Identification of Key Genes Involved in Colorectal Cancer: A Bioinformatics Analysis

Dongxue Lu  
Nanjing University of Chinese Medicine

Jing Yan (yanjing@njucm.edu.cn)  
Clinical Medical College of Nanjing University of Traditional Chinese Medicine  
https://orcid.org/0000-0002-8206-2509

Zhiguang Sun  
Nanjing University of Chinese medicine

Xuan Li  
Affiliated Hospital of Nanjing University of Chinese Medicine: Jiangsu Province Academy of Traditional Chinese Medicine

Feng Liu  
Pukou Hospital of chinese medicine

Kun Zou  
Nanjing University of chinese medicine

Pinpin Ding  
Changzhou Hospital of Traditional chinese medicine

Haixia Liu  
Second Affiliated Hospital of Nanjing university of chinese medicine

Li Zhu  
Second Affiliated Hospital of Nanjing University of chinese medicine

Research article

Keywords: Colorectal cancer, Differentially expressed genes, Microarray, Biomarker, Bioinformatical analysis

DOI: https://doi.org/10.21203/rs.3.rs-95671/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background

Colorectal cancer is the life-threatening tumor with both high prevalence and mortality worldwide. However, the molecular mechanism behind it remains unknown. Methods: Herein, the potential prognostic candidate biomarkers for colorectal cancer were tested by Bioinformatical analysis combined with the CRC clinical samples. Three data sets (GSE32323, GSE44076 and GSE43078) were collected from the gene expression omnibus (GEO). The limma and clusterProfiler packages were used to identify differentially expressed genes (DEGs) and conduct functional enrichment analysis, respectively. To retrieve Interacting Genes (STRING) database, protein–protein interaction (PPI) network was built up using Search Tool, and Cytoscape was applied to carry out the module analysis. Subsequently, the online tool GEPIA was employed to conduct overall survival analysis (http://gepia.cancer-pku.cn/index.html), and the Oncomine database was used to analyze prognostic candidate biomarkers. Finally, 4 key hub genes were selected for validation of their expression levels in 9 patients newly diagnosed with CRC via reverse transcription-quantitative PCR (RT-qPCR). To evaluate the accuracy of prediction, time-dependent receiver operating characteristic (ROC) was applied.

Results

In total, 547 DEGs got identified, inclusive of 475 downregulated and 72 upregulated genes with a significant enrichment in the cellular response to hypoxia, the positive control of ERK1 and ERK2 cascade and the positive control of apoptotic process. The enhanced pathways were Pathways in cancer, PI3K-Akt signaling pathway, cGMP-PKG signaling pathway. Through the extraction of critical modules from the PPI network, 10 hub genes got removed. These 10 hub genes are all up-regulated genes and are highly expressed in colorectal cancer. Survival analysis shows that only CCNB1 and CCNA2 are associated with the survival prediction of colorectal cancer. Moreover, consistence is show between the TCGA data sets and the expression levels of the 4 hub genes. Receiver Operating Characteristic (ROC) curves showed that all CCNB1, CCNA2, AURKA and BUB1B have potential predictive value. Briefly, new hub genes identified can shed light on the underlying mechanism behind CRC carcinogenesis and development, which is conducive to detecting and treating CRC timely.

Background

Colorectal cancer remains the one of leading healthy issues worldwide, with as estimated 1.8 million new cases in 2018[1]. By 2030, the global CRC burden is expected to increase by 60%, more than 2.2 million new cases and 1.1 million deaths. It has been ranked the third and second cancer morbidity of male and female in China, respectively, and the mortality rate ranked the fifth[2]. The K-ras gene is present in approximately 45% of colorectal cancers[3] and is responsive to EGFR-targeted inhibitors. Related studies abroad have shown that for patients with early diagnosis of colorectal cancer, comprehensive treatment can make their 5-year survival rate reach to 82.9% [4], and once the tumor has metastasized, the 5-year
survival rate of the patient is <20%[5]. Lacking sensitive and specific early biomarkers, a high possibility of drug resistance and metastasis is considered to contribute the high mortality of this disease. Therefore, there has a pressing need for identifying the more sensitive and specific biomarkers or drug targets of CRC for developing effective diagnosis and treatment strategies.

Microarray technology provides an all-in-one system biology solution from hardware to software systems. It can simultaneously scan the hybridization signals of tens of thousands of gene probes in the chip and carry out quantitative analysis on the transcriptome profile of samples. Recent advances especially in the algorithms of probe signal detection and analysis, such as the introduction of artificial intelligence technologies, will make the results of microarray more accurate and reliable[6-8]. The microarray technique also provides a powerful tool for exploring the gene regulation pattern and molecular mechanisms involved in oncogenesis and progression of CRC. Recently, different types of biomarkers including coding genes, miRNAs, long non-coding RNAs and circRNAs have been identified in colorectal cancer. Dysregulation of these molecules is involved in the tumor progression or is associated with the prognosis of patients[9-12]. In view of the complexity of the molecular regulatory network of CRC, current studies on tumor biomarkers are not sufficient. Therefore, it is still necessary to identify novel prognostic biomarkers, which will help us develop more sensitive and effective diagnostic and therapeutic strategies. However, limited sample size and significant variability among different projects make it hard to obtain credible results. In this study, three microarray datasets containing mRNA expression data between CRC and non-cancerous tissues were downloaded from Gene Expression Omnibus (GEO) and the differentially expressed genes (DEGs) were screened out. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein-protein interaction (PPI) network analyses were performed to explore the key modules and hub genes involved in CRC progression. In sum, 547 DEGs and 10 hub genes were screened out, which may be candidate biomarkers for CRC.

**Methods**

In line with the process shown in Fig. 1, bioinformatics analysis was carried out. Through getGEO function in package GEOquery, three data sets (GSE32323, GSE44076 and GSE23878) containing the gene expression data on CRC and normal tissues were obtained from GEO ([https://www.ncbi.nlm.nih.gov/geo/](https://www.ncbi.nlm.nih.gov/geo/)) [13]. Table 1 shows the specifics of GEO data sets. With the aforementioned method applied, the raw expression files of three microarray datasets were processed preliminarily [14]. The selected chips were downloaded by the online tool GEO2R ([https://www.ncbi.nlm.nih.gov/geo/geo2r/](https://www.ncbi.nlm.nih.gov/geo/geo2r/)) and the differentially expressed genes were screened. The screening criterion is adj. P. Val < 0.01, |log2 Fold Change| ≥ 1, and the probe name was transformed into the standard gene name. Then, based on annotation information, the array probes were transformed into matching genetic symbols. When there are a number of probes matching the same gene, the value of gene expression was treated as the mean of the probes. When the same probe matches a number of genes, the probe gets removed. The possible regulatory mechanisms behind the occurrence and development of the disease can be revealed by the identification of DEGs at various states and investigation into their correlations. A web tool called ClustVis was applied to build up heatmaps. Differentially up-regulated and differentially
down-regulated genes were subject to screening for the three datasets. Finally, in the online Veeny tool (https://bioinfogp.cnb.csic.es/tools/venny/index.html), as were the differential intersection genes of the three microarrays.

