deletion, O6-methylguanylmethyltransferase (*MGMT*) promoter methylation, and isocitrate dehydrogenase 1 (*IDH1*) mutation. The results showed that 1p/19q codeletion and *IDH1* mutation are prognostic markers following the administration of radiotherapy or chemotherapy<sup>8</sup>. *TP53* mutations have been most commonly identified in A, whereas 1p/19q codeletion is more common in OD. OAs appear to be heterogeneous and typically show either *TP53* mutations or 1p/19q deletion<sup>8</sup>. The 2016 WHO classification of CNS tumours defines tumour entities based on histology and a combination of molecular aberrations,

<sup>1</sup>Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China. <sup>2</sup>Department of Radiology, Zhongnan Hospital, Wuhan University, Wuhan 430071, China. <sup>3</sup>Department of Radiology, The First Affiliated Hospital, Inner Mongolia Medical University, Hohhot 010050, China. Correspondence and requests for materials should be addressed to H.X. (email: [xuhaibo1120@hotmail.com](mailto:xuhaibo1120@hotmail.com)) or P.H. (email: [cjr.hanping@vip.163.com](mailto:cjr.hanping@vip.163.com))

Source accession	Publication year	First author	Country	Platform	Numbers		Reference
Datasets included in the meta-analysis of DLGG vs. NG tissues					DLGG	NG	
GSE68848	2015	Fine H	USA	GPL570, Affymetrix U133 Plus 2.0	99	28	22
GSE16011	2010	Gravendeel LA	Netherlands	GPL8542, Affymetrix U133 Plus 2.0	21	8	23
GSE4290	2006	Fine HA	USA	GPL570, Affymetrix U133 Plus 2.0	45	23	24
GSE12657	2008	Moran LB	United Kingdom	GPL8300, Affymetrix U95 Version 2 Array	7	5	NA
GSE21354	2010	Liu Z	China	GPL570, Affymetrix U133 Plus 2.0	10	4	25
GSE2223	2006	Bredel M	USA	GPL1833, SHFK	6	4	26
GSE70231	2015	Mervi Heiskanen	USA	GPL80, Affymetrix Human Full Length HuGeneFL Array	6	6	27
CGGA	2012	Yan W	China	Agilent Whole Human Genome Array platform	97	5	28
Datasets included in the meta-analysis of A vs. OD tissues					A	OD	
GSE68848	2015	Fine H	USA	GPL570, Affymetrix U133 Plus 2.0	65	30	22
GSE16011	2010	Gravendeel LA	Netherlands	GPL8542, Affymetrix U133 Plus 2.0	13	8	23
GSE4290	2006	Fine HA	USA	GPL570, Affymetrix U133 Plus 2.0	7	38	24
GSE21354	2010	Liu Z	China	GPL570, Affymetrix U133 Plus 2.0	5	5	25
CGGA	2012	Yan W	China	Agilent Whole Human Genome Array platform	58	17	28

**Table 1.** Characteristics of datasets included in the meta-analysis of DLGG vs. NG and A vs. OG tissues. Abbreviations: DLGG, diffusely low-grade glioma; NG, non-glioma; GSE, gene expression omnibus; GPL, gene platform; CGGA, the Chinese Glioma Genome Atlas; A, astrocytoma; OD, oligodendroglioma.

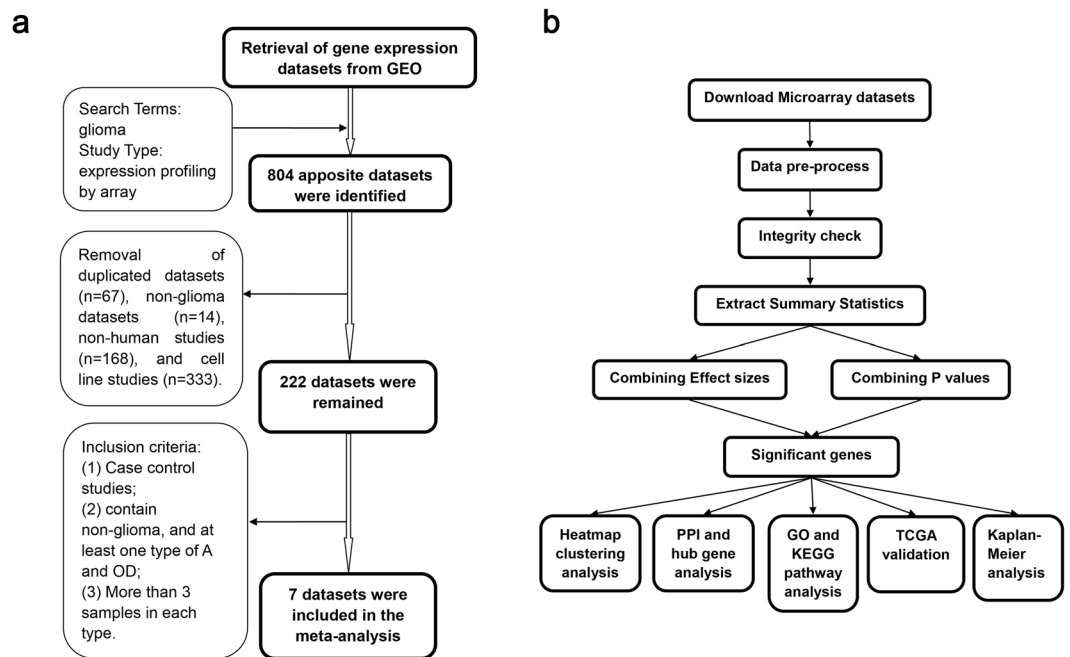
such as IDH mutation, *ATRX* mutation, 1p/19q deletion, and *TP53* mutation<sup>2</sup>. As we gain further insight into molecular biomarkers of glioma, the impact of these markers on diagnosis and treatment continues to evolve.

High-throughput genomics technologies, such as microarrays that provide simultaneous measurements of the expression profiles of thousands of genes, have provided substantial insight into the processes that drive disease development. Although prior studies utilizing microarrays have identified numerous differentially expressed genes (DEGs), inconsistencies exist between studies due to variations in sample size and quality<sup>9,10</sup>. To address this limitation, meta-analyses have been applied to synthesize the information available in publicly available gene expression datasets to identify reliable molecular biomarkers of disease<sup>11</sup>. Importantly, meta-analyses provide enhanced statistical power, allowing the discovery of robust and reliable gene signatures. Prior meta-analyses have been performed to investigate biomarkers in breast cancer<sup>12</sup>, prostate cancer<sup>13</sup>, liver cancer<sup>14</sup>, and lung cancer<sup>15</sup>. Integrative meta-analysis of expression data (INMEX), which allows simultaneous analysis of multiple gene expression datasets, has also been applied<sup>16–18</sup>.

In the present study, we used INMEX to perform meta-analyses of eight eligible microarray datasets to identify key regulators and potential diagnostic and therapy biomarkers associated with DLGG and its clinical subtypes. To the best of our knowledge, this study is the first to explore diagnostic and therapy biomarkers associated with DLGG and its histological subtypes by performing meta-analyses of gene expression datasets.

## Results

**Studies included in the meta-analysis.** A total of 7 studies from the Gene Expression Omnibus (GEO) dataset were included: GSE68848<sup>19</sup>, GSE16011<sup>20</sup>, GSE4290<sup>21</sup>, GSE12657, GSE21354<sup>22</sup>, GSE2223<sup>23</sup>, and GSE70231<sup>24</sup>. Additionally, mRNA expression data from 97 WHO grade II samples, including 58A, 17 OD, 22 OA, and 5 non-glioma (NG) samples, were collected from the Chinese Glioma Genome Atlas (CGGA)<sup>25</sup>. These eight studies were examined using meta-analysis to identify differences between DLGGs and NGs and included a total of 291 cases and 83 controls. To identify possible DEGs between histological DLGG subtypes (A and OD, but not OA, which is not recognized as a separate tumour entity in the 2016 CNS tumour classification system<sup>2</sup>), an additional meta-analysis was performed to examine differences between A and OD samples. For this purpose, five datasets containing information on A and OD (GSE4290, GSE16011, GSE21354, GSE68848, and CGGA) were selected to compare mRNA expression signatures among low-grade glioma subtypes; these datasets included a total of 148A and 98 OD samples. Table 1 provides detailed information on each dataset including the number of each sample type, reference, and the microarray platform used. We obtained expression data and clinical information for 254 patients from the TCGA; these data included 63A, 112 OD, 74 OA, and 5 NG samples. Figure 1 depicts our experimental workflow.



