Identification of differentially expressed key genes between glioblastoma and low-grade glioma by bioinformatics analysis

Yang Xu [1, 2] , **Rongxin Geng** [1, 2] , **Fan'en Yuan** [1, 2] , **Qian Sun** [1, 2] , **Baohui Liu** [Corresp., 1, 2] , **Qianxue Chen** [Corresp. 1, 2]

[1] Department of Neurosurgery, Renmin Hospital of Wuhan University, Wuhan, Hubei, China

[2] Brain Tumor Clinical Center of Wuhan, Wuhan, Hubei, China

Corresponding Authors: Baohui Liu, Qianxue Chen
Email address: bliu666@whu.edu.cn, chenqx666@whu.edu.cn

Gliomas are a very diverse group of brain tumors that are most commonly primary tumor and difficult to cure in central nervous system. It’s necessary to distinguish low-grade tumors from high-grade tumors by understanding the molecular basis of different grades of glioma, which is an important step in defining new biomarkers and therapeutic strategies. We have chosen gene expression profile of GSE52009 from GEO database to detect important differential genes. GSE52009 contains 120 samples, including 60 WHO II samples and 24 WHO IV samples, were selected in our analysis. We used GEO2R tool to pick out differently expressed genes (DEGs) between low-grade glioma and high-grade glioma, and then we used The Database for Annotation, Visualization and Integrated Discovery (DAVID) to perform gene ontology (GO) analysis and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis. Furthermore, Cytoscape with Search Tool for the Retrieval of Interacting Genes (STRING) and Molecular Complex Detection (MCODE) plug-in were applied to achieve the visualization of protein-protein interaction (PPI). We selected 15 hub genes with higher degree of connectivity, including TIMP1 and SAA1 and we used GSE53733 containing 70 glioblastoma samples to conduct Gene Set Enrichment Analysis on. In conclusion, DEGs and hub genes were demonstrated in our bioinformatics analysis, that they might be defined as new biomarkers for diagnosis and to guide the therapeutic strategies of glioblastoma.

Identification of differentially expressed key genes between glioblastoma and low-grade glioma by bioinformatics analysis

Yang Xu [1,2], Rongxin Geng [1,2], Fan'en Yuan [1,2], Qian Sun [1,2], Baohui Liu [1,2*], Qianxue Chen [1,2*]

[1]Department of Neurosurgery, Renmin Hospital of Wuhan University, Hubei 430060, China

[2]Brain Tumor Clinical Center of Wuhan, Hubei 430060, China

Correspondence should be addressed to Baohui Liu; bliu666@whu.edu.cn and Qianxue Chen; chenqx666@whu.edu.cn

Abstract

Gliomas are a very diverse group of brain tumors that are most commonly primary tumor and difficult to cure in central nervous system. It's necessary to distinguish low-grade tumors from high-grade tumors by understanding the molecular basis of different grades of glioma, which is an important step in defining new biomarkers and therapeutic strategies. We have chosen gene expression profile of GSE52009 from GEO database to detect important differential genes. GSE52009 contains 120 samples, including 60 WHO II samples and 24 WHO IV samples, were selected in our analysis. We used GEO2R tool to pick out differently expressed genes (DEGs) between low-grade glioma and high-grade glioma, and then we used The Database for Annotation, Visualization and Integrated Discovery (DAVID) to perform gene ontology (GO) analysis and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis. Furthermore, Cytoscape with Search Tool for the Retrieval of Interacting Genes (STRING) and Molecular Complex Detection (MCODE) plug-in were applied to achieve the visualization of protein-protein

interaction (PPI). We selected 15 hub genes with higher degree of connectivity, including TIMP1 and SAA1 and we used GSE53733 containing 70 glioblastoma samples to conduct Gene Set Enrichment Analysis on. In conclusion, DEGs and hub genes were demonstrated in our bioinformatics analysis, that they might be defined as new biomarkers for diagnosis and to guide the therapeutic strategies of glioblastoma.

1. Introduction

Gliomas are a very diverse group of brain tumors that are most commonly primary tumor and difficult to cure in central nervous system (Louis et al. 2016; Ostrom et al. 2017). They are classified according to their clinical and histopathological characteristics in four grades, including Low-grade gliomas - (1) grade I astrocytomas – pilocytic astrocytomas, (2) grade II diffuse astrocytomas, and (3) grade II oligodendrogliomas; High-grade gliomas - (1) Grade III anaplastic astrocytomas, (2) Grade III anaplastic oligodendrogliomas and (3) grade IV glioblastomas multiforme(GBM) (Sriram & Huse 2015). Low grade gliomas(LGG) (astrocytomas, oligodendrogliomas and oligoastrocytomas) are considered relatively benign, well-differentiated tumors and have 5-year survival rates of 59.9% (Elizabeth & Peter 2010). And GBM is the most common primary malignant brain tumor in adults(Ramos et al. 2018). Despite the multiple therapeutic strategies, including surgery, radiation, and chemotherapy, the average survival time of GBM patients is less than 15 months (Liu et al. 2016). Additionally, approximately 70% of LGG patients develop to GBM within 5–10 years as ending (Furnari et al. 2007). With the development of molecular pathology in gliomas, several biomarkers routinely applied to evaluate gliomas include MGMT promoter methylation, EGFR alterations, IDH1 or IDH2 mutations, and 1p19q co-deletion and many of these markers have become standard of care for molecular testing and prerequisites for clinical trial enrollment(Rodriguez et al. 2016). Therefore, it's necessary to differentiate LGG to GBM by understanding the molecular basis of different grades of glioma, which is an important step in defining new biomarkers and therapeutic strategies.

Gene expression profiling analysis is a useful method with broad clinical application for identifying tumor-related genes in various types of cancer, from molecular diagnosis to pathological classification, from therapeutic evaluation to prognosis prediction, and from drug sensitivity to neoplasm recurrence(De Preter et al. 2010; Freije et al. 2004; Kim et al. 2011; Kulasingam & Diamandis 2008). In recent years, large scales of gene profiling have been made to identify the overwhelming number of genes by the use of microarrays in clinical practice, and complicated and systemic statistical analyses should be made to provide both repeatability and independent validation (Cheng et al. 2016).

In this analysis, GEO2R online tool was applied to look for the differentially expressed genes (DEGs) according to GSE52009 from Gene Expression Omnibus (GEO). Followed by, we produced a heatmap and picked out 15 genes with higher degree of connectivity from the DEGs selected. Furthermore, we analyzed cellular component (CC), biological process (BP), molecular function (MF) and KEGG pathways of the DEGs. In addition, the overall survival (OS) analysis and expression of these hub genes were mad e online. Then, we established PPI network of the DEGs and managed a GSEA using GBM patient gene profiling data (GSE53733).

2. Materials and Methods

2.1. Data of Microarray. Gene expression profile of GSE52009, GSE53733 and GSE4290 were downloaded from GEO database, which is a public and freely accessible database. Based on Agilent GPL6480 platform (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F), GSE52009 dataset included 120 samples, containing 60 WHO II samples and 24 WHO IV samples. And GSE53733 was based on GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), which contained 70 GBM samples. And GSE4290 was based on GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), which included 180 samples, containing 76 WHO II samples and 81 WHO IV samples.

2.2. Screen Genes of Differential Expression. Differentially expressed genes between low grade glioma and high-grade glioma was detected by GEO2R, which was an online analysis tool based on R language (Davis & Meltzer 2007). We set the adjust P value < 0.05 and |logFC| ≥ 2 as the selection criteria to decrease the false positive rate and false discovery rate. Furthermore, the top 15 genes with higher degree of connectivity were selected as hub genes among the 133 discovered DEGs which includes 56 down-regulated genes and 77 up-regulated genes. In addition, we used visual hierarchical cluster analysis to show the two groups by Morpheus online analysis software (https://software.broadinstitute.org/morpheus/) and volcano plot of two groups by ImageGP (http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html) after the relative raw data of TXT files were downloaded.