**GO and KEGG enrichment analysis**

GO functional analysis is a useful method for annotating genes and identifying characteristic biological attributes for high-throughput genome or transcriptome data [15]. KEGG incorporates a wide range of databases, including those on genomes, biological pathways, diseases, drugs and chemical substances [16]. Gene Ontology and KEGG analyses were conducted using enrich GO and enrich KEGG functions of Database for Annotation, Visualization and Integrated Discovery (DAVID; http://david.ncifcrf.gov/), an online bioinformatics database that aims to provide tools for the functional interpretation of genes or proteins [17], respectively. P.adjust (FDR) < 0.05 was considered to be statistically significant.

**Construction of PPI network and module analysis**

The PPI network was constructed by Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/) with interaction score of 0.9 as the threshold [18]. Subsequently, Cytoscape software [19] was used to construct a PPI network and analyze the interactions of the DEGs. The cytoHubba plug-in was used to identify hub genes. The Molecular Complex Detection (MCODE) plug-in was used to screen modules of the PPI network in Cytoscape with a degree cut-off =2, node score cut-off =0.2, k-core =2 and max depth =100. The KEGG pathway enrichment analysis of genes in the top module was performed using DAVID.

**Hub gene analysis**

The Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/index.html) is an online tool, which delivers fast and customizable analysis for RNA sequencing expression data based on The Cancer Genome Atlas and GTEx databases, which include data on 9,736 tumors and 8,587 normal samples [20]. GEPIA provides key interactive and customizable functions, including differential expression analysis between cancer and normal tissues. The seed genes in modules referred to hub genes. The overall survival analyses were performed using online tool GEPIA (http://gepia.cancer-pku.cn/) . The logrank P < 0.05 was considered statistically significant. The association of expression level of hub genes with clinical traits were analysis using the Oncomine Database [21].

**Tissue specimens, RNA extraction and qRT-PCR analysis.**

9 tumor tissues and tumor-adjacent tissues from CRC donors were enrolled in our study to validate the expression levels of co-expressed DEGs. Prior patient consent and ethical approval from the ethics committee of the Jiangsu Provincial Hospital of Traditional Chinese Medicine. All methods were performed in accordance with the ethics guidelines and regulations. We selected 4 co-expressed DEGs, including CCNB1, CCNA2, AURKA and BUB1B, for validation. All tissues were histologically diagnosed. Total RNA from the tissue specimens was isolated using TRIzol reagent (Invitrogen, Carlsbad, California,
USA), and qRT-PCR was performed with SYBR® Green dye (TaKaRa, Shiga, Japan), following the manufacturer’s instructions. The primer sequences are provided in Supplementary Table S1. β-actin was used as a reference gene.

**Ethics statement.** The research protocol was subject to review and approval from the Research Ethics Committee of the First Affiliated Hospital. All of the experiments were carried out on the basis of the instructions from the First Affiliated Hospital, JiangSu Provincial Hospital of Traditional Chinese Medicine. Basic information of clinical patients were shown in supplement Table S2.

**Ethical approval.** All processes conducted in the research with human participants involved conformed to the ethical standards set out by the ethical committee of the First Affiliated Hospital, Nanjing University of Chinese Medicine, the 2019NL-KS1129 Declaration and the subsequent amendments to it. NO tissues were derived from prisoners.

**Ethical standards.** Expressed consent was received from all of those participating in this research.

With Graphpad Prime 8.3 used, the ROC curves were visualized to evaluate how accurate the prediction would be of independent prognostic hub genes such as CCNB1, CCNA2, AURKA and BUB1B. For comparing the discriminatory ability of above-mentioned prognostic genes, the area under the ROC curve (AUC) was calculated.

**Statistical analysis.** Data are indicated by the mean±SD for continuous variables. Both t-test and analysis of variance were conducted to assess the disparities of demographic data. All P values were two-sided, and P<0.05.

**Results**

**Identification of DEGs between CRC and tumor-adjacent tissues**

The three data sets were subject to standardization, with the results presented in Figure 2. Subsequently, the duplicate genes and values with no specific genetic symbols were removed. Totally, 3387 DEGs were obtained in GSE23878. Among them, 784 genes experienced up-regulation and 2603 genes underwent down-regulation. Besides, 4883 DEGs were obtained from GSE32323. There were 919 genes experiencing up-regulation and 3964 genes undergoing down-regulation. In GSE44076, 2485 DEGs were collected. Of them, 1254 were subject to up-regulation and 1231 were subject to down-regulation. Figure 3 presents the DEGs obtained from each data set. Clustering was performed using Euclidean distance. Figure 4 shows the top 100 DEGs conducted by heatmap.

**Data preprocessing and DEG screening**

Three GEO datasets were downloaded, pre-processed and merged into a global dataset which contained 160 CRC and 189 normal samples. Totally, 547 DEGs were identified by online tool GEO2R, including 72
up-regulated genes and 475 down-regulated genes (Figure 5). The most statistically significant up-
regulated and down-regulated genes are listed in Table 2.

**GO and KEGG analysis**

The biological functions and pathways analyses were conducted using DAVID 3.8.0 database. The GO
categories of biological process (BP), cellular component (CC) and molecular function (MF) were
enriched respectively (Figs. 6A–6C). The up-regulated DEGs were mainly associated with cellular
response to hypoxia (BP, GO:0071456), extracellular exosome (CC, GO:0070062) and heparin binding
(MF, GO:0008201). The down-regulated DEGs were most significantly related to positive regulation of
protein phosphorylation (BP, GO:0001934), Z disc (CC, GO:0030018) and magnesium ion binding (MF,
GO:0000287). The DEGs were mainly enriched in pathways of Mineral absorption (hsa04978), cGMP-PKG
signaling pathway (hsa04022), Proximal tubule bicarbonate reclamation (hsa04964), Aldosterone-
regulated sodium reabsorption (hsa04960), Insulin secretion (hsa04911), Bile secretion (hsa04976),
Pathways in cancer (hsa05200), Thyroid hormone signaling pathway (hsa04919) and Cell
cycle(hsa04110)(Table 3; Fig. 6D).

