Identification of potential crucial genes and molecular mechanisms in glioblastoma multiforme by bioinformatics analysis

XIAOJIE CHEN¹, YUANBO PAN², MENGXIA YAN³, GUANSHUI BAO¹ and XUHONG SUN¹

¹Department of Neurology, Shanghai Ninth People’s Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 201999; ²Department of Neurosurgery, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310000, P.R. China

Received August 22, 2019; Accepted April 4, 2020

DOI: 10.3892/mmr.2020.11160

Abstract. Glioblastoma multiforme (GBM) is the most common and malignant brain tumor of the adult central nervous system and is associated with poor prognosis. The present study aimed to identify the hub genes in GBM in order to improve the current understanding of the underlying mechanism of GBM. The RNA-seq data were downloaded from The Cancer Genome Atlas database. The edgeR package in R software was used to identify differentially expressed genes (DEGs) between two groups: Glioblastoma samples and normal brain samples. Gene Ontology (GO) functional enrichment analysis and the Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis were performed using Database for Annotation, Visualization and Integrated Discovery software. Additionally, Cytoscape and Search Tool for the Retrieval of Interacting Genes/Proteins tools were used for the protein-protein interaction network, while the highly connected modules were extracted from this network using the Minimal Common Oncology Data Elements plugin. Next, the prognostic significance of the candidate hub genes was analyzed using UALCAN. In addition, the identified hub genes were verified by reverse transcription-quantitative (RT-q) PCR. In total, 1,483 DEGs were identified between GBM and control samples, including 954 upregulated genes and 529 downregulated genes (P<0.01; fold-change >16) and these genes were involved in different GO terms and signaling pathways. Furthermore, CDK1, BUB1, BUBIB, CENPA and GNG3 were identified as key genes in the GBM samples. The UALCAN tool verified that higher expression level of CENPA was relevant to poorer overall survival rates. In conclusion, CDK1, BUB1, BUBIB, CENPA and GNG3 were found to be potential biomarkers for GBM. Additionally, ‘cell cycle’ and ‘γ-aminobutyric acid signaling’ pathways may serve a significant role in the pathogenesis of GBM.

Introduction

According to the guidelines from the World Health Organization, gliomas are categorized into grades I to IV. Glioblastoma multiforme (GBM) is classified as grade IV and it is a highly malignant form (1,2). GBM has been recognized as the most aggressive type of brain tumor with highly infiltrative ability (3), affecting about ~20,000 people every year in the United States (4). Currently, the treatments for GBM include surgical resection, chemotherapy and radiotherapy (5). However, due to the aggressiveness of gliomas and their resistance to chemotherapy and radiation therapy (6,7), the prognosis of GBM is still poor with a median survival time of 12-15 months for patients with GBM (8,9). In recent years, targeted therapies, such as combinatorial chemotherapy targeting molecular subgroups and gene and immune therapies, have made important progress in preclinical models (10-12). However, the potential molecular mechanism in the pathogenesis of GBM needs to be further investigated. In order to improve the prognosis of GBM patients, it is important to identify molecular mechanism underlying GBM pathogenesis.

In recent years, a number of studies have investigated the potential molecular mechanisms of GBM. It has been reported that the upregulation of EGFR may be responsible for the resistance of glioma cells to the treatments with chemotherapy and radiation (13-15). Additionally, Annexin-A5 over-activation increases Snail expression level via the PI3K/Akt/NF-κB signaling pathway, which is associated with glioblastoma cells migration and invasion (16). In addition, Siebzehnrubl et al (17) indicated that ZEB1 is linked to tumor chemoresistance and invasion of glioblastoma cells by modulating its downstream effectors, including c-MYB, MGMT and ROBO1. Nevertheless, the exact molecular mechanism and gene networks underlying GBM progression remain to be fully elucidated.

The present study aimed to identify the potential key nodes and molecular mechanisms associated with the progression of GBM. Co-expression interactions between the identified...
DEGs were conducted using a protein-protein interaction (PPI) network and several critical genes were identified. In addition, a network module analysis was performed according to the PPI network. In conclusion, the present study may improve the current understanding of the pathogenesis underlying GBM and identified key genes that may represent novel targets for the development of novel treatments of GBM.

Materials and methods

Data download. RNA-seq data were downloaded from The Cancer Genome Atlas (TCGA) database (portal.gdc.cancer.gov) (18). The data of 169 GBM samples and 5 normal brain tissue samples were used and analyzed.