2.3. Gene Ontology and KEGG Pathway Analysis of DEGs. With functions including molecular function, biological pathways, and cellular component, gene ontology (GO) analysis we annotated genes and gene products (2006). KEGG. comprises a set of genome and enzymatic approaches and abiological chemical energy online database (Kanehisa & Goto 2000). It is a resource for systematic analysis of gene function and related high-level genome functional information. DAVID (https://david.ncifcrf.gov/) can provide systematic and comprehensive biological function annotation information for high-throughput gene expression (Dennis et al. 2003). Therefore, we applied GO and KEGG pathway analyses to the DEGs by using DAVID online tools at functional level. We considered P < 0.05 had significant differences. In addition, we used visual analysis to show GO Enrichment plot of two groups by ImageGP (http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html) after the relative raw data of TXT files were downloaded (Geng et al. 2018).

2.4. PPI network and module analysis. The online tool, Search Tool for the Retrieval of Interacting Genes (STRING), is designed to demonstrate the interaction between different proteins(Szklarczyk et al. 2015). STRING in Cytoscape was applied and mapped the DEGs into

STRING to detect the possible relationship among the selected DEGs. We set the confidence score ≥ 0.4, maximum number of interactors = 0 as the selection criteria. In addition, the Molecular Complex Detection (MCODE) was used to screen modules of PPI network in Cytoscape with degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. depth = 100. DAVID was used to perform the signal pathway analysis of genes in the module. 15 hub genes were also mapped into STRING with confidence score ≥ 0.4, maximum number of interactors ≤ 5. The potential information was explored through GO and KEGG pathway analysis.

2.5. Comparing the expression level of the hub genes. The GlioVis (http://gliovis.bioinfo.cnio.es/) is a user-friendly web application for data visualization and analysis to explore brain tumors expression datasets, which was used to analyze the gene expression data of brain tumors and normal samples based on the TCGA datasets. (Bowman et al. 2017). The customizable functions are provided such as analyzing the differences of expression levels between glioblastoma and low-grade glioma, so the expression of these genes was demonstrated. And the relationship could be visualized through the boxplot. All values are presented as the mean±SD. All statistical analyses were performed by SPSS 19.0 software. A difference of $P < 0.05$ was considered statistically significant.

2.6. Gene expression profile and gene set enrichment analysis. The expression profiles of GSE53733 was downloaded from the Gene Expression Omnibus (GEO). And we used gene set enrichment analysis (GSEA) (http://www.broadinstitute.org/gsea) to detect the potential genes influenced by SAA1 and TIMP1 through the Java programming. According to their hub genes expression level (top 50%: high vs bottom 50%: low), we divided the patients into two groups and GSEA was conducted to analyze the effects of selected genes expression level on different biological process. And we set P value of <0.05 and false discovery rates of <0.25 as selection criteria to confirm significant gene sets.

135 *2.7. Human tissue samples.* LGG and GBM tissues were collected from the Department of
136 Neurosurgery, Renmin Hospital of Wuhan University, Wuhan, China. And the clinical glioma
137 specimens were examined and diagnosed by pathologists at Renmin Hospital of Wuhan
138 University. This study was approved by the Institutional Ethics Committee of the Faculty of
139 Medicine at Renmin Hospital of Wuhan University (approval number: 2012LKSZ (010) H).
140 Informed consent was obtained from all patients whose tissues were used.
141
142 *2.8. RNA extraction and quantitative real-time PCR.* Total RNA from cancer tissues was
143 prepared using Trizol reagent (Invitrogen, USA), and cDNA was synthesized using a Pri
144 meScript RT Reagent Kit with gDNA Eraser (RR047A, Takara, Japan). Quantitative real-
145 time PCR (qPCR) for SAA1 and TIMP1 mRNA levels were performed using SYBR Pre
146 mix Ex Taq II (RR820A, Takara) according to the manufacturer's instructions and perfor
147 med in Bio-Rad CFX Manager 2.1 real-time PCR Systems (Bio-Rad, USA). GAPDH was
148 used as internal controls. The data were analyzed by the relative Ct method and expresse
149 d as a fold change compared with the control. The primer sequences included the follow
150 ing: GAPDH 5'-GGAGCGAGATCCCTCCAAAAT-3'(Forward), 5'-GGCTGTTGTCATACT
151 TCTCATGG-3'(Reverse); SAA1 5'-CCTGGGCTGCAGAAGTGATCAGCGA-3'(Forward),5'
152 -AGTCCTCCGCACCATGGCCAAAGAA-3'(Reverse); TIMP1 5'-CTTCTGCAATTCCGAC
153 CTCGT-3'(Forward),5'-ACGCTGGTATAAGGTGGTCTG-3'(Reverse).
154

155 3. Results

156
157 *3.1. Identification of DEGs and hub genes.* 60 WHO II samples and 24 WHO IV samples from
158 GSE52009 were selected in this study. DEGs were detected by applying the GEO2R online
159 analysis tool, setting adjust P value < 0.05 and $|\log FC| \geq 2$ as selection criteria. 133 differential
160 expressed genes, containing 77 up-regulated genes and 56 down-regulated genes, were detected
161 after the analysis of GSE52009. In addition, we selected 15 hub genes with higher degree of

connectivity (Table 1). The results were validated with a DEG expression heatmap and volcano plot of the all downregulated genes and upregulated genes (Figure 1).

3.2. GO function and KEGG pathway enrichment analysis. To explore more particular knowledge of the selected DEGs, we used DAVID to gain the results of GO function and KEGG pathway enrichment analysis. All DEGs were imported to DAVID software and GO analysis results demonstrated that up-regulated and down-regulated DEGs were particularly enriched in the following biological processes (BP): cell migration, locomotion and leukocyte migration, cell motility for up-regulated DEGs, and for downregulated DEGs nervous system development, brain development and regulation of cell projection organization (Table 2, Figure2a, Figure2b). The upregulated DEGs were enriched in phospholipase A2 inhibitor activity, growth factor binding, extracellular matrix structural construction, receptor binding, and the down-regulated DEGs were enriched in calcium ion binding, structural construction of myelin sheath, and protein complex binding for molecular function (MF) (Table 2, Figure2). Moreover, GO cell component (CC) analysis showed that the up-regulated DEGs were enriched in the proteinaceous extracellular matrix, extracellular matrix and cytoplasmic membrane-bounded vesicle lumen, and down-regulated DEGs enriched in neuron part, myelin sheath, and internode region of axon (Table 2, Figure2a, Figure2b).

Interestingly, the most significantly enriched of KEGG pathway only showed in up-regulated pathway, including adhesion, ECM-receptor interaction, amoebiasis and PI3K-Akt signaling pathway (Table3, Figure2c).

3.3. Hub Genes and Module Screening from PPI Network. PPI network of the top 15 hub genes with higher degree of connectivity was made based on the information in the STRING protein query from public databases (Figure 3A). The top module was selected by using MCODE plug-in in the PPI network. (Figure 3B).

189 *3.4. The Kaplan-Meier Plotter of hub genes.* The website, http:// gepia.cancer-pku.cn/, freely
190 provides the prognostic data of the hub genes. It was found that expression of VEGFA (HR 4.2,
191 p<0.001) was associated with worse overall survival (OS) for glioblastoma patients, as well as
192 NDC80 (HR 5.8, p<0.001), CENPA (HR 5.3, p<0.001), CENPF (HR 3.9, p<0.001), NCAPG (HR
193 5.6, p<0.001), ASPM (HR 5, p<0.001), ITGA2 (HR 3, p<0.001), TIMP1 (HR 7, p<0.001)and
194 SAA1 (HR 4.8, p<0.001). (Figure4).
195
196 *3.5. Expression Level and relationship with molecular pathologic diagnosis of Hub genes.* We
197 used data from GlioVis to detect the hub gene expression level between GBM and LGG including
198 astrocytoma, oligodendroglioma and oligoastrocytoma, the expression level of SAA1 and TIMP1
199 significantly increased in GBM (Figure 5A, 5C). And the expression level of SAA1 have no
200 significant difference in three kinds LGG (Figure 5B). However, the expression level of TIMP1 is
201 significantly higher in astrocytoma than oligodendroglioma and oligoastrocytoma (Figure 5D). We
202 further verified our finding in GSE4290 dataset and got consistent result (FigureS1). Then we
203 detected the sample collected in our hospital and found both SAA1 and TIMP1 are significantly
204 increased in GBMs compared with LGG (FigureS2, TableS1). We also detect the relationship
205 between expression level and molecular pathologic diagnosis of hub genes. We found both SAA1
206 and TIMP1 increase in both IDH mutant IDH wild type. The same results were found in MGMT
207 promoter and non-deletion of chromosome 1p.19q. Because of the limited samples in the datasets,
208 we didn't the result of co-deletion of chromosome 1p.19q. (Figure 5E,5F) What's more, we also
209 found that both SAA1 and TIMP1 played important roles in MES-like in the IDH wild type.
210 (Figure 5G,5H)
211
212 *3.6. Gene expression profile and gene set enrichment analysis.* We managed a GSEA by using
213 GBM patient gene profiling data (GSE53733), and showed in figure 5, gene set differences in
214 SAA1 in low versus high glioma patients indicated that SAA1 regulates biology process mainly
215 associated with inflammatory response processes (P<0.001 FDR=0.012) and cytokine mediated