**Construction of PPI network and module analysis**

Protein–protein interaction network reflect the spatiotemporal relationship of macromolecules within the
cell which will provide valuable information about molecular mechanisms in physiological and
pathological process. To explore the molecular mechanisms underlying CRC progression, the online
STRING database was applied to construct the PPI network. The interaction score of 0.4 (highest
confidence) was set as threshold, and nodes without connections were removed from network. Finally, the
PPI network consisted of 542 nodes with 1338 edges, and average local clustering coefficient was 0.367
(PPI enrichment P-value < 1.0E-16) (Fig. 7A). Then, the hub genes were identified via CytoHubba
plugin. The top 10 hub genes, including minichromosome maintenance3 (MCM3), minichromosome
maintenance7 (MCM7), budding uninhibited by benzimidazoles 1 homolog beta (BUB1B), Cyclin
A2(CCN2A2),Cyclin B1(CCNB1), aurora kinase A(AURKA), targeting protein for Xklp2(TPX2), threonine and
tyrosine kinase (TTK), Flap endonuclease I (FEN1), kinesin family member 4A(KIF4A), were identified using
the cytoHubba plug-in, with a higher degree of connectivity (Fig. 7B). Finally, the key modules were
identified via MCODE plugin. A total of 22 functional clusters of modules and related hub genes were
detected. The top three significant modules were presented in Figs. 7C–7E. The KEGG analysis of
module genes revealed that the top three modules were mainly associated with the cell cycle (hsa04110),
Neuroactive ligand-receptor interaction (hsa04080), DNA replication (hsa03030), chemokine signaling
pathway (hsa04062) and Progesterone-mediated oocyte maturation (hsa04914) (Fig. 7F).

**Hub genes analysis**

A total of 10 genes were identified as hub genes. We selected 4 of the hub genes that are most closely
related to the occurrence and development, which were CRC, CCNB1, CCNA2, AURKA and BUB1B, and
analyzed their expression in colorectal cancer (Fig. 8A-D). Sometimes, the overall survival analysis of the
hub genes was performed using the online tool GEPIA(http://gepia.cancer-pku.cn/). Except CCNB1 and CCNA2, CRC patients with upregulated MM3, MCM7, BUB1B, AURKA, TPX2, TTK, FEN1, KIF4A showed no statistically significant on overall survival (Figs. 9A–H). It is worth noting that CCNB1 and CCNA2 are upregulated in CRC patients, but the high expression level is associated with better overall survival (Fig. 9I–J). Subsequently, the expression status of hub genes with Logrank P < 0.05 were further validated using the Oncomine database. The result showed that CCNB1 and CCNA2 were markedly overexpressed in CRC in the different datasets (Fig. 10A–B). In the Oncomine dataset, the alternation of CCNB1 and CCNA2 were associated with tumor grade (Fig. 11A–B), implicating vital roles of these genes in the carcinogenesis or progression of CRC.

The validation of core genes expression in clinical samples.

For ascertaining which gene is crucial to the development of CRC, real-time qPCR was applied to examine the expression of 4 DEGs through clinical sampling, inclusive of CCNB1, CCNA2, AURKA and BUB1B. Notably, relative to tumor-adjacent tissues, CCNB1, CCNA2, AURKA and BUB1B were normally subject to upregulation in colorectal cancer tissues, which conforms to the aforementioned results of bioinformatics analysis (Fig. 12A–D). As revealed by the ROC curve, all CCNB1 (AUC = 0.790, 95% CI: 0.570 to 1.01, P = 0.038) CCNA2 (AUC = 0.864, 95% CI: 0.697 to 1.03, P = 0.009) AURKA (AUC = 0.840, 95% CI: 0.653 to 1.03, P = 0.0152) and BUB1B (AUC = 0.901, 95% CI: 0.754 to 1.05, P = 0.0041) are potentially suitable for prediction (Fig. 12A–D).

Discussion

A number of pre-clinical and clinical studies have been conducted to reveal the underlying mechanisms of CRC in the past decades; however, the incidence and mortality of CRC remain high. This is primarily due to the majority of the studies focusing on a single genetic event, or the results were generated from a single cohort study [1]. In this study, three GEO datasets were analyzed and 547 DEGs were identified, including 72 up-regulated and 475 down-regulated genes. The 547 DEGs were classified into three groups (BP, CC and MF groups) by GO terms, and the KEGG pathway enrichment analysis of the DEGs was conducted using the DAVID database. Finally, the DEGs PPI network was constructed, and the top 10 hub genes and the most significant module was filtered from the PPI network.

The KEGG analysis revealed some DEGs were significantly enriched in Pathways in cancer, PI3K-Akt signaling pathway, cGMP-PKG signaling pathway. These annotation results provided valuable clues to reveal molecular interactions in the development of CRC. Additionally, the most significant module was filtered from the PPI network, among which the majority of the corresponding genes were mostly associated with cell cycle, Neuroactive ligand-receptor interaction and DNA replication. Studies have shown [1-4] that there are countless signaling pathways that play an important role in the cell cycle of cancer and the process of cell proliferation. The signaling pathway of the protein kinase B (PI3K)/ protein kinase B (PKB/ Akt) mammalian target of rapamycin (mTOR) is abnormally induced in a variety of human tumors, including the breast gland cancer, the pancreatic gland cancer, the non-small cell lung
cancer, CRC, etc. [5]. The activation of PI3K-Akt signaling pathway is closely related to colorectal cancer metastasis and poor prognosis [6-7]. Research shows [8] that a deregulated PI3K-AKT signaling pathway in CRC patients, which might serve as therapeutic target(s). The cell cycle is controlled by various mechanisms, which ensure correct cell division; loss of normal cell cycle control is a hallmark of cancer [9]. An increasing number of studies has revealed that targeting the deregulation of the cell cycle in cancer is a potential therapeutic strategy [10]. Therefore, investigating the cell cycle pathway may promote the understanding of carcinogenic mechanisms and insights into CRC treatment options.

