 doi:10.20944/preprints202006.0126.v1

Predictive value of CCNB1, BUB1B and TTK in the progression and prognosis of lung adenocarcinoma

Running title: Predictive value of CCNB1, BUB1B and TTK

Xiong Lecai[1] MD, Bai Yuquan[1] MD, Zhu Minglin[1] MD, Ph.D, Yang Zetian[1] MD, Ph.D, Zhao Jinping[1]MD, Ph.D, Tang Hexiao[1*]MD, Ph.D

[1] Department of Thoracic Surgery, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

* Correspondence to: Tang Hexiao, MD, Ph.D, the Department of Thoracic Surgery, Zhongnan Hospital of Wuhan University, Wuhan 430071, China. Telephone number: 086-15102723563. E-mail: thx1245@sina.com.

Author's contributions:

(I) Conception and design: Tang Hexiao, Xiong Lecai and Bai Yuquan

(II) Administrative support: Zhao Jinping

(III) Collection and assembly of data: Xiong Lecai and Bai Yuquan

(IV) Data analysis and interpretation: Xiong Lecai and Bai Yuquan

(V) Manuscript writing and revision: All authors

(VI) Final approval of manuscript: All authors

 doi:10.20944/preprints202006.0126.v1

Abstract

Lung cancer predominates in cancer-related deaths worldwide, with lung adenocarcinoma (LUAD) being a common histological subtype of lung cancer. The aim at this study was to search for biomarkers associated with the progression and prognosis of LUAD. We have integrated the expression profiles of 1174 lung cancer patients from five GEO datasets (GSE18842, GSE19804, GSE30219, GSE40791 and GSE68465) and identified a set of differentially expressed genes. Functional enrichment analysis showed that these genes are closely related to the progression of LUAD, such as cell cycle, mitosis and adhesion. Cytoscape software was used to establish a protein-protein interaction (PPI) network to analyze important modules using Molecular Complex Detection (MCODE), and finally CCNB1, BUB1B and TTK were selected for further study. The study found that compared with non-tumor lung tissue, CCNB1, BUB1B and TTK are highly expressed in LUAD. Kaplan-Meier analysis showed that CCNB1, BUB1B and TTK were negatively correlated with the overall survival and disease-free survival of patients. Gene set enrichment analysis (GSEA) demonstrated that for the samples of any hub gene highly expressed, most of the functional gene sets enriched in cell cycle. In summary, CCNB1, BUB1B and TTK can be used as biomarkers of poor prognosis of LUAD. The high expression of CCNB1, BUB1B and TTK can accelerate the progression of LUAD and lead to shorter survival, suggesting that they may be potential targets for treatment in LUAD.

Keywords

CCNB1, BUB1B, TTK, lung cancer, protein-protein interaction network

 doi:10.20944/preprints202006.0126.v1

Introduction

In recent years, lung cancer has become the highest cancer incidence and mortality in the world [1]. Due to its high incidence, rapid progression and poor response to treatment, lung cancer has become one of the most serious malignant tumors that threaten human health and life [2,3]. According to the histopathological method, lung cancer can be roughly divided into two categories: non-small cell lung cancer (NSCLC) and small cell lung cancer, while NSCLC includes adenocarcinoma, squamous cell carcinoma and large cell carcinoma [4]. Lung adenocarcinoma, the most common form of NSCLC, accounts for about 40% of primary lung cancers and most of the patients are diagnosed as late stage [5]. Currently, the primary treatments for advanced LUAD remains chemotherapy [6] and targeted therapy [7]. But there are a lot of side effects of chemotherapy and targeted therapy, and the elderly and those with poor general conditions are hard to tolerate. The most notable example of lung cancer targeted therapies is the use of epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKI, mainly jefitinib and erlotinib) [8], but most patients have varying degrees of resistance to TKI. However, most patients with NSCLC have a very poor prognosis, especially in LUAD, with a 5-year survival of only 10-15% [9]. Therefore, the search for more effective biomarkers and new targeted drugs become more urgent.

Recently, genome-wide expression profiling has been used to identify prognostic features in cancer patients [10-12]. However, certain prognostic-related genes identified in one dataset may be difficult to validate in other cohorts [13]. In order to solve these problems, it is very necessary to verify the important role of the hub genes in

 doi:10.20944/preprints202006.0126.v1

independent research or in different populations.

In the present study, we used five datasets to compare changes in gene expression between tumor tissue and adjacent non-tumor lung (NTL) tissues and to overlap those differentially expressed genes (DEGs) that are involved in the development of LUAD. Finally, we identified 126 DEGs in LUAD and all of these belong to five datasets. Functional enrichment analysis of 126 DEGs found that genes were significantly enriched in the cell cycle. And then two important modules were analyzed by Molecular Complex Detection (MCODE) algorithm according to protein-protein interaction (PPI) network mapped by DEGs. Finally, CCNB1, BUB1B and TTK were identified for further study. Therefore, we further investigate and explore the predictive value of CCNB1, BUB1B and TTK for the progression and prognosis of LUAD. Our data show that high expression of CCNB1, BUB1B and TTK can promote the progression and lead to poor prognosis in LUAD.

Materials and Method

1. Data collection

The training set (GSE18842, GSE19804, GSE30219, GSE40791, GSE68465) based on the platform of Affymetrix (Affymetrix HG-U133 Plus 2.0 array and HG-U133A array) and the corresponding clinical data was obtained from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo). Three NSCLC genome-wide expression profiles were extracted from the following three data sets: GSE18842, which contains 46 tumors and 45 paracancer samples; GSE19804, which contains 60 pairs of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

matched tumors and adjacent normal samples; GSE30219, which contains 293 tumor samples and 14 non-tumor samples. Two LUAD whole-gene expression profiles: GSE40791, containing 94 tumor samples and 100 non-tumor samples; GSE68465, containing 443 tumor samples and 19 normal tissues. In this study, the test set GSE10072 was used to identify hub genes. The data set is based on the microarray platform of Affymetrix HT Human Genome U133A Array (HT_HG-U133A), comprising 58 LUAD and 49 NTL.