Data preprocessing and DEG analysis. Based on the annotation information contained in the Ensembl database (asia.ensembl.org/html) (19), 23,269 protein coding genes were identified. Then, the DEGs between GBM samples and normal samples were screened using edgeR package in R (20,21). A P<0.01 and log2|fc|>4 were selected as the cutoff criteria for the identification of DEGs. In addition, a volcano plot was drawn using the gplots package in R software (22).

Functional and pathway enrichment analysis of DEGs. To reveal the main functional pathways of GBM, the Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) was used to analyze Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in the identified DEGs (23). DEG count ≥2 and P<0.05 were used as cutoff to identify the significant biological functions and signaling pathways.

PPI network and module analysis of DEGs. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) was used to assess the interacting partners of the DEGs (24), and the PPI network was constructed with a combined score >0.7. Visualization of the network was performed with Cytoscape (version 3.6.1) software (25). After constructing the PPI network, topology characteristics of the network were analyzed using the Network Analyzer plug-in of Cytoscape software, and degree distribution, distribution of the shortest path, average clustering coefficient and closeness centrality were examined (26). Subsequently, the top 10 significant hub genes or key nodes were identified according to their connectivity degrees by using the CytoHubba tool, a plug-in for Cytoscape (27). The MCODE plug-in was used to analyze the highly interconnected clusters of this network with default parameters (28).

UALCAN survival analysis. UALCAN (ualcan.path.uab.edu/) is an interactive web-software that can be used to perform analyses of tumor subgroup gene expression and survival (29). In UALCAN, samples were categorized into two groups: High expression and Low/Medium expression patients. High expression patients refer to the highest quarter of all patients. Low and medium expression patients refer to the remaining three quarters of patients. The P-values from the Kaplan-Meier analysis were based on log-rank. The effects of candidate key gene expression levels on overall survival (OS) were analyzed using UALCAN.

Sample collection and ethics statement. To determine the expression of hub genes in GBM tissues, samples from 20 patients (age, 26-77 years), including nine males and eleven females, who underwent tumor resection surgery were collected at the Shanghai Ninth People’s Hospital between September 2010 and March 2018. All patients provided written informed consent, and the study conducted in accordance with the Declaration of Helsinki. In addition, 20 normal brain tissues (used as the control group) were collected from patients who suffered from traumatic brain injury and required internal decompression surgery. After collection, all tissues were frozen in liquid nitrogen and stored at -80°C. All participants in the present study signed informed consents and the study was approved by the Ethics Review Committee of Shanghai Ninth People’s Hospital.

RNA extraction and reverse transcription-quantitative (RT-q) PCR. TRIzol (cat. no. 9109; Takara Bio, Inc.) was used to isolate total RNA from each tissue sample according to the manufacturer’s instructions. Briefly, 2 ml TRIzol was added to brain tissue (100 mg) to prepare homogenate. After obtaining RNA, TE buffer was used to dilute the RNA. Nanodrop software (ND2000C; Gene Company, Ltd.) was used to determine the OD value. A value of OD_{260}/OD_{280} within 1.7-2.1 indicated that the extracted RNA was relatively pure. Then, total RNA was reverse transcribed into cDNAs using the PrimeScript RT Master Mix (cat. no. RR036A; Takara Bio, Inc.). RT-qPCR was performed to measure the levels of cDNAs using a SYBR GREEN kit (cat. no. RR420 A; Takara Bio, Inc.). qPCR amplification procedure was performed as follows: Pre-denaturation at 95°C for 30 sec for 1 cycle followed by 40 cycles of 95°C for 5 sec and 60°C for 20 sec. The relative expression level of the five key genes was calculated following comparative CT method, as previously described (30). GAPDH was used to normalize the mRNA expression level. The primer sequences are presented in Table I.

Statistical analysis. Experimental data are presented as the mean ± standard error of the mean. All statistical analyses were performed using SPSS 20.0 (IBM, Corp.) and visualized using GraphPad Prism 6.0 (GraphPad Software Inc.). Differences in expression levels between groups were analyzed by unpaired Student’s t-test. P<0.05 was considered to indicate a statistically significant difference.

Results

Identification of DEGs. Using P<0.01 as a cutoff, a total of 1,483 DEGs (954 upregulated genes and 529 downregulated) were identified between GBM and control samples (Table S1). The volcano plot of DEGs is presented in Fig. 1. Each individual dot represents a DEG.