These data may provide new ideas and directions for the mechanism research and therapeutic strategy of CRC. GO term enrichment analysis showed that the up-regulated DGEs were significantly enriched in cellular response to hypoxia, extracellular exosome, heparin binding, suggesting that some of these DEGs could locate in the nucleus and be involved in cell cycle processes to promote cell proliferation by enhancing DNA helicase activity in colorectal cancer. Increasing experimental evidence indicates that hypoxia is prevalent in the development of solid tumors. In order to ensure their survival, tumor cells under hypoxia can undergo a series of changes in biological behavior to adapt to hypoxia, specifically: (choose genotypes (such as TP53 mutation) that are conducive to survival of hypoxia repair [11], changes in expression of pro-survival genes [12] and inhibition of apoptosis [13], cell support for enhanced autophagy [14] and anabolic conversion [15], etc. Hypoxia can also enhance tumor angiogenesis [16], epithelial cells Interstitialization [17], immunosuppression [18], invasion and metastasis [19-20], etc. Through these changes, tumor cells can fight the body's autoimmunity and existing medical treatments, affecting their prognosis. It can be seen that the up-regulated DGEs play an important role in the occurrence and development of colorectal cancer.

Through integrated bioinformatics analysis, 10 hub genes were identified, including MCM3, MCM7, BUB1B, AURKA, TPX2, TTK, FEN1, cyclin B1 (CCNB1), cyclin A2 (CCNA2) and KIF4A with a high degree of connectivity. Among them, MCM3 and MCM7 belong to the MCMs family, and their biological functions are primarily involved in the DNA replication [2]. Multiple previous studies have demonstrated that minichromosome maintenance (MCM) protein additionally serve a role in the occurrence and development of varies malignancies [3-5]. The MCM participates in DNA synthesis [6] and can be used as a biomarker of oral squamous cell carcinoma [7], melanoma [8], glioma [9] and colon cancer [10]. In this study, the KEGG pathway showed that the DEGs were mainly involved in the cell cycle and DNA replication. CCNB1, one of the highly conserved cyclin family members, is involved in regulating cell cycle at the G2/M transition by forming maturation-promoting factor (MPF) with p34, suggesting that its over-expression can promote the progression of cancers [11-12]. AURKA gene mainly exists on chromosome 1 and encodes cell cycle regulated protein kinase, which is involved in chromosome mitosis and plays a role in tumorigenesis and development. BUB1B is an important constituent protein of the mitotic checkpoint, and is a multidomain protein kinase that responds to centromere tension [13-14]. Studies have demonstrated that BUB1B is overexpressed in various different types of tumor, such as colorectal cancer, renal and breast carcinoma, and its [15-16]. Therefore, further investigation on BUB1B may lead to a greater understanding of its importance in the CRC process, and novel ideas for investigating its molecular mechanisms and establishing more effective treatments [17]. The TPX2 gene is located at
20q11.2 and encodes a microtubule-associated protein involved in spindle assembly during cell mitosis. TPX2 overexpression is common to many tumor types. In hepatocellular carcinoma, it was correlated with increased proliferation, apoptosis inhibition, and induction of EMT [18]. In breast cancer, TPX2 silencing repressed PI3K/AKT and activated p53 signaling, which inhibited proliferation and promoted apoptosis [19]. As a DNA repair protein, Flap endonuclease 1 (FEN1) plays crucial parts in preventing carcinogenesis. Two functional germ line variants (-69G > A and 4150G > T) in the FEN1 gene have been associated with DNA damage levels in coke oven workers and lung cancer risk in general populations. Studies showed lower FEN1 expression may lead to higher risk for malignant transformation of gastrointestinal cells[20-21]. TTK is a class of bispecific kinases first discovered in Saccharomyces cerevisiae, which can phosphorylate serine, threonine, and tyrosine residues. It is the basic component of the spindle assembly checkpoint (SAC), and plays an important role in the replication of mitotic centrosomes and the correct separation of chromosomes[22]. TTK is overexpressed and proposed as a therapeutic target for CRC and other cancers, and closely related to poor clinical prognosis[23-26].

Some reports are contradictory to our findings, and this is potentially attributed to the heterogeneity of the tumor. Thus, it is necessary to further conduct sample functional verification. For the further verification of the expression of the 4 hub gene, RT-qPCR was carried out in 9 clinical samples. As illustrated in Figure 12, we noted up-regulation on the expression of the 4 hub genes in colorectal cancer tissues comparing to tumor-adjacent tissues, which were consistent with our bioinformatics analysis results above, indicating that they would be one of the effective biomarkers for cancer treatment.

Herein, a set of candidate biomarkers possibly significant to the development of CRC were identified. They can be taken as research subjects to investigate their roles in the development of diseases, which can improve our understanding as to the mechanism of CRC on a molecular level. Besides, to figure out their prognostic effects, clinical validation study can be conducted with them as possible prognostic biomarkers. Nevertheless, the study remains subject to some limitations. Firstly, it is premised on bioinformatics analysis of published data. Secondly, it is uncertain whether the correlation between the differential expression of hub genes and disease progression is a causal one. Finally, despite the combination of four GEO data sets, the sample size remains small, which can render the results unreliable. Thus, it is essential to conduct further bioinformatics analysis and experimental verification with a larger sample size.

**Conclusion**

In general, we have excavated biomarkers for predicting the recurrence, metastasis and prognosis of colorectal cancer through the big biological data (GEO, TCGA, Ocomine database). Among these biomarkers, CCNA2 and CCNB1 are closely related to the survival and prognosis of colorectal cancer, and can predict the patient's OS and DFS. More importantly, we selected CCNB1, CCNA2, AURKA and BUB1B for verification on clinical colorectal tumor samples, which were consistent with our bioinformatics analysis results above, indicating that they would be one of the effective biomarkers for cancer.
treatment. However, further prospective and multi-center, large-sample studies should be conducted to ensure these results and provide evidence for individualized treatment.

Declarations

Ethical Approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the ethical committee of the First Affiliated Hospital, Nanjing University of Chinese Medicine were obtained and with the 2019NL-KS1129 Declaration and its later amendments or comparable. NO tissues were procured from prisoners. All procedures performed in the present study involving human participants were in accordance with the ethical standards of institutional. Written informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Availability of Data and Materials

The Primer sequence and clinical sample information used to support the findings of this study are included within the supplementary information file.

Competing interests

The authors declare that they have no conflicts of interest.

Fundings

we have financial support for study, that is National Natural Science Foundation of CN(81673795, 81503536); A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions(PAPD)

Authors' Contribution

ZG S: pr_zhiguangsun@163.com, phD

Contribution: conceived the study question

DX L: 1606388689@qq.com, MD

Contribution: contributed to the study design, data analysis and interpretation, and writing the manuscript.