2. Data preprocessing

Download the original microarray data file (. CEL file) for the five data sets from the GEO database. The original microarray data was based on the "Affy" R package [14] for Robust Multichip Average (RMA) background correction, log2 transformation and normalization. Finally, the probe is annotated by the Affymetrix annotation file.

3. Differentially expressed genes (DEGs) screening

Five training sets were screened for DEGs using the "limma" R package [15 3716] for tumor tissue and NTL tissue groups. The false discovery rate (FDR) <0.05 and | log2 fold change (FC) |> 1.5 as cut-off criteria. Draw the volcanoes and venn diagram by the R package lattice and venn, respectively [16].

4. Functional enrichment analysis

The Database for Annotation, Visualization and Integrate Discovery (DAVID)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

database (http://david.abcc.ncifcrf.gov/) is an online program that provides researchers with a comprehensive set of functional annotation tools to understand the biological implications behind a large number of genes [17]. Gene Ontology (GO) consists of three main categories: Molecular function (MF), biological process (BP) and cellular component (CC). The Encyclopedia of Genes and Genomes (KEGG) is a database that links relevant gene sets with their pathways [18]. Functional annotation with P-value <0.05 was considered statistically significant. In this study, GO and KEGG were used to detect the enrichment of DEGs in biological implications and pathways.

5. PPI network analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) [19] database provides information on protein prediction and protein interactions. In this study, DEGs were mapped into PPI and protein pairs were extracted using combined score > 0.4 as a cut-off. In addition, PPI network were constructed using Cytoscape software version 3.2.1 [20]. Topological properties of the PPI network, including degree [21], closeness [22] and betweenness [23] centralities were determined using the R software package igraph, in order to analyze key genes in the network.

6. Module analysis and validation of hub gene

The network module is one of the features of the protein network and may contain specific biological implications. The most prominent clustering module was analyzed using the Molecular Complex Detection (MCODE) software package in Cytoscape. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

Cytoscape, the MCODE calculation is performed according to the cut-off criteria of degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. Depth = 100, and finally score> 6 as cut-off value to screen key modules. Next, use the DAVID online tool to analyze KEGG pathways for DEGs in key modules.

Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/) [24] was used to verify the progression and prognosis of hub genes in LUAD.

7. Patient tissue specimen

Tissue specimens from 20 patients with LUAD, were collected at the Department of thoracic surgery of Zhongnan Hospital of Wuhan University between August 2017 and February 2018. None of the patients had history of preoperative chemotherapy and radiotherapy.

8. Quantitative real-time PCR

Total RNA from LUAD tissues were isolated using RNeasy Mini kit (cat. no. 74101, Qiagen, Hilden, Germany) according to the manufacturer's instructions. The cDNA was synthesized using 1 μg of total RNA isolated by ReverTra Ace qPCR RT kit (Toyobo, Shanghai, China) and qRT-PCR was performed using 400 ng cDNA per 25 μl reaction. Primers used for CCNB1 , 3- CCTGCCTGCAACAGTACCC; CCNB1,5-CCAACACGATCTCTGGTCGC. BUB1B，3- TAGGGCGTTTATGCAATGAGC；BUB1B , 5- TCCTGAAATATCGCATCTGCTTT. TTK , 3-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

TCATGCCCATTTGGAAGAGTC； TTK， 5- CCACTTGGTTTAGATCCAGGC. GADPH ， 3-AGAAGGCTGGGGCTCATTTG ； GADPH ， 5-GCAGGAGGCATTGCTGATGAT, annealing temperature was 60˚C.

9. Gene set enrichment analysis (GSEA)

In the test set, LUAD samples were divided into two groups based on the expression level of hub genes. To identify potential functions of central genes, GSEA (http://software.broadinstitute.org/gsea/index.jsp) was performed to test whether a series of known biological processes are enriched in hub genes grouped by expression levels [25]. For use with GSEA software, the collection of annotated gene sets of c2.cp.kegg.v6.0.symbols.gmt in Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp) was chosen as the reference gene sets. FDR < 0.05, |enrichment score (ES) | > 0.5 and gene size ≥ 40 were chosen as the cut-off criteria.

Results

1. Identification of differentially expressed genes

Analysis of results shows that 1,387 DEGs (567 up-regulated genes and 820 down-regulated genes) were identified in GSE18842, 463 DEGs (124 up-regulated genes and 339 down-regulated genes) in GSE19804, 1,078 DEGs (326 up-regulated genes and 752 down-regulated genes) in GSE30219, 1,757 DEGs (584 up-regulated genes and 1173 down-regulated genes) in GSE40791, and 1,453 DEGs (768 up-regulated genes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

and 685 down-regulated genes) in GSE48465 (Figure 1A-E).

In addition, we performed an overlap analysis of DEGs in NSCLC and LUAD to identify genes that are specifically expressed in LUAD. A total of 314 genes were significantly differentially expressed in the three NSCLC datasets (Figure 1F). 422 genes were overlaped in the two LUAD datasets (Figure 1G). After the last overlap of these two subgroups of genes was further screened, 126 genes that affect the oncogenic of LUAD were identified (Figure 1H).