signaling pathway (P<0.001 FDR=0.012) (Figure5I, 5J). We considered that SAA1 may negatively regulate inflammatory response and might promote the survival of cancer cells. Seemingly we concluded that TIMP1 might negatively regulates adaptive immune response based on somatic recombination of immune receptors built from leucine-rich (P<0.001 FDR=0.021) and response to interferon (P<0.001 FDR=0.027) may promote the survival of cancer cells. (Figure5K, 5L)

4. Discussion

In this study, we identified 15 DEGs between GBM and LGG and used a series of bioinformatics analyses to screen the key genes and pathways associated with glioma. GSE52009 dataset contains 60 WHO II samples and 24 WHO IV samples. In order to improve the statistical power of DEGs, we defined that the absolute value of the logarithm (base 2) fold change (logFC) greater than 2 and 133 DEGs were obtained. Bioinformatics analysis on DEGs, including GO enrichment, KEGG pathway, PPI network, and survival analysis, expression level, gene set enrichment analysis, found GBM related genes and pathways that play important roles in glioma development.

The GO analysis showed that the up-regulated DEGs were mainly associated with cell migration, extracellular matrix structural construction, cell motility and down-regulated DEGs were involved central nervous system development, calcium ion binding and internode region of axon. Additionally, the KEGG pathways of up-regulated DEGs regulate focal adhesion, ECM-receptor interaction, PI3K-Akt signaling pathway. Among these DEGs, we selected 15 hub genes with higher degree of connectivity. In addition, we found several hub genes with worse overall survival (OS) and higher expression level in glioma patients, including VEGFA, NDC80, TIMP1, SAA1, CENPA, CENPF and NCAPG and we firstly found relationship of SAA1, TIMP1 and molecular pathology in GBM. We could hypothesize that these genes might contribute to the malignance of glioma and SAA1 and TIMP-1 may be biomarkers in GBM.

GBMs are highly vascularized cancers with high levels of VEGF and VEGF-A seems to be the most critical one, mainly operating in the activation of quiescent endothelial cells and promoting cell migration and proliferation (Plate & Dumont 2012). NDC80 is a mitotic regulator and a major element of outer kinetochore which has been reported to drive functions in assembly checkpoint and chromosome segregation of mitosis regulation. NDC80 was mainly enriched in proliferation and procession of cancer in previous studies(Suzuki et al. 2015). Addition, a recently study demonstrated that the expression of NDC80 in HEB was significantly lower than in GBM cell lines and had a negative correlation with the prognosis of patients (Zhong et al. 2018). Interleukin (IL)-8 is a chemokine which was upregulated by NF-κB in GBMs and promotes a more aggressive phenotype mostly through the enhancement of angiogenesis and cell migration. And more and more evidence demonstrated that the IL-8 molecular pathways will allow the generation of both novel therapeutic approaches and diagnostic tools(Kosmopoulos et al. 2017). Non-SMC condensin I complex subunit G (NCAPG)8 is a novel mitotic gene for cell proliferation and migration, which has been less well studied in cancers, and a recently study demonstrated that NCAPG over expressed in GBMs and promote cell proliferation(Liang et al. 2016).

Serum amyloid A1 (SAA1) which was secreted by liver is an acute-phase high density lipo-protein in immune response. Injury, inflammation, and brain trauma can elevate the plasma levels of SAA1 (Lu et al. 2014; Villapol et al. 2015). What's more, it has long been suspected that SAA1 might be a prognostic marker and predictor of cancer risk. Elevated levels of SAA1 in the serum of cancer patients directly correlate with poor prognosis and tumor aggressiveness in various types of cancer, including lung cancer(Cho et al. 2010), cell renal carcinoma(Kosari et al. 2005), melanomas(Findeisen et al. 2009) and so on. Normal brain does not express SAA1(Liang et al. 1997), though an in vitro study demonstrated that microglia and astrocytes are responsive to SAA (Yu et al. 2014). Recently, it has been reported that SAA1 the expression levels in GBM patients are upregulated on both mRNA and protein in human GBMs, and SAA1 involves in angiogenesis via HIF-1α and tumor associated macrophages. And serum levels of SAA1 were associated with the grades of gliomas but did not affect the clinical outcomes of patients with GBM(Knebel et al.

2017). Consistently, SAA1 has been reported as a molecular/metabolic signature that can help identify patients are at high risk of malignant GBM and promotes glioma cells migration and invasion through integrin aVb3(Lin et al. 2018). What's more, although it's unknown why SAA1 upregulated in GBM and other malignant cancers, it has been speculated that SAA proteins play a primary role in the regulation of immunity and invasion processes (Moshkovskii 2012), which is consistent with the result of our study. Thus, we hypothesis SAA1 could be a biomarker of GBM and predict the prognosis of GBM patients. And the mechanism of SAA1 regulate in GBM need further research.

The tissue inhibitors of metalloproteinases (TIMPs, including TIMP-1, TIMP-2, TIMP-3, TIMP-4) are well known play critical roles in both metastasis and invasion through extracellular matrix remodeling which are controlled by the activity of matrix metalloproteinases (MMPs) (Jackson et al. 2017; Ries 2014). Among the four well-known TIMPs characterized so far, most focus has been on TIMP-1 defined as a naturally occurring inhibitor of most MMPs, a family of zinc dependent endopeptidases essential for degrading components of the ECM(Aaberg-Jessen et al. 2009). In addition, TIMP1 shows protease-independent function including anti-apoptotic, antiangiogenic, and differentiation activities in cells (Bridoux et al. 2013; Mandel et al. 2017). Over the past year, more and more studies focus on the influence of TIMP1 in cancers. Serum or urine levels of TIMP1 are also considered as a diagnostic predictor in pancreatic ductal carcinomas containing extensive desmoplasia(Jenkinson et al. 2015; Roy et al. 2014). Increased levels of cytosolic TIMP1 in pretreatment tumor tissue is associated with a significantly shorter overall survival in patients with breast cancer receiving standard adjuvant chemotherapy(Dechaphunkul et al. 2012). And it has been reported that low expression of TIMP-1 in glioblastoma patient predicts longer survival. The shorter survival of glioblastoma patients with a high tumor TIMP-1 level may be explained by the antiapoptotic effect of TIMP-1 preventing apoptosis induced by radiation and chemotherapy(Aaberg-Jessen et al. 2009). More recently, Aaberg-Jessen C et al demonstrated that Co-expression of TIMP-1 and CD63 might have effects in glioblastoma stemness and may predict the poor prognosis of patients through influencing tumor aggressiveness

297 and resistance of therapy(Aaberg-Jessen et al. 2018). We consider TIMP-1 can be identified as a
298 future biomarker for prognosis or monitoring patients' treatment response. However, all these
299 studies didn't demonstrate the specific mechanism, which is the direction for further researches.
300 Additionally, large-scale efforts aimed at characterizing the genomic alterations in human
301 glioblastoma, however, these efforts helped to clarify the role of genomic alterations in the
302 pathogenesis of glioblastoma but were not designed to address intratumor heterogeneity. Recently,
303 Ralph B. et al. described the Ivy Glioblastoma Atlas in which we have assigned key genomic
304 alterations and gene expression profiles to the tumor's anatomic features(Puchalski et al. 2018).
305 The anatomic feature included the leading edge (LE), infiltrating tumor (IT), cellular tumor (CT).
306 And we found the expression levels of SAA1 and TIPM-1 were higher in IT and CT than LE from
307 the atlas. However, the specific mechanisms of these differences aren't clear now, we need to do
308 further research in the future.
309