J Y: yanjing@njucm.edu.cn, phD
Contribution: contributed to the study design, data analysis and guide the writing the manuscript.

F L:399642091@qq.com, MD

Contribution: undertook data collection and data analysis.

Pinpin Ding:991653106@qq.com, MD

Contribution: undertook data collection and data analysis.

HX L:870346309@qq.com, MD

Contribution: undertook data collection and data analysis

K Z:522773765@qq.com, MD

Contribution: undertook clinical samples collection and data analysis

L Z:2013614449@qq.com, MD

Contribution: undertook data collection

X L:mingminglx@126.com,PhD

Contribution: contributed to the study design, data analysis and guide the writing the manuscript.

All authors have read and approved the manuscript.

Acknowledgement

We thank all the team members for their assistance during data collection and writing. We are also grateful to all the members of the faculty and staff in the College of Foreign Languages who have provided me with a lot of help and guidance.

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018, 68(6): 394-424

[2] FERLAY J, SOERJOMATARAM I, DIKSHIT R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns GLOBOCAN 2012[J]. International Journal of Cancer, 2015, 136(5): E359-E386

[3] Prieur A, Cappellini M, Habif G, et al. Targeting the Wnt Pathway and Cancer Stem Cells with Antiprogastri Humanized Antibodies as a Potential Treatment for K-RAS-Mutated Colorectal Cancer [J].
Clinical cancer research : an official journal of the American Association for Cancer Research, 2017, 23(17): 5267-80.

[4] Tsikitis VL, Malireddy K, Green EA, et al. Postoperative surveillance recommendations for early stage colon cancer based on results from the clinical outcomes of surgical therapy trial [J]. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 2009, 27(22): 3671-6.

[5] Adam R, Haller DG, Poston G, et al. Toward optimized front-line therapeutic strategies in patients with metastatic colorectal cancer—an expert review from the International Congress on Anti-Cancer Treatment (ICACT) 2009 [J]. Annals of oncology: official journal of the European Society for Medical Oncology, 2010, 21(8): 1579-84.

[6] Gan Z, Zeng N, Zou F, Chen J, Du M, Liao L-C, Li H, Zhang Y. Multilevel segmentation optimized by physical information for gridding of microarray images [J]. IEEE Access, 2019, 7:32146–32153

[7] Gan Z, Zou F, Zeng N, Xiong B, Liao L, Li H, Luo X, Du M. Wavelet denoising algorithm based on NDOA compressed sensing for fluorescence image of microarray [J]. IEEE Access, 2019b, 7:13338–13346

[8] Peng Y. 2006. A novel ensemble machine learning for robust microarray data classification [J]. Computers in Biology and Medicine, 36(6):553–573

[9] Kaisheng Liu, Yaomin Guo, Kai Zheng, Chang Zou, Haixiong Wu, Shaowang Wang, Ling Ou, Yifei Wang, Bowan Huang, Xiao Wang. Identification of the circRNA-miRNA-mRNA regulatory network of Hsp90 inhibitor-induced cell death in colorectal cancer by integrated analysis [J]. Gene, 2020, 727:144232.

[10] Yuanwen Zheng, Peihua Nie, Shifeng Xu. Long noncoding RNA CASC21 exerts an oncogenic role in colorectal cancer through regulating miR-7-5p/YAP1 axis [J]. Biomed Pharmacother, 2020, 121:109628.

[11] Pinelopi I. Artemaki, Aimilia D. Sklirou, Christos K. Kontos, Aikaterini-Anna Liosi, Despoina D. Gianniou, Jordanis N. Papadopoulos, Ioannis P. Trougakos, Andreas Scorilas. High clusterin (CLU) mRNA expression levels in tumors of colorectal cancer patients predict a poor prognostic outcome [J]. Clin Biochem, 2020, 75:62-69.

[12] Liu XiuLan, Yang Guangji, Huang Juan, Chai Li, Liu Xin, Dai Qiong, Yang ZhiHui. KRAS SNPs are related to colorectal cancer susceptibility and survival in Chinese people [J]. Biomark Med, 2020, 14(1):13-22.

[13] Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor [J]. Bioinformatics, 2007, 23(14):1846–1847

[14] Giulietti M, Occhipinti G, Principato G, Piva F. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development [J]. Cell Oncol (Dordr), 2016, 39(4):379-88.
[15] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: Tool for the unification of biology[J]. Gene Ontology Consortium. Nat Genet, 2000, 25: 25-29.

[16] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H and Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes[J]. Nucleic Acids Res, 1999, 27: 29-34.

[17] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources[J]. Nat Protoc. 2009, 4: 44–57.

[18] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life[J]. Nucleic Acids Res, 2015, 43: D447–D452.

[19] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks[J]. Genome Res, 2003, 13: 2498–2504.

[20] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA. A web server for cancer and normal gene expression profiling and interactive analyses[J]. Nucleic Acids Res, 2017, 45: W98–W102.

[21] Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia, 2007, 9(2): 166–180.

[22] Guo Y, Bao Y, Ma M and Yang W: Identification of key candidate genes and pathways in colorectal cancer by integrated bioinformatical analysis[J]. Int J Mol Sci, 2017, 18: E722.

[23] Tsai Te-Fu, Chen Po-Chun, Lin Yi-Chia, Chou Kuang-Yu, Chen Hung-En, Ho Chao-Yen, Lin Ji-Fan, Hwang Thomas I-S. Benzyl isothiocyanate promotes miR-99a expression through ERK/AP-1-dependent pathway in bladder cancer cell[J]. Environ Toxicol, 2020 Jan; 35(1): 47-54.

[24] Bai Jie, Zhang Xiaoyu, Kang Xiaoning, Jin Lijun, Wag Peng, Wang Zunyi. Screening of core genes and pathways in breast cancer development via comprehensive analysis of multi gene expression datasets. [J]. Oncol Lett, 2019, 18(6): 5821-5830.

[25] Meng Meng, Mingzhu Guo, Pei Wang, Ran Han, Jiaping Zhou, Chunling Wang. Chemopreventive effect of the polysaccharides from Grifola frondosa in colitis-associated colorectal cancer by modulating the Wnt/β-catenin/GSK-3β signaling pathway in C57BL/6 mice[J]. Journal of Functional Foods, 2019, 63.