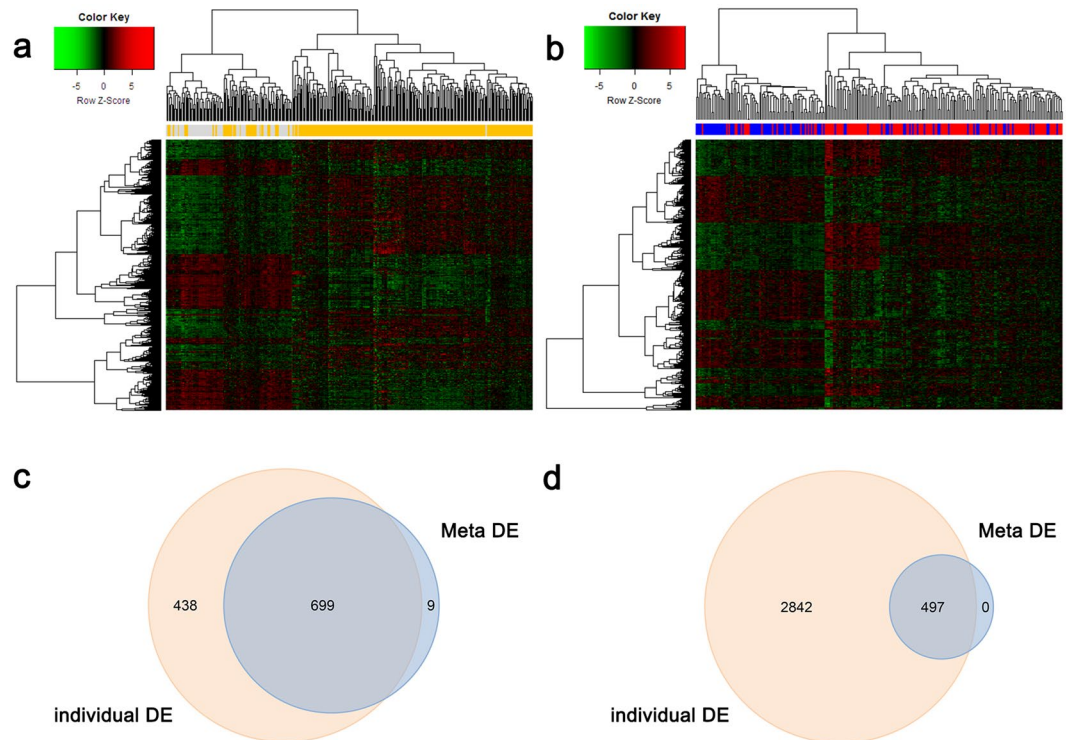
**Figure 1.** Study workflow. **(a)** Identification of eligible gene expression datasets for meta-analysis of DLGG. **(b)** The process used for meta-analysis. Abbreviations: DLGG, diffuse low-grade glioma; GEO, Gene Expression Omnibus; A, astrocytoma; OD, oligodendroglioma; PPI, protein-protein interaction; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, the Cancer Genome Atlas.

**Batch effect adjustment.** Before performing the DLGG and subtype meta-analyses, we corrected for the batch effect using the ComBat procedure in INMEX. The principal component analysis plot showed that each dataset was clearly separated from the others before applying the batch adjustment algorithm. After the adjustment, the samples from all datasets were well intermixed as shown in Supplementary Fig. S1.

**Meta-analysis of gene expression in DLGG.** Gained genes were defined as DEGs with weak but consistent expression profiles across all the datasets, and lost genes were identified as DEGs that appeared in individual analysis, but not in the meta-analysis, or those with large variations across different studies (due to experimental errors or different platforms). In total, there were 9 gained genes and 438 lost genes in the DLGG meta-analysis (Fig. 2c, Supplementary Table S1). From the microarray meta-analysis, we identified 708 DEGs, including 385 upregulated and 323 downregulated genes, as shown in Supplementary Table S1. We conducted two-way hierarchical clustering analysis of all DEGs. The hierarchical clustering map revealed that DLGG and NG samples were non-random partitioned into two major groups (Fig. 2a). In the DLGG datasets, secreted protein acidic and cysteine rich (*SPARC*), methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase (*MTHFD2*), and protein tyrosine phosphatase, receptor type Z1 (*PTPRZ1*) were the most significantly upregulated genes, and retinol-binding protein 4 (*RBP4*), cholecystokinin-B receptor (*CCKBR*), and syntaxin 1A (*STX1A*) were the most significantly downregulated genes. The top 10 upregulated and downregulated genes are shown in Table 2.

Meta-analysis was also used to compare the two subtypes of DLGG among five datasets. There were 0 gained genes and 2842 lost genes in the meta-analysis of the A vs. OD samples (Fig. 2d, Supplementary Table S2). In the microarray meta-analysis, a total of 497 DEGs were identified, including 222 upregulated and 275 downregulated genes, as shown in Supplementary Table S2. The two-way hierarchical clustering map in Fig. 2b shows that A and OD samples were non-random partitioned into two major groups. Bridging integrator 2 (*BIN2*), Thromboxane A synthase 1 (*TBXAS1*), and Transmembrane protein 119 (*TMEM119*) were the most significantly upregulated genes, and elongator acetyltransferase complex subunit 3 (*ELP3*), and Heat shock protein family A (Hsp70) member 9 (*HSPA9*), and Huntingtin interacting protein 1 related (*HIP1R*) were the most significantly downregulated genes between the A and OD samples. The top 10 upregulated and downregulated genes are shown in Table 2.

**Identification of hub genes using network-based meta-analysis.** We constructed a network to identify the critical hub genes among the DEGs identified in the meta-analysis. NetworkAnalyst, available on the web, enables analysis of protein-protein interaction (PPI) networks for multiple gene lists using InnateDB. The database integrates experimental data from IntAct, MINT, BIND, BioGRID and DIP with manually curated protein interaction data from the published literature. The expanded PPI network for DLGG contained 8713 nodes and 31,253 connection edges. We conducted “Zero order” interaction network analysis in a layout format using a force atlas to better visualize the network, which included 535 nodes and 1856 connection edges (Fig. 3a). Hub genes in the network were ranked by degree. Within this network, fibronectin 1 (*FN1*), which had an adjusted *p*-value of



**Figure 2.** Hierarchical clustering and Venn diagram. (a) Two-way hierarchical clustering based on 708 DEGs in DLGG vs. NG tissues across 8 datasets. DLGG (orange label) and NG (grey label) samples fell into two major clusters. (b) Two-way hierarchical clustering based on 497 DEGs in A vs. OD tissues across 5 datasets. A (red label) and OD (blue label) samples fell into two major groups. (c) Venn diagram showing the common and unique DE genes in DLGG vs. NG tissues between the Meta-DE and Individual-DE datasets. (d) Venn diagram showing the common and unique DE genes in A vs. OD tissues between the Meta-DE and Individual-DE datasets. Abbreviations: DLGG, diffuse low-grade glioma; NG, non-glioma; A, astrocytoma; OD, oligodendroglioma; DE, differentially expressed.

3.28E-15 and a combined effect size (ES) of 1.1499, was the most highly ranked hub gene (degree = 80; betweenness = 15,956.77) among the upregulated DEGs, and amyloid beta precursor protein (*APP*), which had an adjusted *p*-value of 4.55E-05 and a combined ES of -1.2673, was the most highly ranked hub gene (degree = 98; betweenness = 27455.55) among the downregulated DEGs. The top ten hub genes are detailed in Supplementary Table S1.

We also conducted network analysis to identify differences between A and OD tissues using NetworkAnalyst. The expanded PPI network for DLGG contained 4732 nodes and 9792 connection edges. The “zero order” interaction network was examined in a force atlas layout format and included 137 nodes and 187 connection edges (Fig. 3b). Within this network, protein tyrosine phosphatase, non-receptor type 6 (*PTPN6*), which had an adjusted *p*-value of 2.89E-05 and a combined ES of 1.3487, was the most highly ranked hub gene (degree = 11; betweenness = 2004.06) among the upregulated DEGs, and cullin 3 (*CUL3*), which had an adjusted *p*-value of 7.00E-07 and a combined ES of -0.89921, was the most highly ranked hub gene (degree = 38; betweenness = 5903.56) among the downregulated DEGs. The top ten hub genes are shown in Supplementary Table S2.