310 ## 5. Conclusion

311

312 We presumed these key genes identified by a series of bioinformatics analyses on DEGs
313 between glioblastoma samples and low-grade glioma samples, probably related to the development
314 of glioma. These hub genes could also affect the survival time of glioma patients. These identified
315 genes and pathways provide a more detailed molecular mechanism for underlying glioma initiation
316 and development. According to the study, SAA1 and TIMP1 can be considered as biomarkers or
317 therapeutic targets or monitoring patients' treatment response for glioblastoma. However, further
318 molecular and biological experiments are required to confirm the functions of the key genes in
319 glioblastoma.
320
321

322 ## References

323

2006. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res* 34:D322-326. 10.1093/nar/gkj021

Aaberg-Jessen C, Christensen K, Offenberg H, Bartels A, Dreehsen T, Hansen S, Schroder HD, Brunner N, and Kristensen BW. 2009. Low expression of tissue inhibitor of metalloproteinases-1 (TIMP-1) in glioblastoma predicts longer patient survival. *J Neurooncol* 95:117-128. 10.1007/s11060-009-9910-8

Aaberg-Jessen C, Sorensen MD, Matos A, Moreira JM, Brunner N, Knudsen A, and Kristensen BW. 2018. Co-expression of TIMP-1 and its cell surface binding partner CD63 in glioblastomas. *BMC Cancer* 18:270. 10.1186/s12885-018-4179-y

Bowman RL, Wang Q, Carro A, Verhaak RGW, and Squatrito M. 2017. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro Oncol* 19:139-141.

Bridoux L, Etique N, Lambert E, Thevenard J, Sowa ML, Belloy N, Dauchez M, Martiny L, and Charpentier E. 2013. A crucial role for Lyn in TIMP-1 erythroid cell survival signalling pathway. *FEBS Lett* 587:1524-1528. 10.1016/j.febslet.2013.03.032

Cheng W, Ren X, Zhang C, Cai J, Liu Y, Han S, and Wu A. 2016. Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology* 86:2226-2234. 10.1212/wnl.0000000000002770

Cho WC, Yip TT, Cheng WW, and Au JS. 2010. Serum amyloid A is elevated in the serum of lung cancer patients with poor prognosis. *Br J Cancer* 102:1731-1735. 10.1038/sj.bjc.6605700

Davis S, and Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23:1846-1847. 10.1093/bioinformatics/btm254

De Preter K, Vermeulen J, Brors B, Delattre O, Eggert A, Fischer M, Janoueix-Lerosey I, Lavarino C, Maris JM, Mora J, Nakagawara A, Oberthuer A, Ohira M, Schleiermacher G, Schramm A, Schulte JH, Wang Q, Westermann F, Speleman F, and Vandesompele J. 2010. Accurate outcome prediction in neuroblastoma across independent data sets using a multigene signature. *Clin Cancer Res* 16:1532-1541. 10.1158/1078-0432.ccr-09-2607

Dechaphunkul A, Phukaoloun M, Kanjanapradit K, Graham K, Ghosh S, Santos C, and Mackey JR. 2012. Prognostic significance of tissue inhibitor of metalloproteinase-1 in breast cancer. *Int J Breast Cancer* 2012:290854. 10.1155/2012/290854

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, and Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:P3.

Elizabeth BCMD, and Peter MBPD. 2010. Survival rates and patterns of care for patients diagnosed with supratentorial low-grade gliomas. *Cancer* 106:1358-1363.

Findeisen P, Zapatka M, Peccerella T, Matzk H, Neumaier M, Schadendorf D, and Ugurel S. 2009. Serum amyloid A as a prognostic marker in melanoma identified by proteomic profiling. *J Clin Oncol* 27:2199-2208. 10.1200/jco.2008.18.0554

Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, Mischel PS, and Nelson SF. 2004. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64:6503-6510. 10.1158/0008-5472.can-04-0452

Furnari FB, Fenton T, Bachoo RM, Mukasa A, Stommel JM, Stegh A, Hahn WC, Ligon KL, Louis DN, Brennan C, Chin L, DePinho RA, and Cavenee WK. 2007. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* 21:2683-2710. 10.1101/gad.1596707

Geng R-X, Li N, Xu Y, Liu J-h, Yuan F-e, Sun Q, Liu B-H, and Chen Q-X. 2018. Identification of Core Biomarkers Associated with Outcome in Glioma: Evidence from Bioinformatics Analysis. *Disease Markers* 2018:16.

10.1155/2018/3215958

Jackson HW, Defamie V, Waterhouse P, and Khokha R. 2017. TIMPs: versatile extracellular regulators in cancer. *Nat Rev Cancer* 17:38-53. 10.1038/nrc.2016.115

Jenkinson C, Elliott V, Menon U, Apostolidou S, Fourkala OE, Gentry-Maharaj A, Pereira SP, Jacobs I, Cox TF, Greenhalf W, Timms JF, Sutton R, Neoptolemos JP, and Costello E. 2015. Evaluation in pre-diagnosis samples discounts ICAM-1 and TIMP-1 as biomarkers for earlier diagnosis of pancreatic cancer. *J Proteomics* 113:400-402. 10.1016/j.jprot.2014.10.001

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

Kim YW, Liu TJ, Koul D, Tiao N, Feroze AH, Wang J, Powis G, and Yung WK. 2011. Identification of novel synergistic targets for rational drug combinations with PI3 kinase inhibitors using siRNA synthetic lethality screening against GBM. *Neuro Oncol* 13:367-375. 10.1093/neuonc/nor012

Knebel FH, Uno M, Galatro TF, Belle LP, Oba-Shinjo SM, Marie SKN, and Campa A. 2017. Serum amyloid A1 is upregulated in human glioblastoma. *J Neurooncol* 132:383-391. 10.1007/s11060-017-2386-z

Kosari F, Parker AS, Kube DM, Lohse CM, Leibovich BC, Blute ML, Cheville JC, and Vasmatzis G. 2005. Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness. *Clin Cancer Res* 11:5128-5139. 10.1158/1078-0432.ccr-05-0073

Kosmopoulos M, Christofides A, Drekolias D, Zavras PD, Gargalionis AN, and Piperi C. 2017. Critical Role of IL-8 Targeting in Gliomas. *Current Medicinal Chemistry* 25:1954-1967.

Kulasingam V, and Diamandis EP. 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* 5:588-599. 10.1038/ncponc1187

Liang JS, Sloane JA, Wells JM, Abraham CR, Fine RE, and Sipe JD. 1997. Evidence for local production of acute phase response apolipoprotein serum amyloid A in Alzheimer's disease brain. *Neurosci Lett* 225:73-76.

Liang ML, Hsieh TH, Ng KH, Tsai YN, Tsai CF, Chao ME, Liu DJ, Chu SS, Chen W, and Liu YR. 2016. Downregulation of miR-137 and miR-6500-3p promotes cell proliferation in pediatric high-grade gliomas. *Oncotarget* 7:19723-19737.

Lin CY, Yang ST, Shen SC, Hsieh YC, Hsu FT, Chen CY, Chiang YH, Chuang JY, Chen KY, Hsu TI, Leong WC, Su YK, Lo WL, Yeh YS, Patria YN, Shih HM, Chang CC, and Chou SY. 2018. Serum amyloid A1 in combination with integrin alphaVbeta3 increases glioblastoma cells mobility and progression. *Mol Oncol* 12:756-771. 10.1002/1878-0261.12196

Liu B, Dong H, Lin X, Yang X, Yue X, Yang J, Li Y, Wu L, Zhu X, Zhang S, Tian D, Wang J, Cai Q, Mao S, Chen Q, and Chang J. 2016. RND3 promotes Snail 1 protein degradation and inhibits glioblastoma cell migration and invasion. *Oncotarget* 7:82411-82423. 10.18632/oncotarget.12396

Louis DN, Perry A, Reifenberger G, Deimling AV, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, and Ellison DW. 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica* 131:803-820.