[26] Sumanta K. Pal, Umang Swami, Neeraj Agarwal. Characterizing the Wnt Pathway in Advanced Prostate Cancer: When, Why, and How[J]. Eur Urol, 2020, 77(1): 22-23.
[27] Lo Russo P. Inhibition of the PI3K / AKT / mTOR pathway in solid tumors [J]. J Clin Oncol, 016, 34(31) : 3803-3815.

[28] Jun Kong. miR-101 and miR-26a inhibit the proliferation and metastasis of colorectal cancer by targeting ASPN to regulate PI3K/AKT signaling [D]. Jiangsu: University of Nantong, 2017.

[29] Jingyi Wu. The effect of B7-H3 and IDH1 on the prognosis of colorectal cancer and its mechanism [D]. Jiangsu, University of Wuxi, 2019.

[30] Zhang T, Ma Y, Fang J, Liu C, Chen L. A Deregulated PI3K-AKT Signaling Pathway in Patients with Colorectal Cancer [J]. J Gastrointest Cancer, 2019, 50(1): 35-41.

[31] Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer [J]. Cell Prolif, 2003, 36: 131-49.

[32] Aarts M1, Linardopoulos S, Turner NC. Tumour selective targeting of cell cycle kinases for cancer treatment [J]. Curr Opin Pharmacol, 2013, 13; 529-35.

[33] Zheng Zi-Hui, Wu Dong-Mei, Fan Shao-Hua, Zhang Zi-Feng, Chen Gui-Quan, Lu Jun. Upregulation of miR-675-5p induced by IncRNA H19 was associated with tumor progression and development by targeting tumor suppressor p53 in non-small cell lung cancer [J]. J Cell Biochem, 2019, 120(11): 18724-18735.

[34] Joshi S, Kumar S, Ponnusamy M P, Batra S K. Hypoxia-induced oxidative stress promotes MUC4 degradation via autophagy to enhance pancreatic cancer cells survival [J]. Oncogene, 2016, 35(45): 5882-5892.

[35] Yun Chul Won, Lee Jun Hee, Lee Sang Hun. Hypoxia-induced PGC-1α Regulates Mitochondrial Function and Tumorigenesis of Colorectal Cancer Cells [J]. Anticancer Res, 2019, 39(9): 4865-4876.

[36] Shentong Yu, Ru Zhou, Tong Yang, Zhuqing Cui, Jing Zhang. Hypoxia regulates autophagy in colorectal cancer cells through SIRT1 [J]. Progress in Modern Biomedicine, 2019, 19(13): 2438-2443.

[37] Xiang Lisha, Semenza Gregg L. Hypoxia-inducible factors promote breast cancer stem cell specification and maintenance in response to hypoxia or cytotoxic chemotherapy [J]. Adv Cancer Res, 2019, 141: 175-212.

[38] Luo Weibo, Wang Yingfei. Hypoxia Mediates Tumor Malignancy and Therapy Resistance [J]. Adv Exp Med Biol, 2019, 1136: 1-18.

[39] Sun Ling-Ling, Chen Chang-Ming, Zhang Jue, Wang Jing, Yang Cai-Zhi, Lin Li-Zhu. Glucose-Regulated Protein 78 Signaling Regulates Hypoxia-Induced Epithelial-Mesenchymal Transition in A549 Cells [J]. Front Oncol, 2019, 9: 137.
[40] Miska Jason, Lee-Chang Catalina, Rashidi Aida, Muroski Megan E, Chang Alan L, Lopez-Rosas Aurora, Zhang Peng, Panek Wojciech K, Cordero Alex, Han Yu, Ahmed Atique U, Chandel Navdeep S, Lesniak Maciej S. HIF-1α Is a Metabolic Switch between Glycolytic-Driven Migration and Oxidative Phosphorylation-Driven Immunosuppression of Tregs in Glioblastoma. [J]. Cell reports, 2019, 27(1): 226-237.

[41] Harper Kelly, Brochu-Gaudreau Karine, Saucier Caroline, Dubois Claire M. Hypoxia Downregulates LPP3 and Promotes the Spatial Segregation of ATX and LPP1 During Cancer Cell Invasion. [J]. Cancers (Basel), 2019, 11(9): 1403.

[42] Jiang Meng-Shi, Yin Xiao-Yi, Qin Bing, Xuan Shao-Yan, Yuan Xiao-Ling, Yin Hang, Zhu Chunqi, Li Xiang, Yang Jie, Du Yong-Zhong, Luo Li-Hua, You Jian. Inhibiting Hypoxia and Chemotherapy-Induced Cancer Cell Metastasis under a Valid Therapeutic Effect by an Assistance of Biomimetic Oxygen Delivery [J]. Mol Pharm, 2019, 16(11): 4530-4541.

[43] Yilun Zhang, Jinxuan Dong, Ziqi Zhang, Haishan Sun, Guohui Mei. Circle vector function applied to ray tracing in MCM for calculating effective emissivities of blackbody cavities [J]. Infrared Phys Technol, 2020, 104: 4538-4551.

[44] Yang Qing, Xie Binhui, Tang Hui, Meng Wei, Jia Changchang, Zhang Xiaomei, Zhang Yi, Zhang Jianwen, Li Heping, Fu Binsheng. Minichromosome maintenance 3 promotes hepatocellular carcinoma radioresistance by activating the NF-κB pathway [J]. J Exp Clin Cancer Res, 2019, 38(1): 263.

[45] Yang Yemei, Ma Shengfang, Ye Zi, Zhou Xianyi. MCM7 silencing promotes cutaneous melanoma cell autophagy and apoptosis by inactivating the AKT1/mTOR signaling pathway [J]. J Cell Biochem, 2020, 121(2): 1283-1294.

[46] Szelachowska Jolanta, Donizy Piotr, Ratajczak-Wielgomas Katarzyna, Halon Agnieszka, Zielecka-Debska Dominika, Lichon Krystian, Maciejczyk Adam, Lata-Wozniak Ewelina, Piotrowska Aleksandra, Matkowski Rafal. The effect of YAP expression in tumor cells and tumor stroma on the prognosis of patients with squamous cell carcinoma of the oral cavity floor and oral surface of the tongue [J]. Oncol Lett., 2019, 18(4): 3561-3570.

[47] D. Maiorano, M. Lutzmann, M. Mechali. MCM proteins and DNA replication [J]. Curr Opin Cell Biol, 2006, 18(2): 130-136.