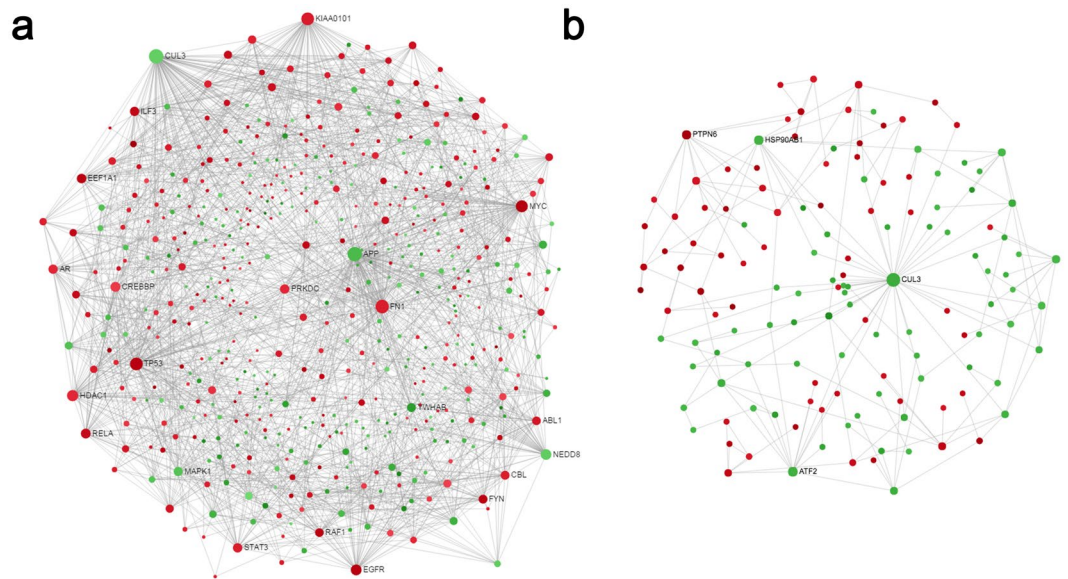
**Functional analysis.** To identify the potential functions of the identified DEGs, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses in the Database for Annotation, Visualization, and Integrated Discovery (DAVID). The GO analysis covered the following three domains: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF).

When comparing the DLGG and NG samples, *pathways in cancer* was the top enriched KEGG pathway among DEGs. The top enriched GO terms included *intracellular signal transduction* (BP), *extracellular exosome* (CC), and *ATP binding* (MF). When comparing the A and OD samples, *cytokine-cytokine receptor interaction* was the top enriched KEGG pathway among DEGs, and *immune response* (BP), *plasma membrane* (CC), and *protein binding* (MF) were the top enriched GO terms among DEGs. The minimum ten enriched and significant GO terms and pathway terms were showed in Fig. 4 (panel 4f showed a minimum of 9 enriched and statistically significant terms).

**The Cancer Genome Atlas (TCGA) dataset validation.** The top three upregulated and downregulated mRNAs in the meta-analyses of DLGG and its subtypes were validated using the TCGA dataset. Notably, we found that *SPARC*, *MTHFD2*, and *PTPRZ1* were significantly overexpressed in the DLGG compared with the NG samples. *RBP4*, *CCKBR*, and *STX1A* were significantly underexpressed in the DLGG compared with the NG

Entrez ID	Gene symbol	Gene name	Combined ES	Adjusted <i>p</i> -value
Top 10 upregulated genes in DLGG vs. NG				
6678	<i>SPARC</i>	Secreted protein acidic and cysteine rich	2.8727	0
10797	<i>MTHFD2</i>	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	2.8634	0
5803	<i>PTPRZ1</i>	Protein tyrosine phosphatase, receptor type Z1	2.7797	1.26E-10
6659	<i>SOX4</i>	SRY-box 4	2.6366	0
9459	<i>ARHGEF6</i>	Rac/Cdc42 guanine nucleotide exchange factor 6	2.4971	0
5375	<i>PMP2</i>	Peripheral myelin protein 2	2.4946	0
6938	<i>TCF12</i>	Transcription factor 12	2.4753	0
6175	<i>RPLP0</i>	Ribosomal protein lateral stalk subunit P0	2.4696	0
7079	<i>TIMP4</i>	TIMP metalloproteinase inhibitor 4	2.4588	0
7078	<i>TIMP3</i>	TIMP metalloproteinase inhibitor 3	2.3411	0
Top 10 downregulated genes in DLGG vs. NG				
5950	<i>RBP4</i>	Retinol-binding protein 4	-2.7344	6.24E-15
887	<i>CCKBR</i>	Cholecystokinin B receptor	-2.6242	6.24E-15
6804	<i>STX1A</i>	Syntaxin 1A	-2.5435	0
23531	<i>MMD</i>	Monocyte to macrophage differentiation associated	-2.5336	0
7781	<i>SLC30A3</i>	Solute carrier family 30 member 3	-2.5278	7.92E-09
1428	<i>CRYM</i>	Crystallin mu	-2.4535	0
1020	<i>CDK5</i>	Cyclin-dependent kinase 5	-2.4533	0
3761	<i>KCNJ4</i>	Potassium voltage-gated channel subfamily J member 4	-2.4439	0
2555	<i>GABRA2</i>	Gamma-aminobutyric acid type A receptor alpha2 subunit	-2.4082	0
529	<i>ATP6V1E1</i>	ATPase H+ transporting V1 subunit E1	-2.4026	0
Top 10 Upregulated Genes in A vs. OD				
51411	<i>BIN2</i>	Bridging integrator 2	1.6907	0.0000223
6916	<i>TBXAS1</i>	Thromboxane A synthase 1	1.6221	0.0000863
338773	<i>TMEM119</i>	Transmembrane protein 119	1.6106	0.000045
7462	<i>LAT2</i>	Linker for activation of T-cells family member 2	1.5738	0.0000291
81704	<i>DOCK8</i>	Dedicator of cytokinesis 8	1.5245	0.0000641
54518	<i>APBB1IP</i>	Amyloid beta precursor protein binding family B member 1 interacting protein	1.4871	0.0000329
5272	<i>SERPINB9</i>	Serpin family B member 9	1.4734	0.00000143
23533	<i>PIK3R5</i>	Phosphoinositide-3-kinase regulatory subunit 5	1.4619	0.00000757
54440	<i>SASH3</i>	SAM and SH3 domain containing 3	1.4526	0.00000121
112616	<i>CMTM7</i>	CKLF-like MARVEL transmembrane domain containing 7	1.4252	0.0000214
Top 10 downregulated genes in A vs. OD				
55140	<i>ELP3</i>	Elongator acetyltransferase complex subunit 3	-1.4435	0.0000694
3313	<i>HSPA9</i>	Heat shock protein family A (Hsp70) member 9	-1.318	8.73E-08
9026	<i>HIP1R</i>	Huntingtin interacting protein 1 related	-1.3054	2.84E-09
23219	<i>FBXO28</i>	F-box protein 28	-1.2516	0.0000275
84894	<i>LINGO1</i>	Leucine rich repeat and Ig domain containing 1	-1.2256	0.0000642
440026	<i>TMEM41B</i>	Transmembrane protein 41B	-1.2254	8.18E-11
79608	<i>RIC3</i>	RIC3 acetylcholine receptor chaperone	-1.2223	8.18E-11
51340	<i>CRNKL1</i>	Crooked neck pre-mRNA splicing factor 1	-1.2067	0.0000443
140767	<i>NRSN1</i>	Neurensin 1	-1.1524	0.0000705
8539	<i>API5</i>	Apoptosis inhibitor 5	-1.1432	1.12E-09

**Table 2.** Top 20 DEGs identified in the meta-analysis of DLGG vs. NG and A vs. OD tissues. The DEGs were ranked according to the combined effect size. Abbreviations: ES, effect size; DLGG, diffused low-grade glioma; NG, non-glioma NG; A, astrocytoma; OD, oligodendroglioma.



**Figure 3.** Network-based meta-analysis of hub genes. **(a)** “Zero order” interaction network of the DEGs identified in the meta-analysis of DLGG vs. NG tissues shown in a force atlas layout format. **(b)** “Zero order” interaction network of the DEGs identified in the meta-analysis of A vs. OD tissues shown in a force atlas layout format. The red and green nodes represent upregulated and downregulated DEGs, respectively. Abbreviations: DLGG, diffuse low-grade glioma; NG, non-glioma; A, astrocytoma; OD, oligodendroglioma; DEG, differentially expressed gene.

samples. *BIN2*, *TBXAS1*, and *TMEM119* were significantly overexpressed in the A samples compared with the OD samples. *ELP3*, *HSPA9*, and *HIP1R* were significantly underexpressed in the A samples compared with the OD samples (Fig. 5).