Functional and pathway enrichment analysis of the DEGs. The GO enrichment analysis results revealed that 954 upregulated DEGs were primarily involved in biological process that included ‘anterior/posterior pattern specification’, ‘embryonic skeletal system morphogenesis’, ‘sister chromatid cohesion’ and ‘cell division’ (Fig. 2). According to the ‘cellular components’ analysis,


DEGs were found to be primarily localized in the nucleus. In addition, the most significantly enriched molecular functions were related to transcriptional activator activity. In addition, as presented in Fig. 3, 529 downregulated DEGs were significantly associated with 'chemical synaptic transmission', 'regulation of ion transmembrane transport' and 'γ-aminobutyric acid signaling pathway'. The top five main biological processes enriched in the upregulated and downregulated deGs are presented in Tables II and III, respectively. The associated signaling pathways significantly enriched in the identified DEGs are presented in Table IV. The KEGG pathway analysis indicated that upregulated DEGs were mostly involved in pathways such as 'cell cycle' and 'p53 signaling' pathways, whereas downregulated genes were associated with 'retrograde endocannabinoid signaling', 'GABAergic synapse' and 'glutamatergic synapse'.

**PPI network construction of DEGs.** Based on the STRING database and Cytoscape software, PPI networks were constructed with combined scores >0.7, including 645 nodes and 1,779 edges (Fig. 4). Subsequently, the node degree distribution of this network was analyzed using a pattern of power-law according to the topology property (Fig. 5). In this network, a degree value >40 was used as the cutoff criterion for the CytoHubba tool. Finally, 10 DEGs were identified as the hub genes with the highest connectivity degree: cyclin dependent kinase 1 (CDK1), centromere protein A (CENPA), G protein subunit γ 3 (GNG3), BUB1 mitotic checkpoint serine/threonine kinase (BUB1), cyclin B2 (CCNB2), kinesin family member 2C (KIF2C), aurora kinase B (AURKB), baculoviral IAP repeat containing 5 (BIRC5), cell division cycle associated 8 (CDCA8) and BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B).

**Module analysis of the PPI network.** Generally, biological networks contain several functional modules and they may interact in various biological process (31). In the present study,

<table>
<thead>
<tr>
<th>cDNA</th>
<th>Forward primer (5’-3’)</th>
<th>Reverse primer (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDK1</td>
<td>CACAAAAACTACAGGTCAAGTGG</td>
<td>GAGAAATTTCCCGAATTCGAGT</td>
</tr>
<tr>
<td>BUB1</td>
<td>GAAAGCATGAGCAATGGGAAA</td>
<td>CCACCTGATGCAACTTCTTATG</td>
</tr>
<tr>
<td>BUB1B</td>
<td>ATGGGTCCCTCTGGAAACTTAG</td>
<td>GGAATGTAGTGTCAAAAACCC</td>
</tr>
<tr>
<td>CENPA</td>
<td>AAGAGCACAACCTCTTGATAA</td>
<td>CATGTAAGGGAGGAGATAAGGC</td>
</tr>
<tr>
<td>GNG3</td>
<td>CGGTGAACAGCACTAGATGTAT</td>
<td>TCACAGTAAGCTACAGGTCTG</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GTGGACCCTGACCTGGGTCTAG</td>
<td>GAGTGGGTGCTGCTGGAGAAGTC</td>
</tr>
</tbody>
</table>

**Table I. Primer sequences for PCR.**

*Figure 1. Volcano plot of differentially expressed genes in glioblastoma multiform. Black, non-differentially expressed genes; red, significantly upregulated genes; green, significantly upregulated genes (based on \(|\log_{2} FC|>4\) and \(P<0.01\). FC, fold-change. FDR, false discovery rate.*
the top two clusters were identified from the PPI network using MCODE analysis. The most significant cluster involved 35 nodes and 572 edges, and CDK1, BUB1, BUB1B, CENPA were highly enriched (Fig. 6A). Cluster 2 is presented in Fig. 6B, consisting of 24 nodes and 276 edges. Specifically, the hub gene GNG3 was enriched in cluster 2, which indicated that it may be involved in the pathogenesis of GBM by cooperating with other genes.

