Cell cycle-related genes expression in hga

Cell cycle-related genes areprognosticmarkersofsurvival in high grade astrocytomas

Highlights:

1. A total of 598 genes were identified as significant risk genes for prognosis.
2. Prognosis genes including *CDC6* and *AURKA* had higher degrees in the PPI network.
3. *CDC6* , *AURKA* and *CHEK1* were mainly enriched in cell cycle and mitotic cell cycle.

Abstract

Objective: High grade astrocytoma (HGA) is a kind of aggressive brain tumor. Moreover, the prognosis of HGA is poor. We aimed to explore the underlying mechanism and screen the risk genes for prognosis of HGA. Method: GSE33331 including 26 brain tumor samples obtained from 26 patients with HGA was downloaded from Gene Expression Omnibus. Risk genes for prognosis of HGA were identified. Protein-protein interaction (PPI) network was constructed using STRING and the network module was identified by MCODE plugin of Cytoscape. Then, function annotation for risk genes for prognosis of HGA and the genes in the network module were performed. Subsequently, drug target was analyzed. Results: A total of 598 genes were identified as significant risk genes for prognosis. All risk genes for prognosis of HGA including *CDC6* (cell division cycle 6), *AURKA* (aurora kinase A) and *CHEK1* (checkpoint kinase 1) were significantly enriched in cell cycle, mitotic as well as mitotic anaphase. While the genes in the network module such as

CDC6, AURKA and CHEK1 mainly participated in functions such as cell cycle, mitotic cell cycle and cell cycle process. Moreover, the genes in the network module mainly participated in the pathways such as cell cycle and cell cycle, mitotic. Drug target analysis showed that 7 genes were recorded in Drugbank database, and there were as many as 32 drug records of *CHEK1* . Conclusion: *CDC6* , *AURKA* and *CHEK1* might be used as prognostic biomarker for HGA. Moreover, our results might provide underlying molecular target for HCC therapy.

Keywords: high grade astrocytoma; prognosis; protein-protein interaction network; function annotation

Introduction

Astrocytoma is the most common primary glial tumors in the central nervous system (CNS) and is graded as low and high grade astrocytomas (HGA) by the World Health Organization (WHO) [1]. The median survival ofHGA, consisting predominantly of glioblastoma (GBM) and anaplastic astrocytoma (AA), is 15 months and 3 years, respectively [2, 3]. Thus, the features of molecular alterations underlying the aggressive behavior of HGA and the identification of new markers are an important step towards a better management for HGA patients.

Initiation and progression of astrocytomas are associated with genetic and chromosomal alterations. He *et. al* have demonstrated that astrocyte elevated gene-1 (AEG-1) is a novel and poor prognostic marker for astrocytomas patients [4]. Kirla *et. al* have reported that low expression of p27 acts as a poor prognostic indicator for survival in patients with HGA via

regulation of cell cycle [5]. Conversely, VRK2 (Vaccinia-related kinase-2) was exhibited to be as a good prognostic marker for patient survival which was associated with longer survival in HGA [6]. Moreover, ADAR2 was regarded as a possible marker for long-term survival of patients with recurrent HGA via mediated RNA editing [7]. Additionally, isocitrate dehydrogenase 1 mutation has been established as a molecular prognostic factor in adult HGA [8]. Despite many studies have identified prognostic genes in HGA, there still remains no predictor of survival which has proved robustly reproducible from study to study [9].

Gene expression microarray analysis has been applied to identify prognostic biomarkers in HGA. This genome-wide method may provide the insight into the biological mechanisms of tumorigenesis which can be utilized for the development of more effective therapies. Thus, more researchers took advantage of bioinformatics methods to analyze the information of GSE33331 [10] offered Donson *et al.* who analyzed the association between immune function-related genes expression as well as immune cell infiltration and survival. However, drug target analysis was not performed.

Hence, we exploited the bioinformatics methods to screen the risk genes for prognosis of HGA, and constructed a PPI network and identified the network module. Moreover, functional enrichment analysis of risk genes for prognosis of HGA and the genes in the network module were performed, followed by drug target analysis.

2 Materials and Methods

2. 1 *Microarray data*

Gene expression microarray data GSE33331 [10] including 26 brain tumor samples obtained from 26 patients with HGA, was downloaded from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (http://www. ncbi. nlm. nih. gov/geo/), which was based on GPL570 platform of [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2. 0 Array (Affymetrix, Santa Clara, CA).