**Kaplan-Meier analysis.** To explore how the identified DEGs affect survival, we conducted Kaplan-Meier analysis using GraphPad Prism 6 with the DLGG cohorts in the TCGA dataset, which contained 246 patients with complete clinical traits and prognostic information. We assessed the prognostic value of the top 3 upregulated and downregulated DEGs in meta-analyses of DLGG vs. NG and A vs. OD. The results from Kaplan-Meier analysis demonstrated that patients with high *TBXAS1* expression had significantly shorter survival than those with high expression of these genes ( $p = 0.0031$ ), and patients with high *HIP1R* expression lived significantly longer than those with low *HIP1R* expression ( $p = 0.0010$ ). The Kaplan-Meier survival curves are shown in Fig. 6, a and b, detailed in Supplementary Fig. S2.

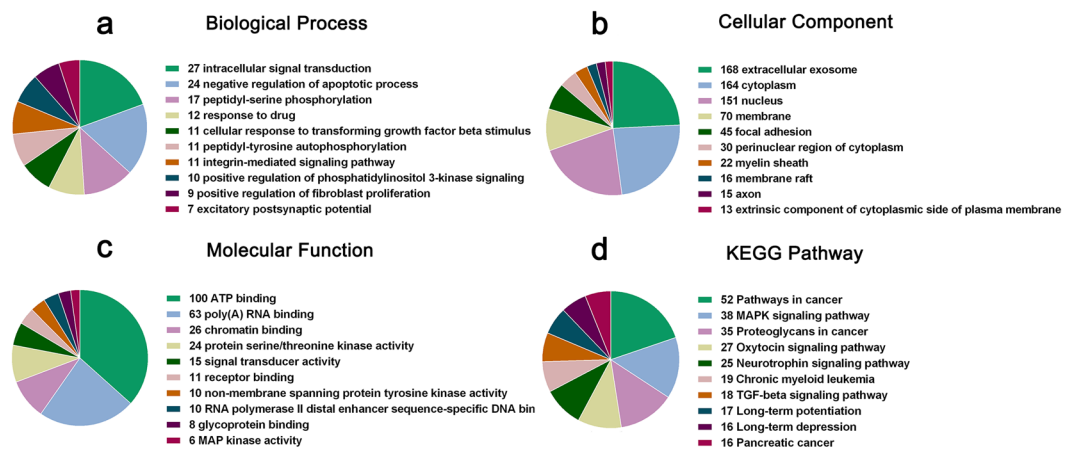
**Immunohistochemical validation.** The protein expression of SPARC, MTHFD2, PTPRZ1, RBP4, CCKBR, and STX1A in DLGG and control samples were analyzed by immunohistochemistry (Fig. 6, c and d). For MTHFD2 and SPARC, weak cytoplasmic positivity were observed in DLGG samples, while control brain tissues demonstrated negative staining ( $p = 0.002$  and  $0.0015$ , respectively). High expression of RBP4 was noted in control brain tissues as against DLGG samples ( $p = 0.015$ ). The immunohistochemical results are detailed in Supplementary Table S3.

Quantitative real-time polymerase chain reaction (qPCR) assay for the expression level of SPARC, MTHFD2, PTPRZ1, RBP4, CCKBR, and STX1A in Hs683 and HEB cells provided a further confirmation of the different expression levels of these DEGs in DLGG vs. NG (Fig. S3). Consistently with meta-analysis, the level of SPARC was significantly upregulated while RBP4 and CCKBR were downregulated in Hs683 cells ( $p = 0.011$  for SPARC,  $p < 0.0001$  for RBP4, and  $p < 0.0001$  for CCKBR, respectively). For MTHFD2, PTPRZ1, and STX1A, the qPCR results were contradicted with results in meta-analysis.

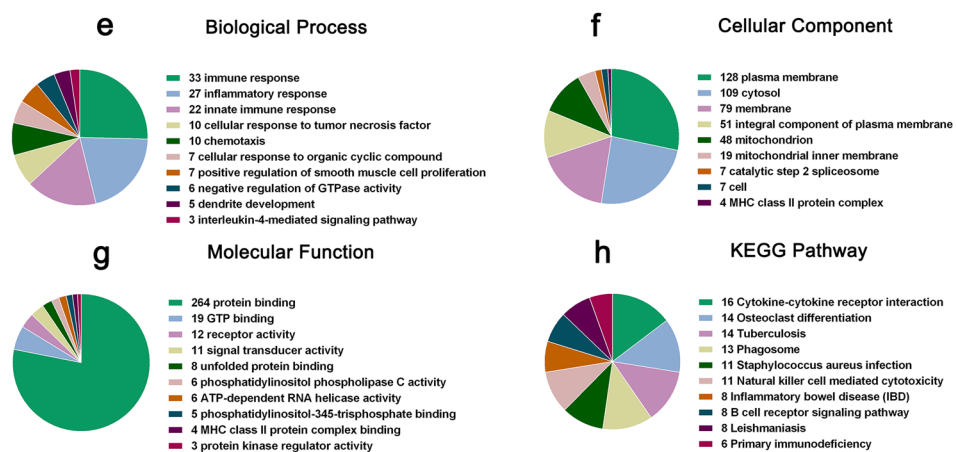
## Discussion

The 2016 version of the WHO classification criteria for glioma combined genotypic parameters with traditional histology and included IDH mutation, *ATRX* mutation, 1p/19q deletion, and *TP53* mutation as classification factors<sup>2,26–28</sup>. With the continued development of high-throughput genomics technology, many studies have explored genetic alterations associated with low-grade gliomas, such as telomerase reverse transcriptase (*TERT*) promoter mutation<sup>29</sup>, CpG island methylator phenotypes (*CIMP*)<sup>30</sup>, O-6-methylguanine-DNA methyltransferase (*MGMT*) promoter methylation<sup>31</sup>, and changes in nestin expression<sup>32</sup>. Although microarray studies have produced a large quantity of data, the sample sizes of most prior studies have been limited, affecting the identification of reliable DEGs. Meta-analysis of multiple microarray datasets makes it possible to increase the effective sample size and improve predictive power, making the identification of DEGs more reliable. Histological diagnosis of DLGG subtype is difficult due to inter-observer variability, particularly in cases lacking the typical symptoms associated with tumours of astrocytic or oligodendroglial lineage. In the present study, we attempted to identify DEGs associated

## DLGG vs. NG



## A vs. OD

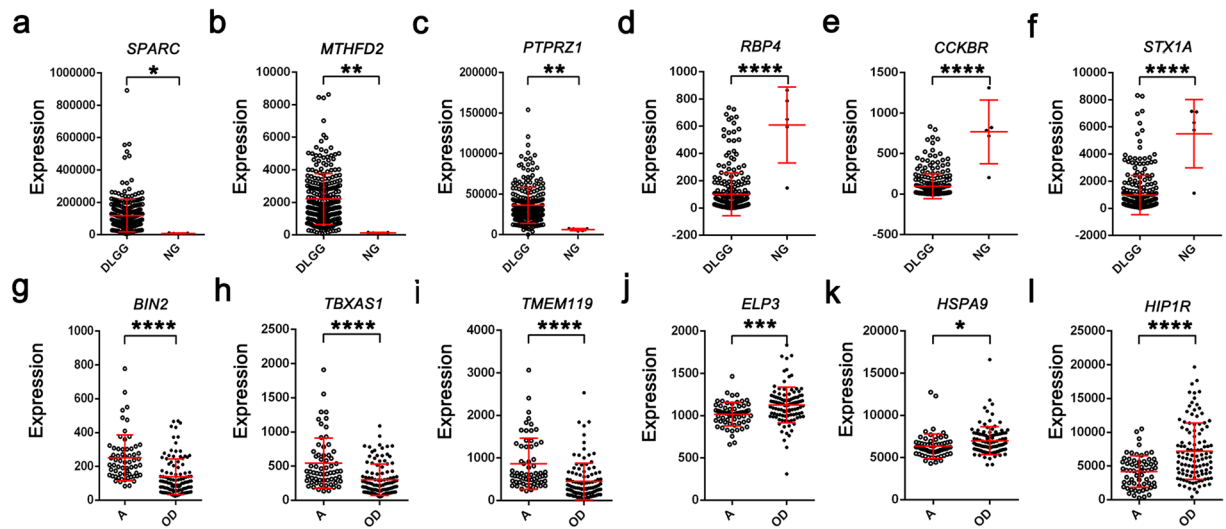


**Figure 4.** GO and KEGG pathway analysis of DEGs. The GO analysis covered the following domains: (a,e) Biological Process, (b,f) Cellular Component, and (c,g) Molecular Function. (d,h) KEGG pathway analysis.  $P$ -value  $< 0.05$  was significant. Abbreviations: GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEG, differentially expressed gene; DLGG, diffuse low-grade glioma; NG, non-glioma; A, astrocytoma; OD, oligodendroglioma.

with DLGG and its histological subtypes by performing meta-analyses of publically available microarray datasets. To the best of our knowledge, this is the first such attempt in DLGG research.