2. Function enrichment of DEGs

GO analysis was performed to determine the biological function of 126 DEGs. In three different areas of GO, genes involved in biological processes are mainly involved in the cell cycle and mitosis. Genes involved in molecular function are primarily involved in the binding of growth factors and chemokine activity. Genes associated with cellular components are primarily involved in spindle and microtubule cytoskeleton (Figure 2A). Then, we further studied the functional significance of DEGs in the development of LUAD by KEGG. The results showed that DEGs were enriched in six pathways. They are ECM-receptor interaction, Cell adhesion molecules (CAMs), PPAR signaling pathway, Complement and coagulation cascades, Cell cycle and Focal adhesion pathway (Figure 2B). Most of the enriched functions and pathways are closely related to the occurrence and development of cancer, which indicates that there is a significant relationship between DEGs and the progression and prognosis of LUAD.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

3. PPI network and module analysis to determine hub genes

126 DEGs were analyzed based on the STRING database, resulting in 378 proteins pairs with a combined score > 0.4 (Figure 3A). The top 12 most representative DEGs are listed according to degree, closeness and betweenness (Table 1). In these DEGs, CCNB1, BUB1B and TTK are involved in cell cycle and mitosis, and are both enriched in the cell cycle pathways.

Two modules with score> 6 (modules 1 and 2) were detected by MCODE (Figure 3B, C). Though the functional enrichment of the genes in the modules (Table 2), we found that the cell cycle pathway was identified as the most important pathway in module 1 (P = 3.99E-5), and the genes CCNB1, BUB1B and TTK in the cell cycle pathway had a higher degree (Table 1). And the genes in module 2 are predominantly enriched in the chemokine pathway (P = 9.02E-4), such as CXCL13, CXCL2, CXCR2.

4. CCNB1, BUB1B and TTK are overexpressed in LUAD

Since CCNB1, BUB1B and TTK are known to play an important role in the regulation of tumor cell cycle and mitosis, they were selected to further investigate their predictive value for the progression and prognosis of LUAD. It was identified in data sets GSE18842, GSE19804, GSE30219, GSE40791 that expression of CCNB1, BUB1B and TTK was significantly increased in LUAD tissues (Figure 4A, B and C). Using qRT-PCR to detect the expression of CCNB1, BUB1B and TTK in 20 pairs of LUAD, we found that CCNB1, BUB1B and TTK were highly expressed in LUAD compared to adjacent normal tissue (all p = 0.000) (Figure 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

5. Associations of CCNB1, BUB1B and TTK expression with progression and prognosis in LUAD

According to the GEPIA database, we found significant differences in CCNB1, BUB1B and TTK expression between different stages of LUAD (Figure 6A). In the training set GSE40719 and the test set GSE10072, a linear regression analysis showed a positive correlation between the three hub genes and the progression of LUAD (P for trend <0.001) (Figure 6B and C). In addition, we also found that the overall survival and disease-free survival of LUAD patients with high expression of CCNB1, BUB1B and TTK were significantly shorter (Figure 6D and E).

6. Gene set enrichment analysis (GSEA)

To identify potential function of the hub genes, GSEA was conducted respectively to search KEGG pathways enriched in the samples with the gene highly expressed. As a result, it was found that the samples with high expression levels of CCNB1, BUB1B and TTK were enriched in the following six pathways (Figure 7, Table 3), namely "Cell cycle", "Pyrimidine metabolism", "Protesome", "P53 signaling pathway" "Oocyte meiosis" and "RNA degradation".

Discussions

High-throughput analyzes are used to determine gene expression signatures for improved accuracy of prognosis [26]. To identify potential biomarkers in the prognosis

 doi:10.20944/preprints202006.0126.v1

and the treatment of LUAD, we integrated the gene expression profiles of 1174 LUAD patients in the five datasets from GEO and then obtained 126 DEGs by analysis. Finally, based on the degree of PPI network and MCODE algorithm to analyze important modules, we found that CCNB1, BUB1B and TTK are important in the cell cycle pathway.

CyclinB1 (CCNB1) is an important member in the cyclin family. Activated CCNB1 can promote cells to enter the M phase from G2 phase and initiate mitotic progression [27,28]. More and more studies have shown that CCNB1 is closely related to the abnormal proliferation of cells and the occurrence of tumors such as CCNB1 overexpression in liver cancer, breast cancer, esophageal cancer and cervical cancer [29-31]. Our study found that CCNB1 is highly expressed in LUAD tissues (Figure 4A, Figure 5A). Aaltonen et al. [32] reported that CCNB1 overexpression in breast cancer is closely related to tumorigenesis, malignant phenotype and poor prognosis. In LUAD patients, we found that CCNB1 overexpression is correlated with shorter overall survival and disease-free survival (Figure 6D and E). These results indicate that CCNB1 as a potential biomarker can predict the prognosis of LUAD patients.

Budding uninhibited by benzimidazoles 1 homolog beta (BUB1B) protein is an important functional protein. It ensures that the chromosome centromere links correctly with the microtubules to maintain genome stability. Prosecuting point defects in monitoring mechanisms can lead to premature segregation of chromosomes, leading to anomalies in the number of chromosomes that contribute to the development of tumors [33,34]. BUB1B was over-expressed in a variety of tumors, including renal and breast

 doi:10.20944/preprints202006.0126.v1

cancers, and there was a significant correlation between mutation and overexpression and chromosomal instability [35]. Our study found that BUB1B was overexpressed in LUAD tissues (Figure 4B, Figure 5B), suggesting that BUB1B is associated with the development of LUAD and can be used as a biomarker to predict the prognosis of LUAD.

Threonine and Tyrosine kinase (TTK) is a dual-specific protein kinase that can phosphorylates threonine and tyrosine [36]. TTK, a core component of the spindle assembly checkpoint (SAC), plays an important role in cell monitoring mechanisms that ensure healthy cell proliferation and precise division [37,38]. Therefore, the abnormal expression of TTK can affect the function of SAC, ultimately affecting the occurrence and progression of tumor. Studies have shown that in many human malignancies such as glioblastoma, thyroid cancer, breast cancer, liver cancer and pancreatic cancer, the expression level of TTK is significantly increased, and there is a significant correlation between this over-expression and poor survival prognosis [39-42]. Our study found that the expression of TTK was significantly elevated in LUAD (Figure 4C, Figure 5C) and the prognosis of LUAD patients with high TTK expression was poor (Figure 6D, E). These results suggest that TTK can be used as a biomarker to predict the prognosis of LUAD.