Lu J, Yu Y, Zhu I, Cheng Y, and Sun PD. 2014. Structural mechanism of serum amyloid A-mediated inflammatory amyloidosis. *Proc Natl Acad Sci U S A* 111:5189-5194. 10.1073/pnas.1322357111

Mandel ER, Uchida C, Nwadozi E, Makki A, and Haas TL. 2017. Tissue Inhibitor of Metalloproteinase 1 Influences Vascular Adaptations to Chronic Alterations in Blood Flow. *J Cell Physiol* 232:831-841. 10.1002/jcp.25491

Moshkovskii SA. 2012. Why do cancer cells produce serum amyloid A acute-phase protein? *Biochemistry (Mosc)* 77:339-341. 10.1134/s0006297912040037

406 Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, and Barnholtz-Sloan JS. 2017.
407 CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United
408 States in 2010-2014. *Neuro Oncol* 19:v1-v88. 10.1093/neuonc/nox158
409 Plate KH, and Dumont DJ. 2012. Tumor angiogenesis and anti-angiogenic therapy in malignant gliomas revisited.
410 *Keio J Med* 124:763-775.
411 Puchalski RB, Shah N, Miller J, Dalley R, Nomura SR, Yoon JG, Smith KA, Lankerovich M, Bertagnolli D, Bickley
412 K, Boe AF, Brouner K, Butler S, Caldejon S, Chapin M, Datta S, Dee N, Desta T, Dolbeare T, Dotson N,
413 Ebbert A, Feng D, Feng X, Fisher M, Gee G, Goldy J, Gourley L, Gregor BW, Gu G, Hejazinia N, Hohmann
414 J, Hothi P, Howard R, Joines K, Kriedberg A, Kuan L, Lau C, Lee F, Lee H, Lemon T, Long F, Mastan N,
415 Mott E, Murthy C, Ngo K, Olson E, Reding M, Riley Z, Rosen D, Sandman D, Shapovalova N, Slaughterbeck
416 CR, Sodt A, Stockdale G, Szafer A, Wakeman W, Wohnoutka PE, White SJ, Marsh D, Rostomily RC, Ng
417 L, Dang C, Jones A, Keogh B, Gittleman HR, Barnholtz-Sloan JS, Cimino PJ, Uppin MS, Keene CD,
418 Farrokhi FR, Lathia JD, Berens ME, Iavarone A, Bernard A, Lein E, Phillips JW, Rostad SW, Cobbs C,
419 Hawrylycz MJ, and Foltz GD. 2018. An anatomic transcriptional atlas of human glioblastoma. *Science*
420 360:660-663. 10.1126/science.aaf2666
421 Ramos AD, Magge RS, and Ramakrishna R. 2018. Molecular Pathogenesis and Emerging Treatment for
422 Glioblastoma. *World Neurosurgery* 116:495-504.
423 Ries C. 2014. Cytokine functions of TIMP-1. *Cell Mol Life Sci* 71:659-672. 10.1007/s00018-013-1457-3
424 Rodriguez FJ, Vizcaino MA, and Lin MT. 2016. Recent Advances on the Molecular Pathology of Glial Neoplasms in
425 Children and Adults. *J Mol Diagn* 18:620-634. 10.1016/j.jmoldx.2016.05.005
426 Roy R, Zurakowski D, Wischhusen J, Frauenhoffer C, Hooshmand S, Kulke M, and Moses MA. 2014. Urinary TIMP-
427 1 and MMP-2 levels detect the presence of pancreatic malignancies. *Br J Cancer* 111:1772-1779.
428 10.1038/bjc.2014.462
429 Sriram V, and Huse JT. 2015. The Evolving Molecular Genetics of Low-grade Glioma. *Advances in Anatomic*
430 *Pathology* 22:94-101.
431 Suzuki A, Badger BL, and Salmon ED. 2015. A quantitative description of Ndc80 complex linkage to human
432 kinetochores. *Nature Communications* 6:8161.
433 Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A,
434 Tsafou KP, Kuhn M, Bork P, Jensen LJ, and von Mering C. 2015. STRING v10: protein-protein interaction
435 networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447-452. 10.1093/nar/gku1003
436 Villapol S, Kryndushkin D, Balarezo MG, Campbell AM, Saavedra JM, Shewmaker FP, and Symes AJ. 2015. Hepatic
437 expression of serum amyloid A1 is induced by traumatic brain injury and modulated by telmisartan. *Am J*
438 *Pathol* 185:2641-2652. 10.1016/j.ajpath.2015.06.016
439 Yu Y, Liu J, Li SQ, Peng L, and Ye RD. 2014. Serum amyloid a differentially activates microglia and astrocytes via
440 the PI3K pathway. *J Alzheimers Dis* 38:133-144. 10.3233/jad-130818
441 Zhong S, Wu B, Dong X, Han Y, Jiang S, Zhang Y, Bai Y, Luo SX, Chen Y, and Zhang H. 2018. Identification of
442 Driver Genes and Key Pathways of Glioblastoma Demonstrates JNJ-7706621 as a novel Anti-Glioblastoma
443 Drug. *World Neurosurgery* 109:e329.
444

Figure 1

Differentially expressed gene expression heat map and volcano plot of glioma.

(A) Differentially expressed gene expression heat map of glioma (all upregulated and downregulated genes). (B) Differentially expressed genes were selected by volcano plot filtering (fold change ≥ 1 and p value ≤ 0.05). The blue point in the plot represents the differentially expressed genes with statistical significance.