2. 2 Data preprocessing and identification of risk genes for prognosis of HGA

The probe data in CEL document were processedto corresponding gene symbols based on the annotation of GPL570 platform. The expression values ofmulti-probes for a given gene symbol were reduced to a single value by computing the average expression value using Aggregate function in the R language (http://www. bioconductor. org). While for probes with missing value, the KNN method [11] of the impute package (http://bioconductor. org/packages/release/bioc/html/impute. html) [12] in the R language (http://www. bioconductor. org) was used to fill the missing value (k value with default value is 10). Quantile normalization was performed using preprocessCore package (http://bioconductor. org/packages/release/bioc/html/preprocessCore. html) [13]. Subsequently, the expression matrix was obtained.

To analyze the relation between genes and survival, the patients were grouped based on the expression values of genes (upper or lower) which were evaluated by AUC of ROC curve. Moreover, the ROC curve was performed by the pROC package (http://cran. r-project. org/web/packages/pROC/index. html) [14]. The patient survivalKaplan–Meier

(KM) curves were plotted by the packages KMsurv (http://CRAN. R-project. org/package= KMsurv) [15] and survival [16] of R language to find genes related to survival, and differences were evaluated by Log Rank test. A p-value < 0. 05 and the AUC values of ROC ≥ 0. 8 were considered as the criteria of risk genes for prognosis of HGA.

2. 3 PPI network construction and identification of modules

To study interaction of genes, PPI network was constucted with STRING software (http://string-db. org/) [17] and displayed by Cytoscape (http://cytoscapeweb. cytoscape. org/) [18] for protein-protein pairs in which combined score was more than 0. 4.

The MCODE plugin (http://baderlab. org/Software/MCODE) of Cytoscape was applied to identify the modules of PPI network.

2. 4 *Function annotation for risk genes for prognosis of HGA and the genes in the network module*

To annotate functions of risk genes for prognosis of HGA and the genes in the network module, TargetMine (http://targetmine. nibio. go. jp) [19] tool was used to perform the GO function and pathway enrichment analysis. The adjusted p value < 0. 05 was considered to be significant which was calculated by the Holm-Bonferroni[20] method.

2. 5 Drug target analysis

Drugbank database (http://www. drugbank. ca/) [21] was applied to screen drug information of key target genes, which was a dual purpose

bioinformatics-cheminformatics database with a strong focus on quantitative, analytic or molecular-scale information about drugs and drug targets. Moreover, the networks of drugs and their corresponding genes were constructed.

3 Results

3. 1 Data preprocessing and i dentification of risk genes for prognosis of HGA

A total of 8134 genes were identified. The data before and after normalization were exhibited in Figure 1A and 1B.

Based on the criteria, 598 genes were identified as significant risk genes for prognosis. Table 1 showed the grouping critical value of genes (*KCNJ6* (potassium inwardly-rectifying channel, subfamily J, member 6), *CHEK1* (checkpoint kinase 1), *LOC283887* (uncharacterized LOC283887) and *LTK* (leukocyte receptor tyrosine kinase)) of AUC greater than 0. 95. The AUC value is larger, the critical value of the gene as a grouping criteria is more reliable, and the accuracy of AUC more than 0. 9 is higher. The ROC and KM curves of genes (*KCNJ6* , *CHEK1* , *LOC283887* and *LTK*) were shown in Figure 2.

3. 3 PPI network of risk genes for prognosis of HGA

The STRING tool was used to obtain the PPI relationships of 598 risk genes for prognosis of HGA. As shown in Figure 3 A, 381 nodes and 1253 edges were involved in the PPI network. Furthermore, 8 hub genes with the higher degrees including *CDC6* (cell division cycle 6) (degree = 48), *POLD1* (polymerase (DNA directed), delta 1, catalytic subunit) (degree = 46), *KNTC1*

(kinetochore associated 1) (degree = 42), *AURKB* (aurora kinase B) (degree = 41), *AURKA* (aurora kinase A) (degree = 41), *RRM2* (ribonucleotide reductase M2) (degree = 40), *MAD2L1* (MAD2 mitotic arrest deficient-like 1 (yeast)) (degree = 40), and *CDC20* (cell division cycle 20) (degree = 40) might play important roles in the development of HGA.