In addition, we also found that the expression of CCNB1, BUB1B and TTK were related with the tumor stage of LUAD based on the chipsets (GSE40791 and GSE10072) and the GEPIA database (Figure 6A, B and C). At the same time, overall survival and disease-free survival were significantly shorter in LUAD patients with high expression

 doi:10.20944/preprints202006.0126.v1

of BUB1B (Figure 6D, E). This indicates that CCNB1, BUB1B and TTK were negatively correlated with the overall survival and disease-free survival of patients. GSEA analysis found that most of the functional gene sets were enriched in the cell cycle pathway (Figure 7, Table 3).

In summary, this study shows that CCNB1, BUB1B and TTK are overexpressed in LUAD tissues and their upregulation can promote the progression of LUAD and lead to low survival and disease-free survival. Therefore, CCNB1, BUB1B and TTK can be used as prognostic indicators in LUAD patients.

Funding

None

Conflict of interest

All the authors declare that they have no conflict of interest.

Statement of Ethical

This article does not contain any studies with animals and humans performed by any of the authors.

Reference

1. Chuang KY, Chuang YC, Ho YS (2015) Global influence of Cancer Statistics articles. Curr Sci India 109 (9):1552-1554

2. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R,

Jemal A (2016) Cancer treatment and survivorship statistics, 2016. CA: a cancer journal for clinicians 66 (4):271-289. doi:10.3322/caac.21349

3. Cetin K, Ettinger DS, Hei YJ, O'Malley CD (2011) Survival by histologic subtype in stage IV nonsmall cell lung cancer based on data from the Surveillance, Epidemiology and End Results Program. Clinical epidemiology 3:139-148. doi:10.2147/CLEP.S17191

4. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba, II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES, Meyerson M (2007) Characterizing the cancer genome in lung adenocarcinoma. Nature 450 (7171):893-898. doi:10.1038/nature06358

5. Vijayvergia N, Mehra R (2014) Clinical challenges in targeting anaplastic lymphoma kinase in advanced non-small cell lung cancer. Cancer chemotherapy and pharmacology 74 (3):437-446. doi:10.1007/s00280-014-2517-6

6. Gao GH, Jiang JW, Liang XH, Zhou XL, Huang RF, Chu ZH, Zhan Q (2009) A meta-analysis of platinum plus gemcitabine or vinorelbine in the treatment of advanced non-small-cell lung cancer. Lung cancer 65 (3):339-344. doi:10.1016/j.lungcan.2008.11.019

7. Saito M, Suzuki H, Kono K, Takenoshita S, Kohno T (2018) Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. Surgery today 48 (1):1-8. doi:10.1007/s00595-017-1497-7

8. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB (2013) Screening for lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 143 (5 Suppl):e78S-e92S. doi:10.1378/chest.12-2350

9. Hassan KA, Wang L, Korkaya H, Chen G, Maillard I, Beer DG, Kalemkerian GP, Wicha MS (2013) Notch pathway activity identifies cells with cancer stem cell-like properties and correlates with worse survival in lung adenocarcinoma. Clinical cancer research : an official journal of the

American Association for Cancer Research 19 (8):1972-1980. doi:10.1158/1078-0432.CCR-12-0370

10. Andersson E, Arvidsson Y, Sward C, Hofving T, Wangberg B, Kristiansson E, Nilsson O (2016) Expression profiling of small intestinal neuroendocrine tumors identifies subgroups with clinical relevance, prognostic markers and therapeutic targets. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 29 (6):616-629. doi:10.1038/modpathol.2016.48

11. Chen F, Xiang CX, Zhou Y, Ao XS, Zhou DQ, Peng P, Zhang HQ, Liu HD, Huang X (2015) Gene expression profile for predicting survival of patients with meningioma. International journal of oncology 46 (2):791-797. doi:10.3892/ijo.2014.2779

12. Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JM, Macdonald J, Thomas D, Moskaluk C, Wang Y, Beer DG (2006) Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer research 66 (15):7466-7472. doi:10.1158/0008-5472.CAN-06-1191

13. Lohr M, Kollmann C, Freis E, Hellwig B, Hengstler JG, Ickstadt K, Rahnenfuhrer J (2012) Optimal strategies for sequential validation of significant features from high-dimensional genomic data. Journal of toxicology and environmental health Part A 75 (8-10):447-460. doi:10.1080/15287394.2012.674912

14. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy - analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20 (3):307-315. doi:10.1093/bioinformatics/btg405

15. Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43 (7). doi:ARTN e47

10.1093/nar/gkv007

16. Wang Y, Qian CY, Li XP, Zhang Y, He H, Wang J, Chen J, Cui JJ, Liu R, Zhou H, Xiao L, Xu XJ, Zheng Y, Fu YL, Chen ZY, Chen X, Zhang W, Ye CC, Zhou HH, Yin JY, Liu ZQ (2015) Genome-scale long noncoding RNA expression pattern in squamous cell lung cancer. Scientific reports 5. doi:Artn 11671

10.1038/Srep11671

17. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID:

Database for annotation, visualization, and integrated discovery. Genome biology 4 (9). doi:Artn R60

Doi 10.1186/Gb-2003-4-9-R60

18. Altermann E, Klaenhammer TR (2005) PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC genomics 6. doi:Artn 60 10.1186/1471-2164-6-60

19. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research 43 (D1):D447-D452. doi:10.1093/nar/gku1003

20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome research 13 (11):2498-2504. doi:10.1101/gr.1239303

21. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411 (6833):41-42. doi:Doi 10.1038/35075138

22. Estrada E, Rodriguez-Velazquez JA (2005) Subgraph centrality in complex networks. Phys Rev E 71 (5). doi:Artn 056103

10.1103/Physreve.71.056103

23. Goh KI, Oh E, Kahng B, Kim D (2003) Betweenness centrality correlation in social networks. Phys Rev E 67 (1). doi:Artn 017101