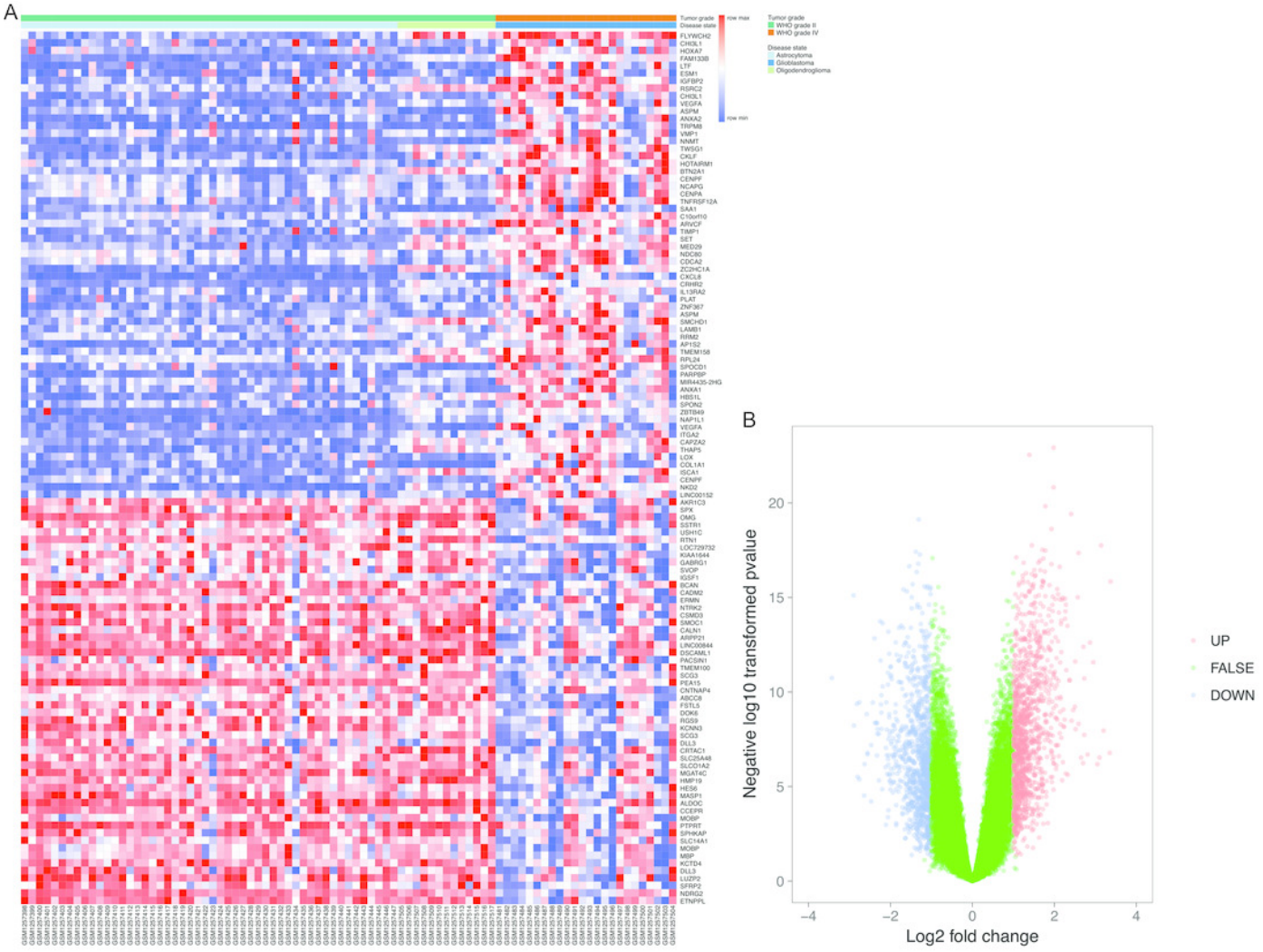
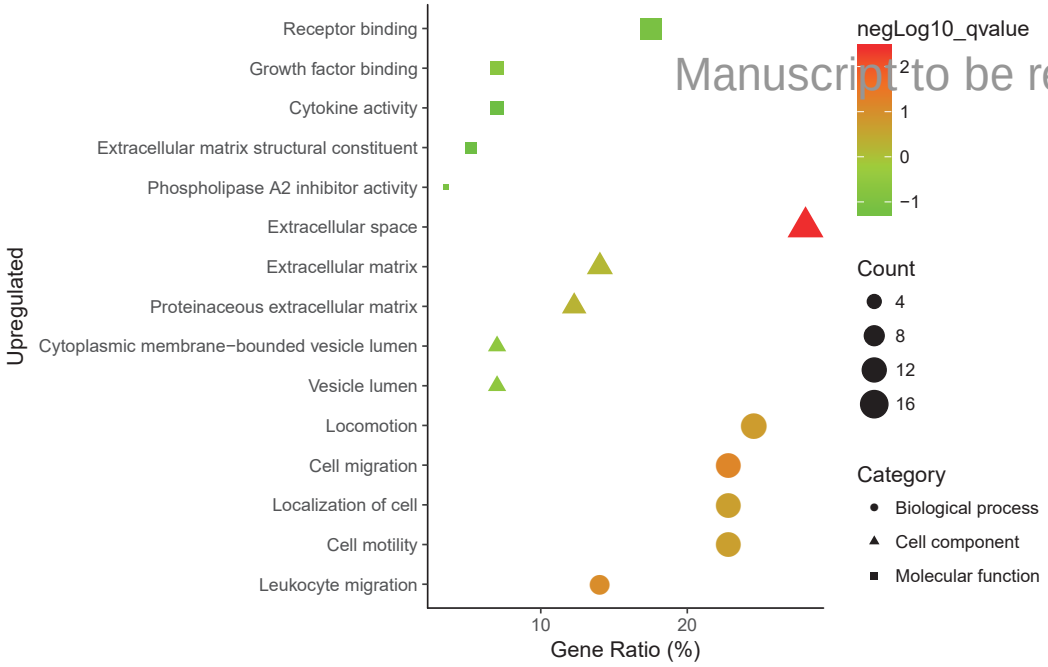


Figure 2(on next page)

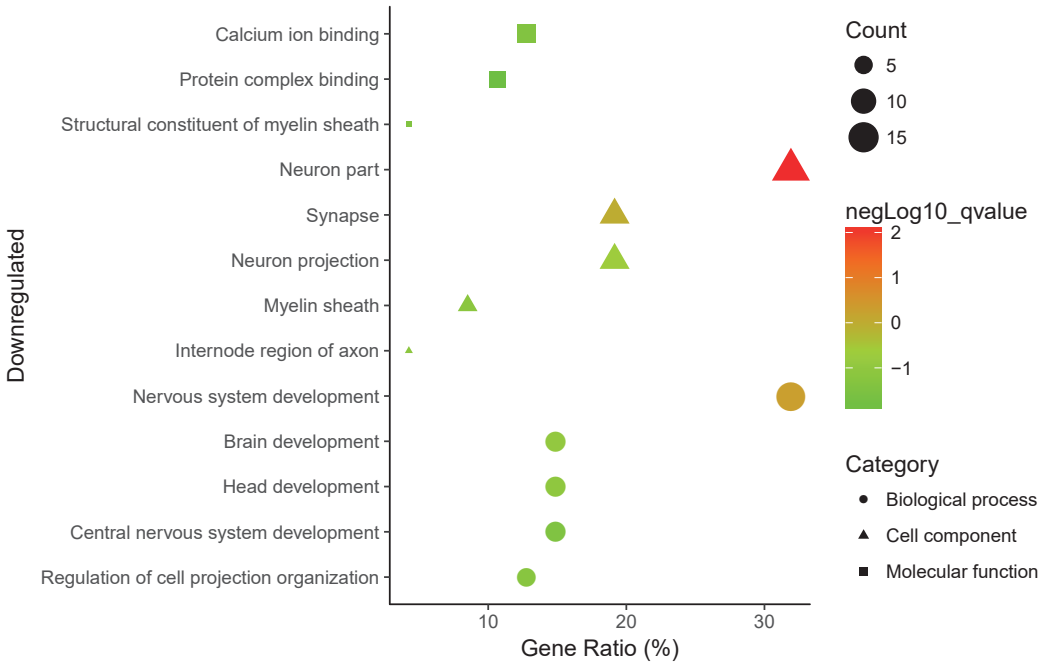
GOanalysis results showed that upregulated DEGs

(A) and downregulated DEGs (B) were particularly enriched in BP, MF, and CC. (C) The most significantly enriched KEGG pathway of the upregulated DEGs. GO: gene ontology; BP: biological process; MF: molecular function; CC: cell component; KEGG: Kyoto Encyclopedia of Genes and Genomes.

A



B



C

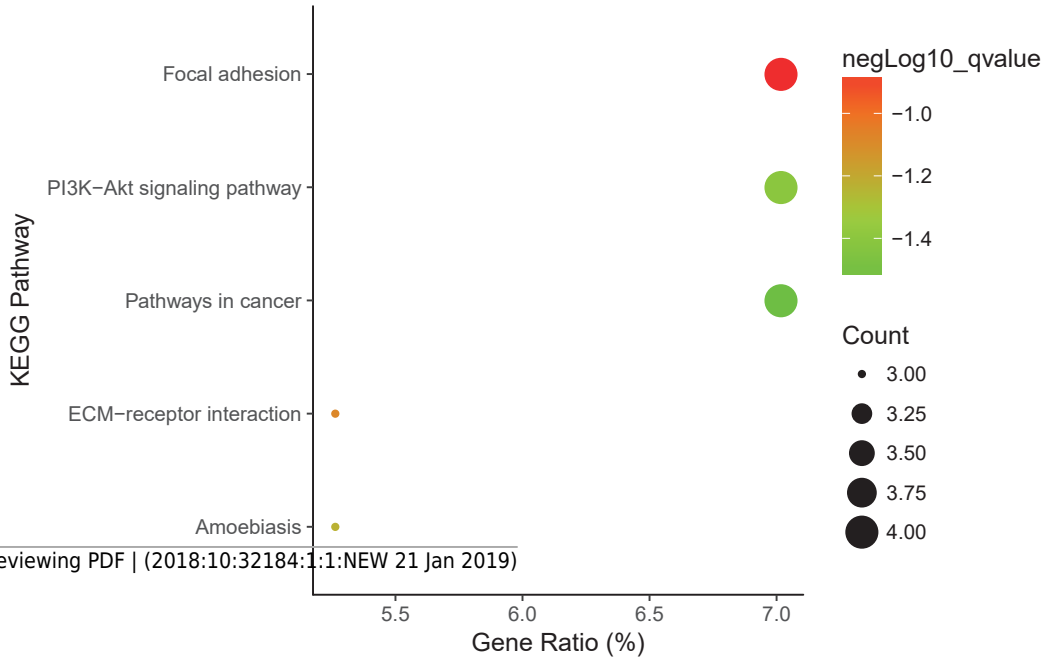


Table 1(on next page)