Subsequently, the network module of the PPI network was obtained using MCODE plugin of Cytoscape (Figure 3 B). In the present study, 31 genes and 435 edges were involved in this module. Moreover, there was an interaction among two genes.

3. 4 GO and pathway enrichment analysis of risk genes for prognosis of HGA and the genes in the network module

As shown in Table 2, 598 risk genes for prognosis of HGA were significantly enriched in cell cycle of GO function and pathways such as cell cycle, mitotic as well as mitotic anaphase. While the genes in the network module mainly participated in GO function such as cell cycle, mitotic cell cycle and cell cycle process. Moreover, the genes in the network module mainly participated in the pathways such as cell cycle and cell cycle, mitotic.

3. 5 Drug target analysis

We searched 31 genes of the network module in PPI network in Drugbank databaseï¼Œand found that 7 genes were recorded in Drugbank database. Notably, there were as many as 32 drug records of *CHEK1* (Figure 4). Moreover, the AUC value of *CHEK1* was 0. 95833.

Discussion

HGA is a kind of aggressive brain tumor. Despite advances in the therapeutic management, the median survival was only few months [22]. Moreover, the molecular mechanism is still poorly known. In the current study, we exploited the bioinformatics methods to explore the underlying mechanism and screen the risk genes for prognosis of HGA. A total of 598 genes were identified as significant risk genes for prognosis. Eight risk genes for prognosis of HGA including *CDC6* and *AURKA* hadhigherdegrees in the PPI network. Moreover, *CDC6* , *AURKA* and *CHEK1* were mainly enriched in cell cycle and mitotic cell cycle.

Deregulation of cell cycle control is a main mechanism in the development of astrocytic gliomas [23]. As we all know, specific genes play crucial roles in cell cycle control, cell proliferation and growth [24]. In the present study, *CDC6* , *AURKA* and *CHEK1* were mainly enriched in cell cycle and mitotic cell cycle. *CDC6* is an essential and highly regulated component of DNA replication complex [25]. Moreover, *CDC6* is an oncogene, and aberrant expression of *CDC6* results in DNA damage and genetic instability [26]. Notably, the depletion of *CDC6* could inhibit the initiation of DNA replication in human tumor cells [27]. In addition, *CDC6* is involved in cell cycle checkpoint which coordinates S phase and mitotic entry. Deregulation of cell cycle checkpoints appears to be a universal phenomenon in cancers. *CHEK1* is a conserved Ser/Thr kinase mediating cell cycle arrest and is essential to preserve genome stability [28]. Luo et al. have demonstrated that *CHEK1* is a regulator of the checkpoints in G2/M and S phase in mammals [29]. Moreover, Syljuasen et al. reported that *CHEK1* was needed during normal S phase to avoid abnormally increased initiation of DNA replication [30].

CHEK1 has been reported to highly express in several tumor and cell lines and is associated with drug resistance and poor prognosis [31, 32]. The over-expressed CHEK1 was associated with poor

prognosis of hepatocellular carcinoma via target spleen tyrosine kinase [33]. Thus, we speculate that CHEK1 might be a poor prognosis biomarker in HGA and might be a underlying molecular target for HCC therapy.

AURKA , a serine-threonine kinase, is pivotal for mitosis involved in cell cycle regulation, centrosome duplication, spindle assembly and mitotic entry [34-36]. The over-expression and knockdown of AURKA cause the formation of abnormal mitotic spindles and genomic instability, which is thought to be an important mechanism of progression to malignancy [37, 38]. Recent studies have demonstrated that AURKA is over-expressed in gliomas [39, 40]. Furthermore, the over-expression of AURKA appears to be associated with poorer prognosis [41]. Norman L. Lehman et al. have reported that AURKA is significantly related with shorter survival in glioblastoma and Aurora A is a potential new drug target for the treatment of glial neoplasms [42]. Pramila Ramani et al. have also demonstrated that high AURKA is connected with adverse prognostic factors and shorter survival [43]. In light of these results, we infer that AURKA might be a promising prognostic factor of survival in HGA.

Taken together, our results imply that CDC6 , AURKA and CHEK1 might be used as prognostic biomarker for HGA and meanwhile our results might provide underlying molecular target for HCC therapy. However, subsequent

validations are needed to verify the role of *CDC6* , *AURKA* and *CHEK1* in the development of glioma in our study

Figures Legends

Figure 1A: Box plot of gene expressions of 26 samples before normalization. B: Box plot of gene expressions of 26 samples after normalization. The X axis stands for samples while the Y axis stands for expression level of genes after log 2 transformation. The black line in the center was the median of expression value, and the consistent distribution indicated a good standardization.