10.1103/Physrev.E.67.017101

24. Tang ZF, Li CW, Kang BX, Gao G, Li C, Zhang ZM (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research 45 (W1):W98-W102. doi:10.1093/nar/gkx247

25. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics 34 (3):267-273. doi:10.1038/ng1180

26. Ko JH, Ko EA, Gu W, Lim I, Bang H, Zhou T (2013) Expression profiling of ion channel

genes predicts clinical outcome in breast cancer. Molecular cancer 12 (1):106. doi:10.1186/1476-4598-12-106

27. Sengupta N, Gill KA, MacFie TS, Lai CS, Suraweera N, McDonald S, Silver A (2008) Management of colorectal cancer: a role for genetics in prevention and treatment? Pathology, research and practice 204 (7):469-477. doi:10.1016/j.prp.2008.03.003

28. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, Alteri R, Robbins AS, Jemal A (2014) Cancer treatment and survivorship statistics, 2014. CA: a cancer journal for clinicians 64 (4):252-271. doi:10.3322/caac.21235

29. Egloff AM, Vella LA, Finn OJ (2006) Cyclin B1 and other cyclins as tumor antigens in immunosurveillance and immunotherapy of cancer. Cancer research 66 (1):6-9. doi:10.1158/0008-5472.CAN-05-3389

30. Wang ZQ, Fan M, Candas D, Zhang TQ, Qin LL, Eldridge A, Wachsmann-Hogiu S, Ahmed KM, Chromy BA, Nantajit D, Duru N, He FC, Chen M, Finkel T, Weinstein LS, Li JJ (2014) Cyclin B1/Cdk1 Coordinates Mitochondrial Respiration for Cell-Cycle G2/M Progression. Dev Cell 29 (2):217-232. doi:10.1016/j.devcel.2014.03.012

31. Sanchez I, Dynlacht BD (2005) New insights into cyclins, CDKs, and cell cycle control. Semin Cell Dev Biol 16 (3):311-321. doi:10.1016/j.semcdb.2005.02.007

32. Aaltonen K, Amini RM, Heikkila P, Aittomaki K, Tamminen A, Nevanlinna H, Blomqvist C (2009) High cyclin B1 expression is associated with poor survival in breast cancer. British journal of cancer 100 (7):1055-1060. doi:10.1038/sj.bjc.6604874

33. Pinto M, Vieira J, Ribeiro FR, Soares MJ, Henrique R, Oliveira J, Jeronimo C, Teixeira MR (2008) Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas. Cellular oncology 30 (5):389-395. doi:10.3233/Clo-2008-0439

34. Myrie KA, Percy MJ, Azim JN, Neeley CK, Petty EM (2000) Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines. Cancer letters 152 (2):193-199. doi:Doi 10.1016/S0304-3835(00)00340-2

35. Scintu M, Vitale R, Prencipe M, Gallo AP, Bonghi L, Valori VM, Maiello E, Rinaldi M, Signori E, Rabitti C, Carella M, Dallapiccola B, Altomare V, Fazio VM, Parrella P (2007) Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast

carcinoma. Cancer letters 254 (2):298-307. doi:10.1016/j.canlet.2007.03.021

36. Fisk HA, Mattison CP, Winey M (2004) A field guide to the Mps1 family of protein kinases. Cell cycle 3 (4):439-442

37. Musacchio A (2011) Spindle assembly checkpoint: the third decade. Philosophical transactions of the Royal Society of London Series B, Biological sciences 366 (1584):3595-3604. doi:10.1098/rstb.2011.0072

38. Rutledge SD, Cimini D (2016) Consequences of aneuploidy in sickness and in health. Current opinion in cell biology 40:41-46. doi:10.1016/j.ceb.2016.02.003

39. Maia AR, de Man J, Boon U, Janssen A, Song JY, Omerzu M, Sterrenburg JG, Prinsen MB, Willemsen-Seegers N, de Roos JA, van Doornmalen AM, Uitdehaag JC, Kops GJ, Jonkers J, Buijsman RC, Zaman GJ, Medema RH (2015) Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 26 (10):2180-2192. doi:10.1093/annonc/mdv293

40. Miao R, Luo H, Zhou H, Li G, Bu D, Yang X, Zhao X, Zhang H, Liu S, Zhong Y, Zou Z, Zhao Y, Yu K, He L, Sang X, Zhong S, Huang J, Wu Y, Miksad RA, Robson SC, Jiang C, Zhao Y, Zhao H (2014) Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. Journal of hepatology 61 (4):840-849. doi:10.1016/j.jhep.2014.05.025

41. Liu X, Liao W, Yuan Q, Ou Y, Huang J (2015) TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. Oncotarget 6 (33):34309-34320. doi:10.18632/oncotarget.5295

42. Kaistha BP, Honstein T, Muller V, Bielak S, Sauer M, Kreider R, Fassan M, Scarpa A, Schmees C, Volkmer H, Gress TM, Buchholz M (2014) Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. British journal of cancer 111 (9):1780-1787. doi:10.1038/bjc.2014.460

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

Figures

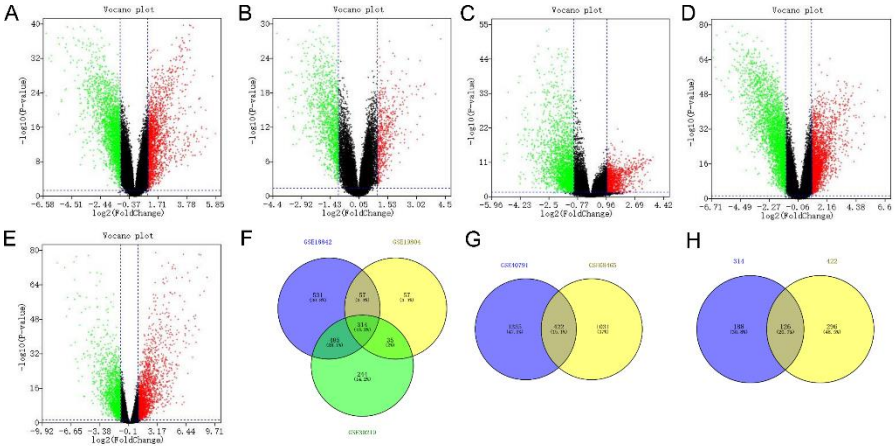

Figure 1. Identification of expression differences between tumor and NTL.