[48] F. Rezazadeh, R. Ebrahimi, A. Andisheh-Tadbir, M. J. Ashraf, B. Khademi. Evaluation of the Ki-67 and MCM3 expression in cytologic smear of oral squamous cell carcinoma [J]. Journal of Dentistry (Shiraz), 2017, 18(3): 207-211.

[49] B. Nodin, M. Fridberg, L. Jonsson, J. Bergman, M. Uhlén, and K. Jirström, “High MCM3 expression is an independent biomarker of poor prognosis and correlates with reduced RBM3 expression in a prospective cohort of malignant melanoma” [J]. Diagn Pathol, 2012, 7: 82.
[50] C. Hua, G. Zhao, Y. Li, L. Bie. Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas[J]. BMC Cancer, 2014, 14: 526.

[51] Y. Ishibashi, T. Kinugasa, Y. Akagi et al. Minichromosome maintenance protein 7 is a risk factor for recurrence in patients with Dukes C colorectal cancer[J]. Anticancer Res, 2014, 34(8): 4569-75.

[52] Wang Shu, Sun Hongmei, Zhan Xiaokai, Wang Qiwen. MicroRNA-718 serves a tumor-suppressive role in non-small cell lung cancer by directly targeting CCNB1[J]. Int J Mol Med, 2020, 45(1): 33-44.

[53] Li Bin, Zhu Hai-Bo, Song Gui-Dong, Cheng Jian-Hua, Li Chu-Zhong, Zhang Ya-Zhuo, Zhao Peng. Regulating the CCNB1 gene can affect cell proliferation and apoptosis in pituitary adenomas and activate epithelial-to-mesenchymal transition[J]. Oncol Lett, 2019, 18(5): 4651-4658.

[54] Jun DY, Lee JY, Park HS, Lee YH and Kim YH. Tumor suppressor protein p53-mediated repression of human mitotic centromere-associated kinesin gene expression is exerted via down-regulation of Sp1 level[J]. PLoS One, 2017, 12(2): e0189698.

[55] Pinto M, Vieira J, Ribeiro FR, Soares MJ, Henrique R, Oliveira J, Jerónimo C and Teixeira MR: Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas[J]. Cell Oncol, 2008, 30(5): 389-395.

[56] Myrie KA, Percy MJ, Azim JN, Neeley CK, Petty EM. Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines[J]. Cancer Lett, 2000, 152(2): 193-9.

[57] Scintu M, Vitale R, Prencipe M, Gallo AP, Bonghi L, Valori VM, Maiello E, Rinaldi M, Signori E, Rabitti C, et al: Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma[J]. Cancer Lett, 2007, 254(2): 298-307.

[58] Hahn Marc-Manuel, Vreede Lilian, Bemelmans Sonja A S A, van der Looij Erica, van Kessel Ad Geurts, Schackert Hans K, Ligtenberg Marjolijn J L, Hoogerbrugge Nicoline, Kuiper Roland P, de Voer Richarda M. Prevalence of germline mutations in the spindle assembly checkpoint gene BUB1B in individuals with early-onset colorectal cancer[J]. Genes, chromosomes & cancer, 2016, 55(11): 855-63.

[59] Liang B, Jia C, Huang Y, He H, Li J, Liao H, Liu X, Liu X, Bai X, Yang D. TPX2 Level Correlates with Hepatocellular Carcinoma Cell Proliferation[J]. Apoptosis, and EMT. Dig Dis Sci. 2015; 60: 2360–72.

[60] Yang Y, Li DP, Shen N, Yu XC, Li JB, Song Q, Zhang JH. TPX2 promotes migration and invasion of human breast cancer cells[J]. Asian Pac J Trop Med, 2015, 8: 1064–70.

[61] Liu Li, Zhou Changchun, Zhou Liqing, Peng Li, Li Dapeng, Zhang Xiaojiao, Zhou Mo, Kuang Pengqun, Yuan Qipeng, Song Xianrang. Yang Ming. Functional FEN1 genetic variants contribute to risk of
hepatocellular carcinoma, esophageal cancer, gastric cancer and colorectal cancer[J]. Carcinogenesis, 2012, 33(1): 119-23.

[62] Sun H, He L, Wu H, Pan F, Wu X, Zhao J, Hu Z, Sekhar C, Li H, Zheng L, Chen H, Shen B H, Guo Z. The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation[J]. Oncogene, 2017, 36(2): 194-207.

[63] Jacqueline M. Masona, Xin Weia, Graham C. Fletchera, Reza Kiarasha, Richard Brokxa, Richard Hodgsona, Irina Beletskayaa, Mark R. Brayaa, and Tak W. Mak. Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer[J]. Proc. Natl. Acad. Sci. U.S.A, 2017, 114(12): 3127-3132.

[64] Zhang Li, Jiang Baofei, Zhu Ni, Tao Mingyue, Jun Yali, Chen Xiaofei, Wang Qilong, Luo Chao. Mitotic checkpoint kinase Mps1/TTK predicts prognosis of colon cancer patients and regulates tumor proliferation and differentiation via PKCα/ERK1/2 and PI3K/Akt pathway[J]. Med Oncol, 2019, 37(1): 5.

[65] Liu X D, Yao D W, Xin F. TTK contributes to tumor growth and metastasis of clear cell renal cell carcinoma by inducing cell proliferation and invasion[J]. Neoplasma, 2019, 66(6): 946-953.

[66] Chen Xiangling, Yu Chengli, Gao Jing, Zhu Hongwen, Cui Binghai, Zhang Tao, Zhou Yanting, Liu Qian, He Han, Xiao Ruoxuan, Huang Ruimin, Xie Hua, Gao Daming, Zhou Hu. A novel USP9X substrate TTK contributes to tumorigenesis in non-small-cell lung cancer[J]. Theranostics, 2018, 8(9): 2348-2360.