UALCAN survival analysis. UALCAN was used to analyze the OS of 169 patients with GBM based on TCGA data. Briefly, survival analysis indicated that the higher expression of CENPA gene was significantly correlated with shorter OS time in GBM patients, whereas the expression of CDK1, BUB1, BUB1B and GNG3 was not correlated with OS (Fig. 7).

Expression of hub genes in GBM. To further verify the expression level of hub genes in GBM samples, RT-qPCR was performed to calculate the mRNA levels of the five hub genes identified in the present study (CDK1, BUB1, BUB1B, CENPA and GNG3) in GBM samples. As illustrated in Fig. 8, the expression of CDK1, BUB1, BUB1B and CENPA were
significantly upregulated in GBM samples compared with normal control tissues. By contrast, the expression level of GNG3 was higher in the control tissues compared with tumor tissues. The present RT-qPCR results were in line with the aforementioned bioinformatics analysis, suggesting that these key genes may be linked to the molecular mechanism underlying GBM.

Discussion

Glioblastoma multiforme (GBM) is the most common aggressive brain tumor, and is associated with a poor patient survival rate (32). However, the most important challenge in treating GBM is the presence of significant intra-tumor heterogeneity (13,33,34). Although previous studies have reported numerous potential biomarkers associated with the progression of GBM, the potential molecular mechanism underlying its pathogenesis has not been comprehensively investigated (35-37). In the present study, a total of 1,483 DEGs were identified, containing 954 upregulated genes and 529 downregulated genes. The present results suggested that these DEGs were primarily enriched in cell division, mitotic nuclear division and chemical synaptic transmission. In addition, the upregulated and downregulated genes were related to cell cycle, p53 signaling pathway and synapse. A PPI network was constructed and the highly connected module were identified. After module analysis, several hub genes with higher degree of connectivity were identified, including CDK1, BUB1, BUB1B, CENPA and GNG3.

In the TCGA database, there were only 5 control samples. Although it is of a smaller size when compared with the 169 GBM samples, it should not have great influence on the results of the present study, as there was relatively little individual difference in the control simples from different patients, whereas the difference was much more evident in tumor samples due to intratumoral heterogeneity between different patients.
Table IV. Results of KEGG pathway analysis.

A, Upregulated genes

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04110: Cell cycle</td>
<td>12</td>
<td>3.26E-07</td>
</tr>
<tr>
<td>hsa05322: Systemic lupus erythematosus</td>
<td>9</td>
<td>2.49E-04</td>
</tr>
<tr>
<td>hsa04115: p53 signaling pathway</td>
<td>6</td>
<td>0.001428</td>
</tr>
<tr>
<td>hsa04512: ECM-receptor interaction</td>
<td>6</td>
<td>0.004496</td>
</tr>
<tr>
<td>hsa05202: Transcriptional misregulation in cancer</td>
<td>8</td>
<td>0.004710</td>
</tr>
</tbody>
</table>

B, Downregulated genes

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04080: Neuroactive ligand-receptor interaction</td>
<td>44</td>
<td>5.35E-28</td>
</tr>
<tr>
<td>hsa04723: Retrograde endocannabinoid signaling</td>
<td>20</td>
<td>4.27E-14</td>
</tr>
<tr>
<td>hsa04020: Calcium signaling pathway</td>
<td>24</td>
<td>3.72E-13</td>
</tr>
<tr>
<td>hsa04727: GABAergic synapse</td>
<td>15</td>
<td>7.40E-10</td>
</tr>
<tr>
<td>hsa04724: Glutamatergic synapse</td>
<td>15</td>
<td>3.76E-08</td>
</tr>
</tbody>
</table>

KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 4. The giant protein-protein interaction network. Key nodes in the network are presented in different colors: Blue stands for the hub genes, red corresponds to the upregulated genes and green corresponds to the downregulated genes in glioblastoma multiforme.
CDK1, also named cell division control protein 2, is an important cell cycle regulator functioning as a serine/threonine kinase (38,39). During G2 and early mitosis, active CDK1-cyclin complexes can phosphorylate various downstream proteins, leading to re-organization of the cytoskeleton, nuclear envelope breakdown and chromosome condensation (40,41). Previous studies have reported that downregulation of CDK1 could inhibit proliferative ability of human glioma cells, while overexpression of CDK1 contributed to senescence escape of the cells and promoted oncogenesis of human gliomas (42-44). In the present study, CDK1 was also observed to be significantly upregulated in the module analysis of the PPI network constructed. The present results suggested that CDK1, interacting with other genes identified in the present module analysis, may be associated with the progression of GBM by modulating the cell cycle.