A-E. Volcano plot of the differential mRNA expression analysis. X-axis: log2 fold change; Y-axis: - log10 (P-value) for each probes; Vertical dotted lines: fold change ≥ 1.5 or ≤ 1.5; Horizontal dotted line: the significance cutoff (FDR P-value = 0.05). (A) There were 1,387 genes identified to be differentially expressed in GSE18842, including 567 up-regulated and 820 down-regulated genes. (B) 463 genes (124 up-regulated and 339 down-regulated genes) differentially expressed in GSE19804. (C) 1,078 genes (326 up-regulated and 752 down-regulated genes) differentially expressed in GSE30219. (D) 1,757 genes (584 up-regulated and 1,173 down-regulated genes) differentially expressed in GSE40791. (E) 1,453 genes (768 up-regulated and 685 down-regulated genes) differentially expressed in GSE68465. F-H. Overlap analysis between different datasets. (F) A total of 314 genes were significantly differentially expressed in the three NSCLC datasets. (G) 422 genes were overlapped in the two LUAD datasets. (H) There were 126 overlapping genes significantly differentially expressed between tumor and NTL in all five datasets.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

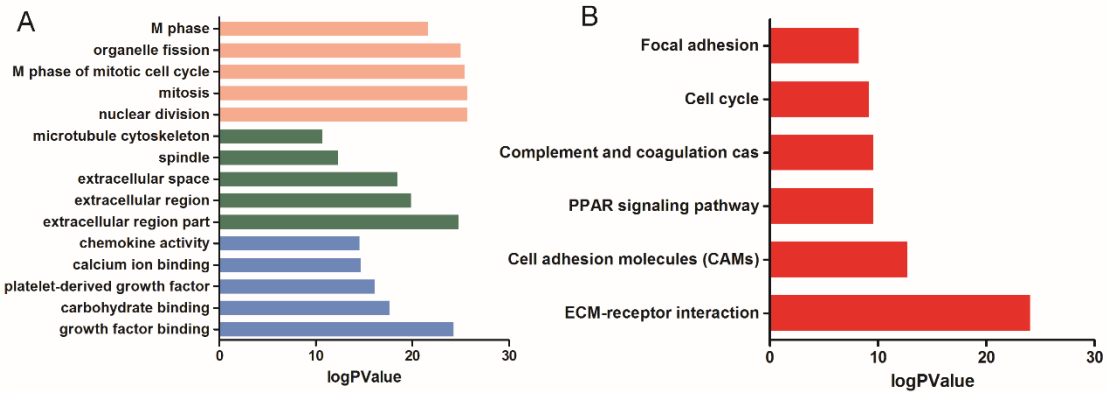

Figure 2. GO and pathway analysis of significant differentially expressed genes. (A) The top five significantly enriched GO categories were calculated. Red: Biological process; Green: Cellular component; Blue: Molecular function. (B) Gene networks identified through KEGG analysis of the differentially expressed genes.

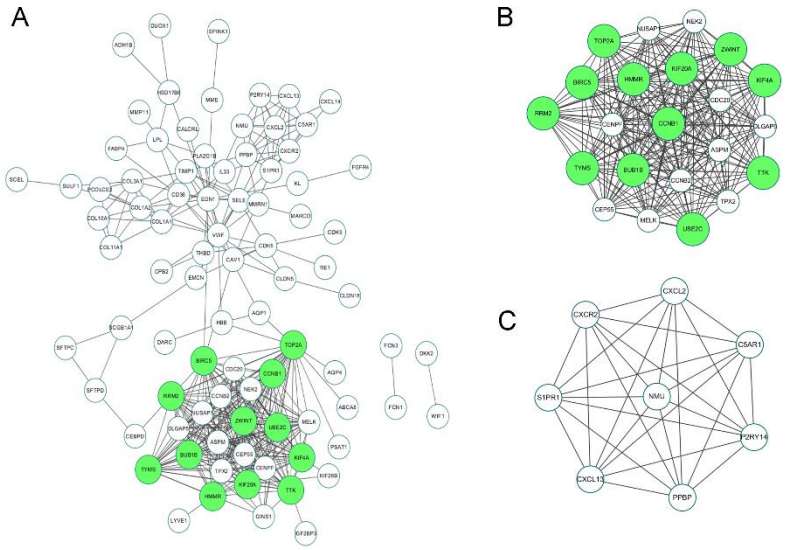

Figure 3. Two significant modules identified from the protein-protein interaction network using the molecular complex detection method with a score of >6.0. (A) Protein-protein interaction network of 126 DEGs. (B) Module 1: MCODE score=22. (C) Module 2: MCODE score=8. The green node is identified as the top 12 nodes according to the degree.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

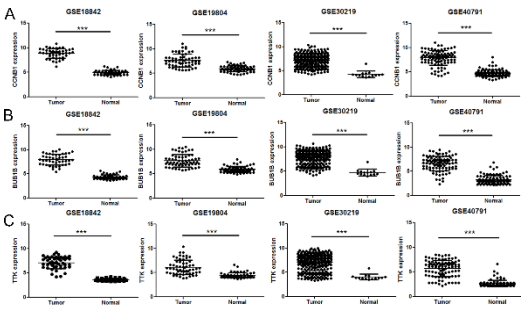

Figure 4. Identification of the differentially expressed genes. (A) Identification of mRNA expression of CCNB1 in four datasets, respectively. (B) Identification of mRNA expression of BUB1B in four datasets, respectively. (C) Identification of mRNA expression of TTK in four datasets, respectively. *** corresponds to P <0.001.