Figure2The ROC and KM curves of *CHEK1* , KCNJ6, *LOC283887* and *LTK* . The left drawing is ROC curve. The X axis stands for specificity while the Y axis stands for sensibility. The right drawing is KM curve. The X axis stands for survival time while the Y axis stands for survival probability.

Figure3A: Protein-protein interaction (PPI) network of risk genes for prognosis ofhigh grade astrocytomas. The thickness of the edges is proportional to combined score; the node size is proportional to the connectivity degree; the degree of color depth of nodes is proportional to AUC value of ROC curve, which means the color darker, the AUC value is higher. B: the network module of PPI network.

Figure4The recorded drug information genes in the module of PPI network in Drugbank database. Rectangular nodes stand for drug which is representatives with ID of Drugbank database. Oval nodes stand for genes.

Table 1Genes of AUC value more than 0. 95 of ROC curve

gene	limit_exp	AUC
KCNJ6	3. 100654675	0. 970833333
CHEK1	7. 722366626	0. 958333333
LOC28388 7	1. 370895764	0. 958333333
LTK	1. 835803618	0. 954166667

limit_exp: critical value of gene expression.

Table2Gene Ontology functional enrichment of the risk genes for prognosis of high grade astrocytomas

Ontology	Term	Pvalue	count
GO-BP	cell cycle [GO: 0007049]	4. 63E-07	74
GO-BP	mitotic cell cycle [GO: 0000278]	4. 64E-07	56
GO-BP	mitotic cell cycle process [GO:	0.	42

	1903047]	001353
GO-BP	regulation of cell cycle [GO: 0051726]	0. 001996	43
GO-BP	regulation of mitosis [GO: 0007088]	0. 004203	15
GO-CC	membrane-enclosed lumen [GO: 0031974]	8. 15E-05	107
GO-CC	organelle lumen [GO: 0043233]	1. 29E-04	105
GO-CC	intracellular organelle part [GO: 0044446]	2. 38E-04	177
GO-CC	nucleus [GO: 0005634]	5. 84E-04	158
GO-CC	nuclear part [GO: 0044428]	9. 74E-04	92
GO-MF	protein binding [GO: 0005515]	0. 007229	249

Table3The pathway enrichment of the risk genes for prognosis of high grade astrocytomas

pathway	Pvalue	count

Cell Cycle, Mitotic [REACT_152]	2. 58E-06	41
Cell Cycle [REACT_115566]	1. 56E-05	43
Mitotic Anaphase [REACT_1275]	0. 002613	19
Mitotic Metaphase and Anaphase [REACT_150314]	0. 002851	19
Mitotic M-M/G1 phases [REACT_21300]	0. 006075	27

Table4Gene Ontology functional enrichment of genes of the module in PPI network

Ontology	Term	Pvalue	count
GO-BP	cell cycle [GO: 0007049]	6. 40E-18	26
GO-BP	mitotic cell cycle [GO: 0000278]	1. 22E-17	23
GO-BP	cell cycle process [GO: 0022402]	1. 58E-	23

15

	pathway	pvalue	count
GO-BP	mitotic cell cycle process [GO: 1903047]	2. 16E-13	19
GO-BP	regulation of cell cycle [GO: 0051726]	1. 46E-11	18
GO-CC	non-membrane-bounded organelle [GO: 0043228]	8. 90E-07	21
GO-CC	intracellular non-membrane-bounded organelle [GO: 0043232]	8. 90E-07	21
GO-CC	chromosome [GO: 0005694]	1. 12E-06	12
GO-CC	spindle [GO: 0005819]	1. 34E-06	8
GO-CC	chromosome, centromeric region [GO: 0000775]	1. 72E-06	7

Table5The pathway enrichment of genes of the module in PPI network

pathway	pvalue	count
Cell Cycle [REACT_115566]	2. 90E-17	21

Cell Cycle, Mitotic [REACT_152]	7. 47E-17	20
Resolution of Sister Chromatid Cohesion [REACT_150425]	1. 99E-08	9
Aurora B signaling [aurora_b_pathway]	3. 02E-08	7
Mitotic Prometaphase [REACT_682]	3. 97E-08	9