Across the eight analysed studies, we identified 708 genes that were consistently differentially expressed in DLGG vs. NG tissues (385 upregulated and 323 downregulated). Interestingly, we identified 9 gained genes with weak but consistent expression profiles across all the datasets, which were not discovered in the prior individual analyses. Among the upregulated DEGs, *SPARC* had the highest combined ES of 2.87, consistent with a report by Huang *et al.*, which demonstrated that *SPARC* was upregulated by more than 2-fold in 20–60% of DLGG cases. Further immunohistochemical staining for *SPARC* was strongly positive in the cytoplasm of neoplastic cells and reactive astrocytes and negative in adjacent normal brain tissues<sup>33</sup>, which was in accordance with our immunohistochemical analysis. qPCR assay for the expression level of *SPARC* in Hs683 and HEB cells provided a further validation. *RBP4*, a recently identified adipokine, has been associated with many types of cancer. Uehara *et al.*<sup>34</sup> reported that *RBP4* was over-expressed in prostate cancer cells and associated with the growth of these cells *in vitro*. In addition, higher circulating levels of *RBP4* have been associated with colon adenoma<sup>35</sup>, oral squamous cell cancer<sup>36</sup>, and breast cancer. In a study reported by Lorkova *et al.*<sup>37</sup>, low concentrations of *RBP4* were identified in sera collected from patients with epithelial ovarian cancer. However, no studies have investigated the relationship between *RBP4* and DLGG. *RBP4* had the highest combined ES (−2.73) among the examined downregulated DEGs, and *RBP4* was negative staining in DLGG samples as against NG samples and downregulated in Hs683 cells by qPCR assay. The results implied that *RBP4* might be associated with DLGG, having the potential to be a potential target for diagnosis and therapy in DLGG.

*MTHFD2* is a metabolic enzyme that participates in mitochondrial folate one-carbon metabolism. Overexpression of *MTHFD2* has been associated with poor prognosis in patients with breast cancer<sup>38</sup>, and knock-down of *MTHFD2* in breast cancer cell lines reduced cell viability; increased apoptosis<sup>39</sup>; decreased migration,



**Figure 5.** TCGA dataset validation. Expression levels of the top three upregulated (a–c) and downregulated (d–f) DEGs in the DLGG vs. NG tissues in the TCGA cohorts. Expression levels of the top three upregulated (g–i) and downregulated (j–l) DEGs in the A vs. OD tissues in the TCGA cohorts. Abbreviations: DLGG, diffuse low-grade glioma; NG, non-glioma; A, astrocytoma; OD, oligodendroglioma; DEG, differentially expressed gene. TCGA, the Cancer Genome Atlas. \*\*\*\*Indicates  $p$ -value  $< 0.0001$ ; \*\*indicates  $p$ -value  $< 0.01$ ; \*indicates  $p$ -value  $< 0.05$ .

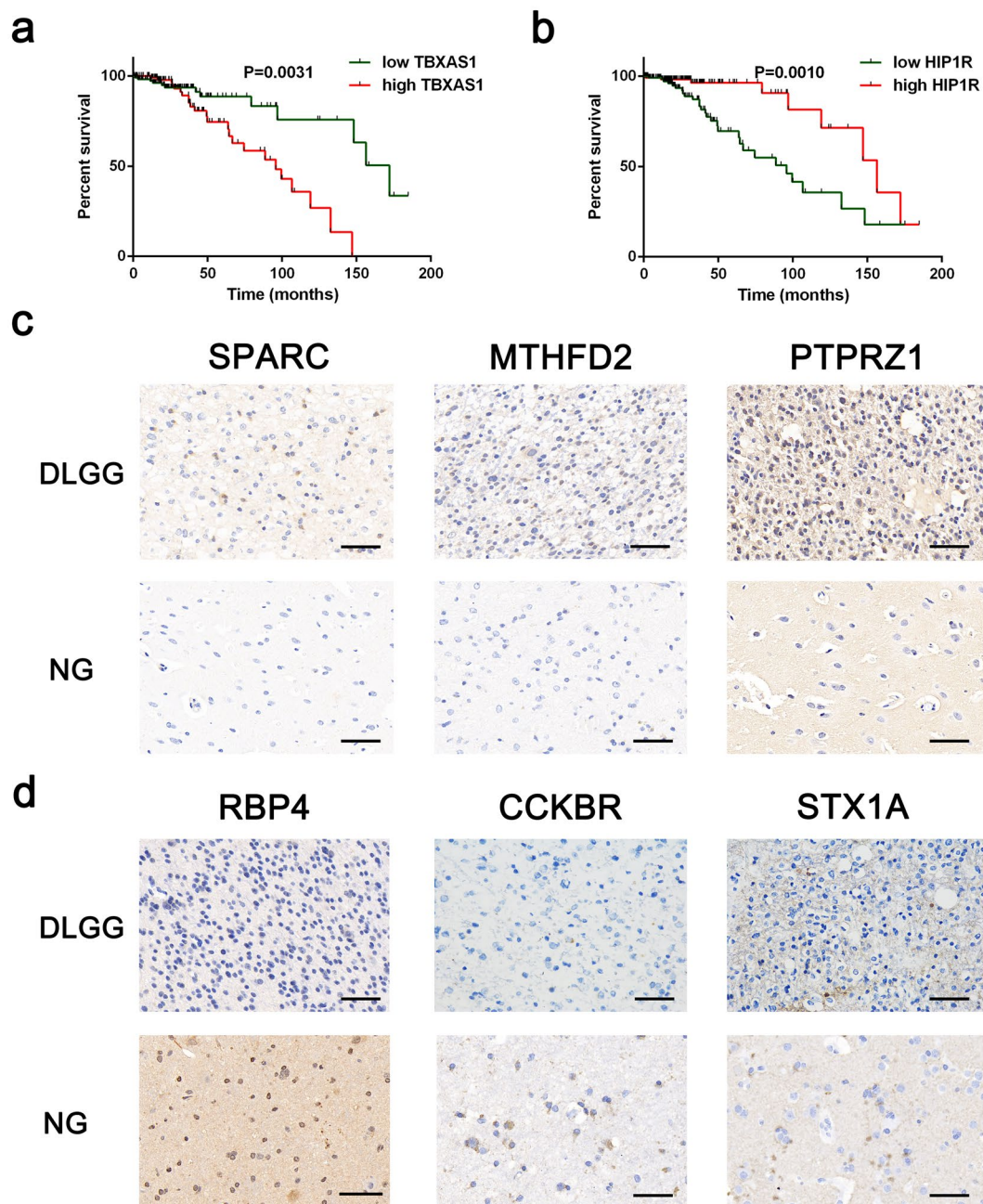
invasion, and metastasis<sup>40</sup>, and increased the expression of cancer stem cell markers<sup>40</sup>. Pikman *et al.* reported that knockdown of *MTHFD2* decreased cell growth, induced differentiation, and impaired colony formation in primary acute myeloid leukaemia (AML) blasts<sup>41</sup>. Despite the critical roles of *MTHFD2* in breast cancer and AML, the functions of this enzyme in DLGG have not been reported. The immunohistochemical results showed that *MTHFD2* was positively expressed in all DLGG samples, whereas control brain tissues demonstrated negative staining. *PTPRZ1* was previously reported to be upregulated in glioblastoma compared with normal brain at both the mRNA and protein levels<sup>42</sup>. This finding is similar to our current results showing that *PTPRZ1* was overexpressed in DLGG vs. NG tissues. *CCKBR*, also known as *CCK2R*, is widely expressed in the human gastrointestinal tract, pancreas, and lung, as well as in some neuroendocrine tissues<sup>43</sup>, and may regulate cellular proliferation, differentiation, and morphology<sup>44</sup>. However, the role of *CCKBR* in DLGG have not been uncovered. In our study, *CCKBR* was downregulated in DLGG vs. NG and Hs683 vs. HEB cells, suggesting that *CCKBR* may be a potential suppressor gene in DLGG development. *STX1A* is a member of the syntaxin superfamily, and studies have shown that the expression of this protein is correlated with Williams's syndrome, cystic fibrosis<sup>45</sup> and Alzheimer's disease<sup>46</sup>. Ulloa *et al.*<sup>47</sup> reported that inhibition of *STX1A* reduced glioblastoma tumour proliferation and cell invasion. In the present study, *STX1A* was downregulated in DLGG compared with control brain tissues. The association between *STX1A* and DLGG should be investigated further.