In addition, CDK1, BUB1, BUB1B and CENPA were also identified as key nodes in the current study. BUB1 and BUB1B are spindle assembly checkpoint (SAC) genes that serve as a controller of mitotic checkpoints and chromosome segregation (32,45). BUB1 mRNA was reported to be upregulated
in glioma samples and the expression level was positively correlated with glioma grade (46). In addition, BUB1B was identified to be enriched in glioblastoma cells and associated with radio-resistance and recurrence of glioblastoma (47,48). In accordance with these previous studies, the present study identified that the expression levels of both BUB1 and BUB1B in GBM samples were higher compared with the negative controls. Collectively, accumulating evidence suggested that BUB1 and BUB1B may exert a significant influence on the progression of GBM by regulating mitotic spindle assembly checkpoint and sister chromosome segregation during mitosis. It has been previously reported that SAC inhibition promoted the response of glioblastoma cells to antimitotic drugs and enhanced the efficacy of tumor treating fields, which impair mitosis by disturbing the spindle formation (49,50). Therefore, inhibition of BUB1 or BUB1B may be a therapeutic strategy to treat GBM by...
influencing SAC during the process of cell division. In addition, *CENPA*, which encodes a centromere-associated protein, plays an important role in cell division by directing the assembly of active kinetochores (51). Failure in this process could lead to dysfunction of chromosome segregation, which has been found to be associated with initiation and progression of cancer (52-54). Notably, the present study identified a higher expression level of *CENPA* in brain tissue samples from GBM patients and high levels of *CENPA* were significantly associated with shorter OS. In addition, GO enrichment analyses showed that *CENPA* was involved in the biological function of sister chromatid cohesion, nucleosome assembly and mitotic cytokinesis. The present results suggested that *CENPA* gene may regulate the mitosis of glioma cells by interacting with other genes in the cluster 1, thus suggesting an association with pathogenesis and prognosis in GBM. The present study identified a novel potential target molecule in GBM, and inhibition of *CENPA* may be a novel potential therapeutic strategy for GBM.

The present results showed that *GNG3* may be associated with the GABAergic synapse pathway. *GNG3* is a gene that encodes the γ subunits of G proteins (55), and the disruption of *GNG3* may induce the dysfunction of the GABAB1 receptor signaling pathway (56). Notably, a previous study reported that switching GABA catabolism toward γ-hydroxybutyric acid production could suppress glioblastoma cell tumorigenic properties (57). Therefore, it was hypothesized that low expression of *GNG3* may be associated with the pathogenesis of GBM by regulating related signaling pathways (Table SII). However, further studies are required to confirm the role of *GNG3* in GBM.

Collectively, the present study has identified that several hub genes (*CDK1, BUB1, BUB1B, CENPA* and *GNG3*) were involved in the mechanism of GBM. The present results suggested that *CDK1, BUB1, BUB1B* and *CENPA* may play an essential role in the cell cycle pathway, leading to GBM progression. In addition, *GNG3* was found to be a potential therapeutic target to treat GBM. The present results may provide novel insights into the molecular mechanisms of GBM. Additional studies investigating the hub genes identified in the present study are required to examine their detailed function and interaction in GBM. However, several classical molecular traits, such as IDH1 mutation status, were not analyzed with these hub genes in our study. The authors believe that the IDH mutation analysis for these hub genes, such as *CENPA*, need to be investigated in the future. Furthermore, it is also appealing to investigate whether these hub genes are regulated by oncogenic mutations of themselves or other genes in the future.

**Acknowledgements**

Not applicable.

**Funding**

This research was financially supported by The Shanghai Science & Technology Research Program (grant no. 14411972200) and Shanghai Wu Mengchao Medical Science Foundation (grant no. JJHM-2019009).
Availability of data and materials
The datasets analyzed in the present study are available from The Cancer Genome Atlas database (https://portal.gdc.cancer.gov/).

Authors' contributions
GB and XS conceived and designed the study. YP downloaded and analyzed the RNA-seq data. XC and MY collected the samples. XC and MY analyzed the data. XC wrote the manuscript. All authors read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate
Informed consent was obtained from all patients and the present study was approved by the Ethics Review Committee of Shanghai Ninth People's Hospital.

Patient consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interest.

References