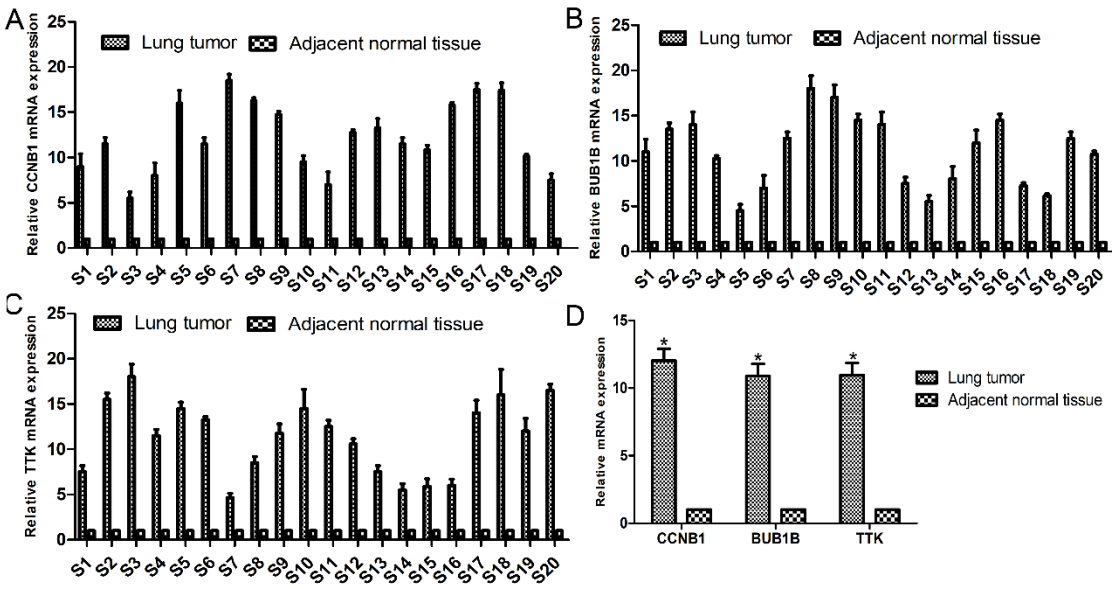

Figure 5. The expression of CCNB1, BUB1B and TTK in LUAD patient. (A, B, C) Relative mRNA expression of CCNB1, BUB1B and TTK in 20 patients with LUAD. (D) Quantitative analysis the differences of CCNB1, BUB1B, and TTK in patients with LUAD. * indicates that there is a difference compared to the adjacent normal tissue.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

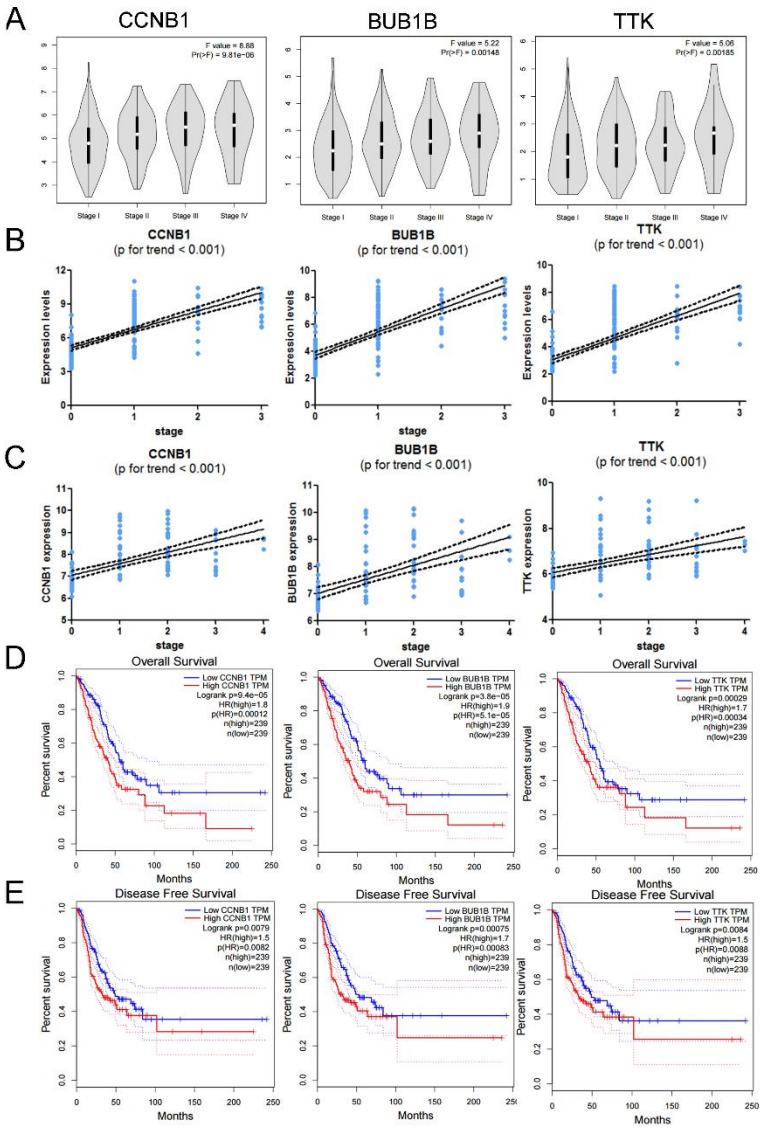

Figure 6. The role of CCNB1, BUB1B and TTK in the progression and prognosis of LUAD. (A) GEPIA database indicated that CCNB1, BUB1B and TTK had a strong correlation with the progression of LUAD based on TCGA data. (B) CCNB1, BUB1B and TTK expression were correlated with the disease progression of LUAD (GSE40791). (C) CCNB1, BUB1B and TTK expression were correlated with the disease progression of LUAD (GSE10072). (D-E) Kaplan-Meier survival curve obtained GEPIA database revealed that LUAD patients with higher expression of CCNB1, BUB1B and TTK had a significantly shorter (D) overall survival time and (E)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

disease free survival time.