To identify possible DEGs between A and OD tissues, we performed a meta-analysis of five datasets. Across these datasets, we identified 497 genes that were consistently differentially expressed in A vs. OD tissues (222 upregulated and 275 downregulated). Among the upregulated DEGs, *BIN2*, also called *BRAP1*, had the highest combined ES of 1.69. *BIN2* is a putative membrane-remodelling protein located at chromosome 4q22.1 and is primarily expressed in hematopoietic cells<sup>48</sup>. Some studies have reported frequent disruptions in the 4q22.1 region in breast cancer and hepatocarcinoma<sup>49,50</sup>. However, Kai *et al.* did not identify any important functions associated with *BIN2* in the context of hepatoma cell growth<sup>48</sup>. Thus, the mechanism underlying the role of *BIN2* is unclear. *ELP3* is involved in various functions, such as transcriptional elongation, tRNA modification, histone acetylation, and cell migration<sup>51,52</sup>. Wang *et al.*<sup>53</sup> revealed that *ELP3* was significantly under-expressed in invasive ductal breast carcinoma compared to paired normal tissues. In the current study, *ELP3* had the highest combined ES (−1.44) among the downregulated DEGs in A vs. OD, suggesting that *ELP3* have the potential to function as a therapeutic biomarker of OD.

The expression levels of the top three upregulated and downregulated DEGs were also analysed using TCGA cohorts, and the results were consistent with those obtained from the meta-analysis. Moreover, Kaplan-Meier analysis revealed that high expression of *HIP1R* significantly correlated with longer overall survival, whereas high expression of *TBXAS1* significantly correlated with shorter overall survival. These findings confirm that these genes are potential prognostic biomarkers for DLGG as well as for distinguishing between astrocytic and oligodendroglial DLGG.

Network-based meta-analysis was performed on the list of DEGs to identify hub genes based on network centrality scoring. *FNI* and *APP* were the most important hub genes among the upregulated and downregulated genes, respectively, in the meta-analysis of DLGG. *FNI* encodes fibronectin, which is involved in cell adhesion and migration. Some studies have found that *FNI* is the hub gene in glioma, a result that is consistent with our findings; as such, *FNI* is a potential target for diagnosis and therapy<sup>54,55</sup>. *APP* mutations cause Alzheimer's disease, and





**Figure 6.** Kaplan-Meier analysis and immunohistochemistry. **(a,b)** Kaplan-Meier estimates of overall survival of patients with grade II gliomas. Red lines represent the high expression of DEGs and green lines represent the low expression of DEGs. The tick marks on the Kaplan-Meier survival curves represent the censored subjects. **(c)** Representative sections with immunohistochemical staining against the top 3 up-regulated DEGs in DLGG vs. NG. **(d)** Representative sections with immunohistochemical staining against the top 3 down-regulated DEGs in DLGG vs. NG. Bar = 50  $\mu$ m. Abbreviations: DLGG, diffuse low-grade glioma; NG, non-glioma; DEG, differentially expressed gene.

over-expression of *APP* has a link with shortened survival in patients with breast cancer<sup>56</sup>. Nizzari *et al.* showed that *APP* has a potential role in tumorigenesis mostly through its actions in activating extracellular signal-regulated protein kinase<sup>57</sup>. In the current study, *APP* was downregulated in DLGG relative to NG tissues, therefore the association between *APP* and DLGG should be explored further.

In the meta-analysis of DLGG subtypes, *PTPN6* and *CUL3* were the most important hub genes among upregulated and downregulated genes, respectively. *PTPN6* is a tyrosine phosphatase that participates in the regulation of numerous intracellular signalling cascades that modulate cell proliferation, differentiation, and apoptosis. In a report by Sooman *et al.*, high *PTPN6* expression was found to contribute to worse prognosis in patients with anaplastic glioma, and there was an association between high *PTPN6* expression and worse survival in a subgroup of patients with anaplastic oligodendroglioma ( $p = 0.053$ )<sup>58</sup>. *CUL3* mediates neurofibromin destabilization, which

underlies glioblastoma pathogenesis<sup>59</sup>. *CUL3* is also critical to the full activation of the Ras/ERK pathway, which is essential for cell proliferation, arrest, differentiation, survival, and apoptosis<sup>60</sup>.

According to the GO and KEGG pathway analysis, mRNAs were targeted to *pathway in cancer*, *MAPK signaling pathway*, *proteoglycans in cancer*, and *TGF-beta signaling pathway* in DLGG vs. NG tissues, which are known to contribute to tumorigenesis. Based on these findings, the GO categories and KEGG pathways identified in this study merit further study and validation.

Our study has several limitations. First, heterogeneity and confused factors may have distorted the analysis. To correct this limitation, we performed data preprocessing and normalization of each dataset as well as batch effect adjustment. Second, the samples included in the meta-analysis were relatively small, particularly for the NG tissues. The small sample size may not own enough power to detect true gene expression changes associated with DLGG. Third, we did not produce direct experimental evidence to verify the function of the identified DEGs, as the primary focus of our study was to perform meta-analyses of public datasets to discover novel or important mRNAs in DLGG and its histological subtypes. We have conducted immunohistochemical analysis and qPCR validation for six DEGs in DLGG vs. NG, however, the differences between cell lines and human tissues may lead to the contradictory results, so qPCR validation in human surgical specimens and more extensive investigations into these candidates will be performed in the future.

## Methods

**Identification of eligible gene expression datasets for low-grade glioma.** We searched the National Center for Biotechnology Information (NCBI) GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) to identify studies profiling low-grade glioma gene expression patterns. The keyword “glioma” was used for the search. Studies were included in the analysis if they met the following conditions: (1) they profiled gene expression, (2) they contained NG samples as well as at least one type of A and OD, (3) they contained more than three samples per type. We excluded studies examining non-human tissues or cell lines and studies not reporting microarray data. The following information was extracted from each identified study: GEO accession number, platform information, sample number, references, and gene expression data. We also collected DLGG samples from TCGA (<https://tcga-data.nci.nih.gov/tcga/>) and CGGA (<http://www.cgca.org.cn/>)<sup>25</sup>.

**Batch effect adjustment and individual data analysis.** The batch effect correction option in INMEX<sup>16</sup> (<http://www.networkanalyst.ca/faces/home.xhtml>) was applied to reduce potential study-specific batch effects. The ComBat procedures in INMEX were used to stabilize the expression ratios of genes with too high or too low ratios using Empirical Bayes methods and to stabilize individual gene variances by shrinking variances across all other genes<sup>61</sup>. Principal component analysis plots were used to visualize the sample clustering patterns before and after conducting the ComBat procedures.

**Meta-analysis of microarray datasets.** The INMEX program (<http://www.networkanalyst.ca/faces/home.xhtml>), a web interface for integrative meta-analysis<sup>16</sup>, was used to conduct the microarray meta-analysis of DLGG. Every dataset was preprocessed using log<sub>2</sub> transformation, and tables containing relative expression values were constructed showing gene information in rows and sample information in columns. Each dataset was visualized using box plots to ensure that identical distribution was present among the samples and to identify potential outliers. After uploading the datasets into the INMEX program, we annotated the data by converting different gene symbols to Entrez IDs. After all datasets were uploaded, processed, and annotated, the data integrity of each dataset was checked before proceeding to the meta-analysis stage. The meta-analysis of DLGG and NG tissues was conducted after estimating the ES to generate more biologically consistent results. We used a random effects model rather than a fixed effects model when the between-study heterogeneity based on Cochran's Q test was significant<sup>62</sup>. This method is based on the moderated ES and was performed using the metaMA package<sup>63</sup>. The two-way hierarchical clustering analysis of DEGs were conducted by the gplots package in the R software.

**Functional analysis.** To explore the functions of the identified DEGs, GO and KEGG pathway analyses were conducted using DAVID (<https://david.ncifcrf.gov/>). Significant GO terms and KEGG pathways were selected as the enriched terms based on values with  $p < 0.05$ .

**Network-Based meta-analysis.** Network-based meta-analysis was carried out using NetworkAnalyst<sup>64</sup>. The Hub Explorer in NetworkAnalyst contains detailed information on nodes within the current network, including degree, betweenness centrality, and expression<sup>65</sup>. The degree was defined as the number of connections to other nodes. The betweenness centrality was the number of shortest paths going through a node. The expression was defined as the log fold change value of the corresponding node. Nodes with the highest degree or betweenness values represent the critical hubs of the network.

**Kaplan-Meier analysis.** Kaplan-Meier survival curves were plotted using graphpad prism 6, which enables interactive exploration of survival correlations using the log-rank test. According to the median expression level of each DEG, we divided the patients with DLGG in the TCGA cohorts into groups with low and high DEG expression.  $P < 0.05$  was considered statistically significant.