Gene ontology analysis ofdifferentially expressed genes associated with LGG and HGG

1 **Table 2: Gene ontology analysis of differentially expressed genes associated with LGG and HGG**

Expression	Category	Term	Count	%	PValue	FDR
Up-regulated	GOTERM_BP_FAT	GO:0016477~cell migration	13	22.80702	3.95E-05	0.067618
	GOTERM_BP_FAT	GO:0050900~leukocyte migration	8	14.03509	5.28E-05	0.090452
	GOTERM_BP_FAT	GO:0040011~locomotion	14	24.5614	1.09E-04	0.185762
	GOTERM_BP_FAT	GO:0051674~localization of cell	13	22.80702	1.22E-04	0.208673
	GOTERM_BP_FAT	GO:0048870~cell motility	13	22.80702	1.22E-04	0.208673
	GOTERM_MF_FAT	GO:0019838~growth factor binding	4	7.017544	0.004366	5.428929
	GOTERM_MF_FAT	GO:0019834~phospholipase A2 inhibitor activity	2	3.508772	0.010298	12.37092
	GOTERM_MF_FAT	GO:0005102~receptor binding	10	17.54386	0.010815	12.95228
	GOTERM_MF_FAT	GO:0005201~extracellular matrix structural constituent	3	5.263158	0.018986	21.69308
	GOTERM_MF_FAT	GO:0005125~cytokine activity	4	7.017544	0.019477	22.19193
	GOTERM_CC_FAT	GO:0005615~extracellular space	16	28.07018	3.09E-06	0.003877

	GOTERM_CC_FAT	GO:0005578~proteinaceous extracellular matrix	7	12.2807	4.24E-04	0.531176
	GOTERM_CC_FAT	GO:0031012~extracellular matrix	8	14.03509	5.41E-04	0.676623
	GOTERM_CC_FAT	GO:0060205~cytoplasmic membrane-bounded vesicle lumen	4	7.017544	0.002985	3.680925
	GOTERM_CC_FAT	GO:0031983~vesicle lumen	4	7.017544	0.003066	3.779192
	GOTERM_BP_FAT	GO:0007399~nervous system development	15	31.91489	3.29E-04	0.533895
	GOTERM_BP_FAT	GO:0007420~brain development	7	14.89362	0.00548	8.553316
	GOTERM_BP_FAT	GO:0060322~head development	7	14.89362	0.007003	10.80691
	GOTERM_BP_FAT	GO:0031344~regulation of cell projection organization	6	12.76596	0.011466	17.11031
	GOTERM_BP_FAT	GO:0007417~central nervous system development	7	14.89362	0.019918	27.91963
Down-regulated	GOTERM_MF_FAT	GO:0005509~calcium ion binding	6	12.76596	0.021579	24.70502
	GOTERM_MF_FAT	GO:0019911~structural constituent of myelin sheath	2	4.255319	0.02239	25.5132
	GOTERM_MF_FAT	GO:0032403~protein complex binding	5	10.6383	0.092825	71.83703
	GOTERM_CC_FA	GO:0097458~neuron part	15	31.91489	6.61E-06	0.007993

T					
GOTERM_CC_FAT	GO:0045202~synapse	9	19.14894	8.96E-04	1.078808
GOTERM_CC_FAT	GO:0043005~neuron projection	9	19.14894	0.00424	5.011629
GOTERM_CC_FAT	GO:0033269~internode region of axon	2	4.255319	0.01097	12.49482
GOTERM_CC_FAT	GO:0043209~myelin sheath	4	8.510638	0.01206	13.65505

2

Figure 3(on next page)

protein-proteininteraction network and top module of hub genes

(A) The protein-protein interaction network of top 15 hub genes. (B) Top module from the protein-protein interaction network.

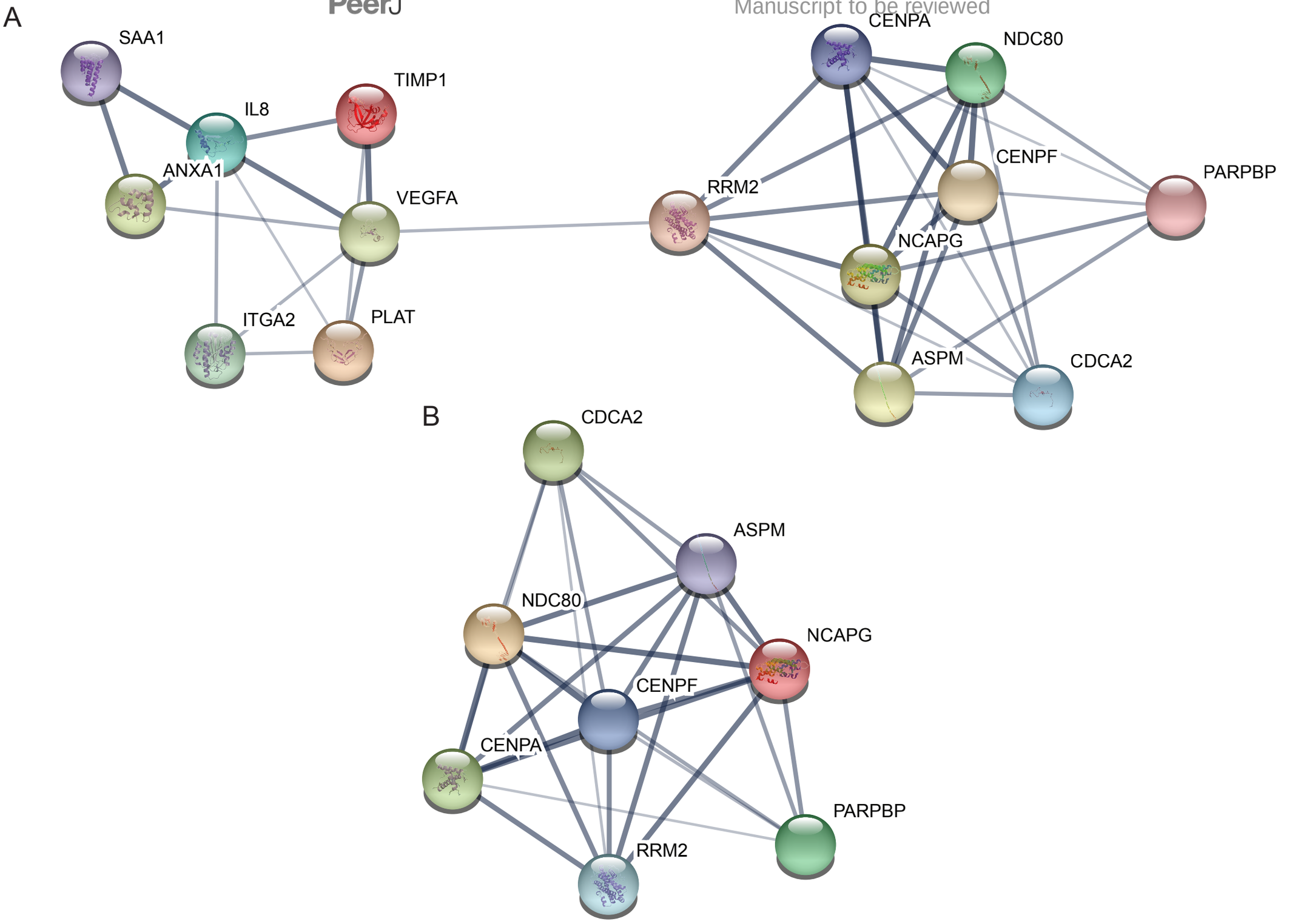

A
SAA1
IL8
TIMP1
ANXA1
VEGFA
ITGA2
PLAT
CENPA
NDC80
CENPF
PARPBP
RRM2
NCAPG
ASPM
CDCA2
B
CDCA2
ASPM
NDC80
NCAPG
CENPF
CENPA
PARPBP
RRM2

Table 2(on next page)

Top 15 hub genes with higher degree ofconnectivity

Table1. Top 15 hub genes with higher degree of connectivity

Gene	Degree	P-value
VEGFA	13	7.55E-07
NDC80	8	7.16E-09
IL8	8	5.15E-07
CENPA	7	3.31E-09
CENPF	7	6.06E-10
NCAPG	7	9.09E-10
ASPM	7	1.86E-08
RRM2	7	1.13E-09
ITGA2	6	2.45E-09
ANXA1	6	6.32E-08
CDCA2	6	6.42E-09
PLAT	5	2.64E-08
PARPBP	5	6.34E-15
TIMP1	4	2.06E-06
SAA1	4	2.40E-06

Figure 4(on next page)

Prognosticvalue of hub genes in glioma patients.