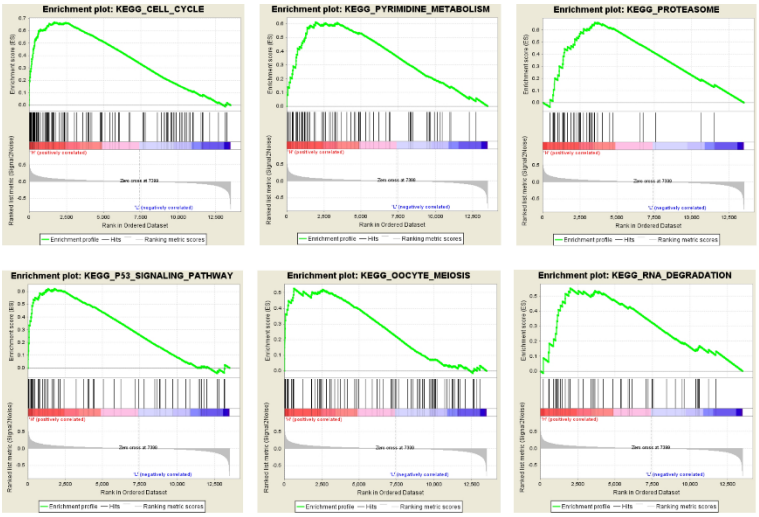

Figure 7. Gene set enrichment analysis (GSEA). Only listed the six most common functional gene sets enriched in LUAD samples with CCNB1 highly expressed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

Tables

Nodes	Degree	Nodes	Betweenness	Nodes	Closeness
TOP2A	27	EDN1	84	KL	0.000274
BIRC5	23	SELE	77.66667	CALCRL	0.000268
RRM2	23	VWF	69.16667	EDN1	0.000265
TTK	23	CD36	65.33333	CLDN5	0.000232
CCNB1	22	CAV1	63	THBD	0.000227
UBE2C	22	CDH5	57.5	TIE1	0.000226
KIF4A	22	BIRC5	56	NUSAP1	0.000225
KIF20A	22	PPBP	45.83333	CDH5	0.000224
BUB1B	22	COL1A2	45.16667	TOP2A	0.000221
ZWINT	22	CXCR2	45	CXCL2	0.000195
HMMR	22	S1PR1	17	CAV1	0.000189
TYMS	22	LPL	17	SELE	0.000186

Table 1. Nodes with higher values in degree, closeness and betweenness centrality.

Terms	Desription	PValue	Genes
	Moudle 1		
hsa04110	Cell cycle	3.99E-05	CCNB1, CCNB2, BUB1B, TTK, CDC20
hsa04115	p53 signaling pathway	0.005971	CCNB1, CCNB2, RRM2
hsa04114	Oocyte meiosis	0.015117	CCNB1, CCNB2, CDC20
	Moudle 2		
hsa04062	Chemokine signaling pathway	9.02E-04	PPBP, CXCL13, CXCL2, CXCR2
hsa04060	Cytokine-cytokine receptor interaction	0.002408	PPBP, CXCL13, CXCL2, CXCR2
hsa04080	Neuroactive ligand-receptor interaction	0.033105	C5AR1, S1PR1, P2RY14

Table 2. KEGG pathway enriched by differentially expressed genes in different modules (P<0.05).

CCNB1			
NAME	SIZE	ES	FDR q-val
KEGG_CELL_CYCLE	114	0.656634	0
KEGG_SPLICEOSOME	100	0.516428	0.007238
KEGG_OOCYTE_MEIOSIS	91	0.500848	0.013704
KEGG_PYRIMIDINE_METABOLISM	75	0.618308	0
KEGG_P53_SIGNALING_PATHWAY	60	0.636315	0
KEGG_RNA_DEGRADATION	46	0.554071	0.014575
KEGG_NUCLEOTIDE_EXCISION_REPAIR	42	0.538825	0.03323
KEGG_PROTEASOME	40	0.66882	1.45E-04

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2020 doi:10.20944/preprints202006.0126.v1

KEGG_GLUTATHIONE_METABOLISM	40	0.580912	0.01561
BUB1B			
NAME	SIZE	ES	FDR q-val
KEGG_CELL_CYCLE	114	0.669095	0
KEGG_OOCYTE_MEIOSIS	91	0.521107	0.012467
KEGG_PYRIMIDINE_METABOLISM	75	0.613621	0
KEGG_P53_SIGNALING_PATHWAY	60	0.649158	0
KEGG_RNA_DEGRADATION	46	0.562881	0.014839
KEGG_NUCLEOTIDE_EXCISION_REPAIR	42	0.559288	0.027549
KEGG_PROTEASOME	40	0.665843	3.35E-04
KEGG_GLUTATHIONE_METABOLISM	40	0.574304	0.025232
TTK			
NAME	SIZE	ES	FDR q-val
KEGG_CELL_CYCLE	114	0.67045	0
KEGG_PYRIMIDINE_METABOLISM	75	0.612532	3.74E-04
KEGG_PROTEASOME	40	0.661959	4.60E-04
KEGG_P53_SIGNALING_PATHWAY	60	0.622031	4.99E-04
KEGG_OOCYTE_MEIOSIS	91	0.527021	0.007285
KEGG_GLUTATHIONE_METABOLISM	40	0.593787	0.012024
KEGG_RNA_DEGRADATION	46	0.552928	0.026657

Table 3. A total of functional gene sets was enriched in the samples with high expression levels of CCNB1, BUB1B and TTK.