**Clinical samples and immunohistochemistry.** Six DLGG samples and six NG samples were derived from patients undergoing surgical procedures at the Union Hospital of Tongji medical college, Wuhan, China. Human study protocols were approved by the ethical committee of Tongji Medical College, Huazhong University of Science and Technology in accordance with the Declaration of Helsinki. All patients in the study gave written informed consent.

Immunohistochemical analysis for SPARC, MTHFD2, PTPRZ1, RBP4, CCKBR, and STX1A on formalin-fixed, paraffin-embedded tissues were performed. Six DLGG and six NG tissues were embedded in paraffin wax and sections were deparaffinized with xylene, and rehydrated. The sections were immersed in a 0.01 mol/L citrate buffer solution at pH 6.0, and heated to repair antigen. After blocking the endogenous peroxidase activity with 3% H<sub>2</sub>O<sub>2</sub> for 25 minutes, the sections were treated for 30 minutes with 3% bovine serum albumin. The sections were incubated with anti-SPARC antibody (1:50; Santa Cruz), anti-MTHFD2 antibody (1:50, Santa Cruz), anti-PTPRZ1 antibody (1:100; Santa Cruz), anti-RBP4 antibody (1:50; Santa Cruz), anti-CCKBR antibody (1:100; Santa Cruz), and anti-STX1A antibody (1:100; Santa Cruz), and then incubated with HRP-conjugated secondary antibodies for 50 min. The slides were counterstained with Harris's hematoxylin, dehydrated through graded alcohols, and cleared in xylene prior to slide mounting.

**Cell culture and quantitative real-time polymerase chain reaction.** The low-grade Hs683 glioma cells and normal human glial HEB cells were obtained as gifts from Prof. Yiping Li (Institute of Human Virology, Zhongshan School of Medicine, Sun Yat-sen University North Campus). These cell lines were cultured in DMEM (Hyclone) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Invitrogen) at 37 °C with 5% CO<sub>2</sub>.

Expression of *SPARC*, *MTHFD2*, *PTPRZ1*, *RBP4*, *CCKBR*, and *STX1A* were detected by qPCR. Total RNA was extracted from Hs683 and HEB cells by Trizol reagent (Aidlab) according to the manufacturer's instruction. Then 1 µg of total RNA from each sample was reverse transcribed into cDNA, which was used as the template for PCR amplification, and *GAPDH* was used as standard control. The amplification program used was performed as follows: 95 °C for 10 min, followed by 40 cycles at 95 °C for 30 sec, 60 °C for 30 sec. PCR products were visualized with ethidium bromide on 1.5% agarose gel. The primers were listed in Supplementary Table S4. The expression of each mRNA relative to *GAPDH* was calculated as  $2^{-[(Ct \text{ of target genes}) - (Ct \text{ of GAPDH})]}$ .

**Statistical analysis.** ES is a standardized difference that considers both the direction and the magnitude of gene expression changes. We used a random effects model of ES combination for the meta-analysis, and a stringent threshold ( $p < 0.0001$ ) was used to identify DEGs. The Benjamini-Hochberg false discovery rate was used to correct the P values. Significantly enriched GO terms and pathways were identified using hypergeometric tests with  $p < 0.05$  as the threshold value. The statistical analysis of immunohistochemical results used the fisher's exact test ( $p < 0.05$  was significant). qPCR assay for the expression levels of DEGs in cell lines were calculated using one way ANOVA analysis.

**Data availability.** All data generated or analysed during this study are available from the corresponding author on reasonable request.

## References

- Louis, D. N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica*. **114**, 97–109, <https://doi.org/10.1007/s00401-007-0243-4> (2007).
- Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta neuropathologica*. **131**, 803–820, <https://doi.org/10.1007/s00401-016-1545-1> (2016).
- Leighton, C. *et al.* Supratentorial low-grade glioma in adults: an analysis of prognostic factors and timing of radiation. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. **15**, 1294–1301, <https://doi.org/10.1200/JCO.1997.15.4.1294> (1997).
- Macdonald, D. R. Low-grade gliomas, mixed gliomas, and oligodendrogliomas. *Seminars in oncology*. **21**, 236–248 (1994).
- Okamoto, Y. *et al.* Population-based study on incidence, survival rates, and genetic alterations of low-grade diffuse astrocytomas and oligodendrogliomas. *Acta neuropathologica*. **108**, 49–56, <https://doi.org/10.1007/s00401-004-0861-z> (2004).
- Ohgaki, H. & Kleihues, P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of neuropathology and experimental neurology*. **64**, 479–489 (2005).
- Kesari, S. *et al.* Phase II study of protracted daily temozolomide for low-grade gliomas in adults. *Clinical cancer research: an official journal of the American Association for Cancer Research*. **15**, 330–337, <https://doi.org/10.1158/1078-0432.CCR-08-0888> (2009).
- Hartmann, C. *et al.* Molecular markers in low-grade gliomas: predictive or prognostic? *Clinical cancer research: an official journal of the American Association for Cancer Research*. **17**, 4588–4599, <https://doi.org/10.1158/1078-0432.CCR-10-3194> (2011).
- Ramasamy, A., Mondry, A., Holmes, C. C. & Altman, D. G. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS medicine*. **5**, e184, <https://doi.org/10.1371/journal.pmed.0050184> (2008).
- Siddiqui, A. S. *et al.* Sequence biases in large scale gene expression profiling data. *Nucleic acids research*. **34**, e83, <https://doi.org/10.1093/nar/gkl404> (2006).
- Rung, J. & Brazma, A. Reuse of public genome-wide gene expression data. *Nature reviews. Genetics*. **14**, 89–99, <https://doi.org/10.1038/nrg3394> (2013).
- Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 8418–8423, <https://doi.org/10.1073/pnas.0932692100> (2003).
- Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D. & Chinnaiyan, A. M. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res*. **62**, 4427–4433 (2002).
- Hoshida, Y. *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res*. **69**, 7385–7392, <https://doi.org/10.1158/0008-5472.CAN-09-1089> (2009).
- Chen, R. *et al.* A meta-analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma. *Cancer Res*. **74**, 2892–2902, <https://doi.org/10.1158/0008-5472.CAN-13-2775> (2014).
- Xia, J. *et al.* INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic acids research*. **41**, W63–70, <https://doi.org/10.1093/nar/gkt338> (2013).
- Santiago, J. A. & Potashkin, J. A. Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, 2257–2262, <https://doi.org/10.1073/pnas.1423573112> (2015).