Prognostic value of hub genes (VEGFA, NDC80, CENPA, CENPF, NCAPG, ASPM, ITGA2, TIMP1, and SAA1) in glioma patients. HR: hazard ratio.

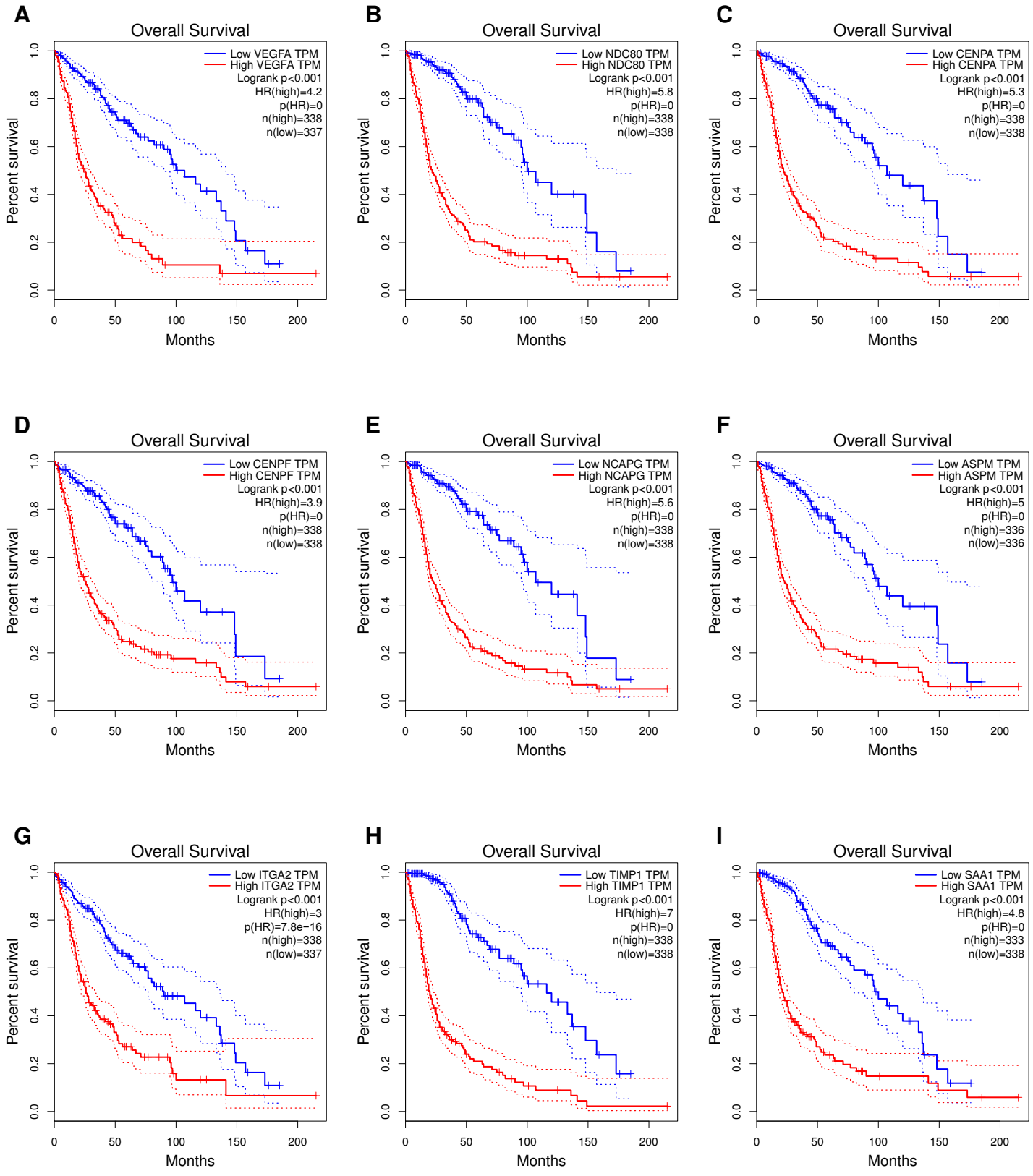

A
Overall Survival
Low VEGFA TPM
High VEGFA TPM
Logrank p<0.001
HR(high)=4.2
p(HR)=0
n(high)=338
n(low)=337
B
Overall Survival
Low NDC80 TPM
High NDC80 TPM
Logrank p<0.001
HR(high)=5.8
p(HR)=0
n(high)=338
n(low)=338
C
Overall Survival
Low CENPA TPM
High CENPA TPM
Logrank p<0.001
HR(high)=5.3
p(HR)=0
n(high)=338
n(low)=338
D
Overall Survival
Low CENPF TPM
High CENPF TPM
Logrank p<0.001
HR(high)=3.9
p(HR)=0
n(high)=338
n(low)=338
E
Overall Survival
Low NCAPG TPM
High NCAPG TPM
Logrank p<0.001
HR(high)=5.6
p(HR)=0
n(high)=338
n(low)=338
F
Overall Survival
Low ASPM TPM
High ASPM TPM
Logrank p<0.001
HR(high)=5
p(HR)=0
n(high)=336
n(low)=336
G
Overall Survival
Low ITGA2 TPM
High ITGA2 TPM
Logrank p<0.001
HR(high)=3
p(HR)=7.8e−16
n(high)=338
n(low)=337
H
Overall Survival
Low TIMP1 TPM
High TIMP1 TPM
Logrank p<0.001
HR(high)=7
p(HR)=0
n(high)=338
n(low)=338
I
Overall Survival
Low SAA1 TPM
High SAA1 TPM
Logrank p<0.001
HR(high)=4.8
p(HR)=0
n(high)=333
n(low)=338
Percent survival
Months
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200

Table 3(on next page)

KEGG pathway analysis ofdifferentially expressed genes associated with HGG

Table 3: KEGG pathway analysis of differentially expressed genes associated with HGG

Expression	Category	Term	Count	%	P Value	Genes	FDR
Up-regulated	KEGG_PATHWAY	hsa04510: Focal adhesion	4	7.017544	0.007454	LAMB1, VEGFA, ITGA2, COL1A1,	7.094772
	KEGG_PATHWAY	hsa04512:ECM-receptor interaction	3	5.263158	0.012926	LAMB1, ITGA2, COL1A1	12.01164
	KEGG_PATHWAY	hsa05146:Amoebiasis	3	5.263158	0.018811	LAMB1, CXCL8, COL1A1	17.03775
	KEGG_PATHWAY	hsa04151:PI3K-Akt signaling pathway	4	7.017544	0.02979	LAMB1, VEGFA, ITGA2, COL1A1	25.72978
	KEGG_PATHWAY	hsa05200:Pathways in cancer	4	7.017544	0.041599	LAMB1, VEGFA, ITGA2, CXCL8	34.15806

Figure 5(on next page)

The expression level and potential function of SAA1 and TIMP1

(A) SAA1 significantly increased in glioblastomas; (B) The expression level of SAA1 have no significant difference in LGG; (C) TIMP1 significantly increased in glioblastomas; (D) TIMP1 is significantly higher in astrocytoma than oligodendroglioma and oligoastrocytoma; (E, F) SAA1 and TIMP1 increase in both IDH mutant IDH wild type. The same results were found in MGMT promoter and non-deletion of chromosome 1p.19q; G, H. SAA1 and TIMP1 played important roles in MES-like in the IDH wild type; (I, J) SAA1 regulates biology process associated with inflammatory response processes and cytokine mediated signaling pathway; (K, L). TIMP1 negatively regulates adaptive immune response based on somatic recombination of immune receptors built from leucine-rich and response to interferons.