18. Jha, P. K., Vijay, A., Sahu, A. & Ashraf, M. Z. Comprehensive Gene expression meta-analysis and integrated bioinformatic approaches reveal shared signatures between thrombosis and myeloproliferative disorders. *Sci Rep.* **6**, 37099, <https://doi.org/10.1038/srep37099> (2016).
19. Madhavan, S. *et al.* Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res.* **7**, 157–167, doi:<https://doi.org/10.1158/1541-7786.MCR-08-0435> (2009).
20. Gravendeel, L. A. *et al.* Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* **69**, 9065–9072, <https://doi.org/10.1158/0008-5472.CAN-09-2307> (2009).
21. Sun, L. *et al.* Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell.* **9**, 287–300, <https://doi.org/10.1016/j.ccr.2006.03.003> (2006).
22. Liu, Z., Yao, Z., Li, C., Lu, Y. & Gao, C. Gene expression profiling in human high-grade astrocytomas. *Comparative and functional genomics.* **2011**, 245137, <https://doi.org/10.1155/2011/245137> (2011).
23. Bredel, M. *et al.* Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. *Cancer Res.* **65**, 8679–8689, <https://doi.org/10.1158/0008-5472.CAN-05-1204> (2005).
24. Rickman, D. S. *et al.* Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res.* **61**, 6885–6891 (2001).
25. Yan, W. *et al.* Molecular classification of gliomas based on whole genome gene expression: a systematic report of 225 samples from the Chinese Glioma Cooperative Group. *Neuro-oncology.* **14**, 1432–1440, <https://doi.org/10.1093/neuonc/nos263> (2012).
26. Cancer Genome Atlas Research, N. *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *The New England journal of medicine.* **372**, 2481–2498, <https://doi.org/10.1056/NEJMoa1402121> (2015).
27. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell.* **164**, 550–563, <https://doi.org/10.1016/j.cell.2015.12.028> (2016).
28. Eckel-Passow, J. E. *et al.* Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *The New England journal of medicine.* **372**, 2499–2508, <https://doi.org/10.1056/NEJMoa1407279> (2015).
29. Chan, A. K. *et al.* TERT promoter mutations contribute to subset prognostication of lower-grade gliomas. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc.* **28**, 177–186, <https://doi.org/10.1038/modpathol.2014.94> (2015).
30. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell.* **17**, 510–522, <https://doi.org/10.1016/j.ccr.2010.03.017> (2010).
31. Watanabe, T. *et al.* Aberrant hypermethylation of p14ARF and O6-methylguanine-DNA methyltransferase genes in astrocytoma progression. *Brain pathology.* **17**, 5–10, <https://doi.org/10.1111/j.1750-3639.2006.00030.x> (2007).
32. Hatanpaa, K. J. *et al.* High expression of the stem cell marker nestin is an adverse prognostic factor in WHO grade II-III astrocytomas and oligoastrocytomas. *Journal of neuro-oncology.* **117**, 183–189, <https://doi.org/10.1007/s11060-014-1376-7> (2014).
33. Huang, H. *et al.* Gene expression profiling of low-grade diffuse astrocytomas by cDNA arrays. *Cancer Res.* **60**, 6868–6874 (2000).
34. Uehara, H., Takahashi, T. & Izumi, K. Induction of retinol-binding protein 4 and placenta-specific 8 expression in human prostate cancer cells remaining in bone following osteolytic tumor growth inhibition by osteoprotegerin. *Int J Oncol.* **43**, 365–374, <https://doi.org/10.3892/ijo.2013.1954> (2013).
35. Abola, M. V. *et al.* Serum levels of retinol-binding protein 4 and risk of colon adenoma. *Endocrine-related cancer.* **22**, L1–4, <https://doi.org/10.1530/ERC-14-0429> (2015).
36. Chen, Y. *et al.* Identification of host-immune response protein candidates in the sera of human oral squamous cell carcinoma patients. *PLoS One.* **9**, e109012, <https://doi.org/10.1371/journal.pone.0109012> (2014).
37. Lorkova, L. *et al.* Decreased concentrations of retinol-binding protein 4 in sera of epithelial ovarian cancer patients: a potential biomarker identified by proteomics. *Oncol Rep.* **27**, 318–324, <https://doi.org/10.3892/or.2011.1513> (2012).
38. Liu, F. *et al.* Increased MTHFD2 expression is associated with poor prognosis in breast cancer. *Tumour Biol.* **35**, 8685–8690, <https://doi.org/10.1007/s13277-014-2111-x> (2014).
39. Selcuklu, S. D. *et al.* MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells. *The Journal of biological chemistry.* **287**, 29516–29528, <https://doi.org/10.1074/jbc.M111.335943> (2012).
40. Lehtinen, L. *et al.* High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion. *Oncotarget.* **4**, 48–63, <https://doi.org/10.18632/oncotarget.756> (2013).
41. Pikman, Y. *et al.* Targeting MTHFD2 in acute myeloid leukemia. *The Journal of experimental medicine.* **213**, 1285–1306, <https://doi.org/10.1084/jem.20151574> (2016).
42. Lu, K. V. *et al.* Differential induction of glioblastoma migration and growth by two forms of pleiotrophin. *The Journal of biological chemistry.* **280**, 26953–26964, <https://doi.org/10.1074/jbc.M502614200> (2005).
43. Zhang, R. *et al.* MiR-148a regulates the growth and apoptosis in pancreatic cancer by targeting CCKBR and Bcl-2. *Tumour Biol.* **35**, 837–844, <https://doi.org/10.1007/s13277-013-1115-2> (2014).
44. Bierkamp, C. *et al.* Expression of cholecystokinin-2/gastrin receptor in the murine pancreas modulates cell adhesion and cell differentiation *in vivo*. *The American journal of pathology.* **165**, 2135–2145, [https://doi.org/10.1016/S0002-9440\(10\)63263-4](https://doi.org/10.1016/S0002-9440(10)63263-4) (2004).
45. Gao, M. C. *et al.* Intelligence in Williams Syndrome is related to STX1A, which encodes a component of the presynaptic SNARE complex. *PLoS One.* **5**, e10292, <https://doi.org/10.1371/journal.pone.0010292> (2010).
46. Jacobs, E. H., Williams, R. J. & Francis, P. T. Cyclin-dependent kinase 5, Munc18a and Munc18-interacting protein 1/X11alpha protein up-regulation in Alzheimer's disease. *Neuroscience.* **138**, 511–522, <https://doi.org/10.1016/j.neuroscience.2005.11.017> (2006).
47. Ulloa, F. *et al.* Blockade of the SNARE protein syntaxin 1 inhibits glioblastoma tumor growth. *PLoS One.* **10**, e0119707, <https://doi.org/10.1371/journal.pone.0119707> (2015).
48. Ge, K. & Prendergast, G. C. Bin2, a functionally nonredundant member of the BAR adaptor gene family. *Genomics.* **67**, 210–220, <https://doi.org/10.1006/geno.2000.6216> (2000).
49. Schwendel, A. *et al.* Chromosome alterations in breast carcinomas: frequent involvement of DNA losses including chromosomes 4q and 21q. *British journal of cancer.* **78**, 806–811 (1998).
50. Rashid, A. *et al.* Genetic alterations in hepatocellular carcinomas: association between loss of chromosome 4q and p53 gene mutations. *British journal of cancer.* **80**, 59–66, <https://doi.org/10.1038/sj.bjc.6690321> (1999).
51. Svejstrup, J. Q. Elongator complex: how many roles does it play? *Current opinion in cell biology.* **19**, 331–336, <https://doi.org/10.1016/j.ceb.2007.04.005> (2007).
52. Close, P. *et al.* Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Molecular cell.* **22**, 521–531, <https://doi.org/10.1016/j.molcel.2006.04.017> (2006).
53. Wang, D. *et al.* Promoter hypermethylation may be an important mechanism of the transcriptional inactivation of ARRDC3, GATA5, and ELP3 in invasive ductal breast carcinoma. *Molecular and cellular biochemistry.* **396**, 67–77, <https://doi.org/10.1007/s11010-014-2143-y> (2014).
54. Long, H. *et al.* Prediction and Analysis of Key Genes in Glioblastoma Based on Bioinformatics. *Biomed Res Int.* **2017**, 7653101, <https://doi.org/10.1155/2017/7653101> (2017).
55. Jiang, C. M. *et al.* Analysis of differentially expressed genes based on microarray data of glioma. *International journal of clinical and experimental medicine.* **8**, 17321–17332 (2015).
56. Takagi, K. *et al.* Amyloid precursor protein in human breast cancer: an androgen-induced gene associated with cell proliferation. *Cancer science.* **104**, 1532–1538, <https://doi.org/10.1111/cas.12239> (2013).

57. Nizzari, M. *et al.* Amyloid precursor protein and Presenilin1 interact with the adaptor GRB2 and modulate ERK 1,2 signaling. *The Journal of biological chemistry*. **282**, 13833–13844, <https://doi.org/10.1074/jbc.M610146200> (2007).
58. Sooman, L. *et al.* PTPN6 expression is epigenetically regulated and influences survival and response to chemotherapy in high-grade gliomas. *Tumour Biol.* **35**, 4479–4488, <https://doi.org/10.1007/s13277-013-1590-5> (2014).
59. McGillicuddy, L. T. *et al.* Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer Cell*. **16**, 44–54, <https://doi.org/10.1016/j.ccr.2009.05.009> (2009).
60. Hollstein, P. E. & Cichowski, K. Identifying the Ubiquitin Ligase complex that regulates the NF1 tumor suppressor and Ras. *Cancer Discov.* **3**, 880–893, <https://doi.org/10.1158/2159-8290.CD-13-0146> (2013).
61. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. **8**, 118–127, <https://doi.org/10.1093/biostatistics/kxj037> (2007).
62. Cochran, W. G. The combination of estimates from different experiments. *Biometrics*. **10**, 101–129, <https://doi.org/10.2307/3001666> (1954).
63. Marot, G., Foulley, J. L., Mayer, C. D. & Jaffrezic, F. Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics*. **25**, 2692–2699, <https://doi.org/10.1093/bioinformatics/btp444> (2009).
64. Xia, J., Benner, M. J. & Hancock, R. E. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic acids research*. **42**, W167–174, <https://doi.org/10.1093/nar/gku443> (2014).
65. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*. **430**, 88–93, <https://doi.org/10.1038/nature02555> (2004).

## Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (81372369 to HX; 81371661 to PH; and 81101042 to FL). We thank the department of pathology in Union hospital of Tongji medical college for providing help for this study. We also thank Prof. Yiping Li for having supplied us Hs683 and HEB cell lines necessary to carry on this work.

## Author Contributions

Conceived and designed the study: S.W. and P.H. Downloaded, analyzed the data and wrote the paper: S.W., W.F., F.J., and F.L.; S.W., X.H., Y.Z., and H.X. edited the manuscript and contributed to data interpretation and results discussion.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12087-y>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017