[67] Pachis Spyridon T, Kops Geert J P L. Leader of the SAC: molecular mechanisms of Mps1/TTK regulation in mitosis[J]. Open Biol, 2018, 8(8): 180109

Tables

| Table 1 The detail information of three GEO Datasets |
| ID | Tissue | Platform | Normal | Tumor |
|-----|--------|----------|--------|-------|
| GSE32323 | CRC | GPL570 | 17 | 27 |
| GSE44076 | CRC | GPL13667 | 148 | 98 |
| GSE23878 | CRC | GPL570 | 24 | 35 |

GEO: gene expression omnibus; CRC: colorectal cancer
### Table 2 Top ten up- and down-regulated DEGs.

| Up-regulated DEGs | Gene symbol | logFC | Adjust P | Down-regulated DEGs | Gene symbol | logFC | Adjust P |
|-------------------|-------------|-------|----------|----------------------|-------------|-------|----------|
|                   | FABP6       | 2.5236035 | 0.00000002 | SST                  | -5.0158362 | 6.83E-12 |
|                   | SLC01B3     | 3.1640579 | 0.000142 | CHRDL1               | -4.7149542 | 9.14E-12 |
|                   | SLC7A5      | 2.4605643 | 2.09E-10 | CA4                  | -4.7812383 | 0.000000286 |
|                   | CLDN1       | 3.8365233 | 0.000347 | LGALS2               | -4.8012694 | 0.00000015 |
|                   | ASCL2       | 2.7737687 | 0.000138 | MMP28                | -4.5484346 | 7.19E-13 |
|                   | CRNDE       | 4.2755368 | 1.59E-12 | AKR1B10              | -5.1512725 | 0.00000145 |
|                   | TRIB3       | 2.866854 | 1.32E-09 | MT1F                 | -5.2634923 | 1.22E-08 |
|                   | ADH1B       | 2.4665976 | 0.000947 | PADI2                | -6.2011828 | 2.27E-08 |
|                   | ABCA8       | 2.722522 | 0.00324 | LAMA1                | -5.3343777 | 0.000000294 |
|                   | GCNT2       | 2.3043753 | 0.00435 | SLC25A34             | -4.7128069 | 2.61E-08 |

### Table 3 KEGG enriched pathways of DEGs

| ID     | Description                                      | Count | p adjust (FDR) |
|--------|--------------------------------------------------|-------|----------------|
| hsa04978 | Mineral absorption                               | 9     | 6.96E-05       |
| hsa04022 | cGMP-PKG signaling pathway                       | 13    | 0.005089719    |
| hsa04964 | Proximal tubule bicarbonate reclamation         | 5     | 0.005859887    |
| hsa04960 | Aldosterone-regulated sodium reabsorption        | 6     | 0.008022768    |
| hsa04911 | Insulin secretion                                | 8     | 0.019882876    |
| hsa04976 | Bile secretion                                   | 7     | 0.023805229    |
| hsa05200 | Pathways in cancer                               | 21    | 0.029453784    |
| hsa04151 | PI3K-Akt signaling pathway                       | 19    | 0.03041787     |
| hsa04919 | Thyroid hormone signaling pathway                | 9     | 0.032753173    |
| hsa04110 | Cell cycle                                       | 9     | 0.047771444    |
Figure 1

Flow diagram of bioinformatics analysis.
Figure 2

Standardization of gene expression (A) Standardization of GSE23878, (B) standardization of GSE32323, (C) standardization of GSE44076. The blue bar represents the data before normalization, and the red bar represents the data after normalization.
Figure 3

Heatmap of the top 100DEGs identified in (A) GSE23878, (B) GSE32323 and (C) GSE44076. Genes up-regulated are in red. Genes down-regulated are in blue. The differences are set as |FC| > 2 and corrected P-value < 0.05.
Figure 4

Volcano plots of DEGs between CRC and tumor-adjacent tissues (A) GSE23878, (B) GSE32323 and (C) GSE44076. Data points in red represent up-regulated and green represent down-regulated genes. Genes without any significant difference are in grey. The differences are set as $|FC| > 2$ and corrected P-value <0.05.
Figure 5

Identification of consistent DEGs between primary CRC tumor tissues and normal tissues. (A) Venn plot for consistent up-regulated DEGs. (B) Venn plot for consistent down-regulated DEGs

Figure 6

GO and KEGG analysis of DEGs. (A–C) The top 15 terms of GO categories of biological process (BP), cellular component (CC) and molecular function MF, respectively. (D) KEGG pathway analysis of DEGs, p.adjust (FDR) < 0.05 was considered significantly. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate.
Figure 7

Protein–protein interaction network of DEGs and modules analysis. (A) PPI network of DEGs; (B) Top 10 hub genes with a higher degree of connectivity. (C–E) the top 3 modules of PPI network. (F) The KEGG enrichment analysis of the genes in the top three modules. PPI, protein–protein interaction; DEG, differentially expressed gene; KEGG, Kyoto Encyclopedia of Genes and Genomes.
Figure 8

The expression level of the selected 4 hub genes in the CRC tissues and tumor-adjacent tissues from TCGA. (A) CCNB1; (B) CCNA2; (3) AURKA; (4) BUB1B.
Figure 9

The overall survival analysis of the 10 hub genes (A-J).

| Median Rank | p-Value    | Gene   |
|-------------|------------|--------|
| 236.0       | 4.54E-8    | CCNB1  |
| 448.0       | 6.49E-14   | CCNA2  |

Figure 10

Oncomine analysis of cancer vs. normal tissue of CCNB1 and CCNA2. Heat maps of CCNB1 (A), CCNA2 (B) gene expression in colorectal cancer samples vs. normal tissues. 1. Rectal Adenocarcinoma vs. Normal Gaedcke Colorectal, Genes Chromosomes Cancer, 2010; 2. Colorectal Adenoma Epithelia vs. Normal Gaspar Colon, Am J Pathol, 2008; 3. Colon Adenoma vs. Normal Sabates-Bellver Colon, Mol Cancer Res, 2007; 4. Rectal Adenoma vs. Normal Sabates-Bellver Colon, Mol Cancer Res, 2007; 5. Cecum Adenocarcinoma vs. Normal TCGA Colorectal, No Associated Paper, 2011; 6. Colon Mucinous
Adenocarcinoma vs. Normal TCGA Colorectal, No Associated Paper, 2011;
7. Rectal Mucinous Adenocarcinoma vs. Normal TCGA Colorectal, No Associated Paper, 2011;
8. Rectosigmoid Adenocarcinoma vs. Normal TCGA Colorectal, No Associated Paper, 2011.

Figure 11

The association between the expression of selected hub genes and tumor stage. (A-B) The expressions of CCNB1, CCNA2 were correlated with tumor stage in the Oncomine dataset. 0: No value; 1: Stage I; 2: Stage II; 3: Stage IIA; 4: Stage II B; 5: Stage III; 6: Stage III A; 7: Stage III B; 8: Stage III C; 9: Stage IV; 10: Stage IVA.

Figure 12
The expression level of the selected 4 hub genes in the CRC tissues and tumor-adjacent tissues from clinical samples and ROC curve analysis of the 4 hub genes. (A) CCNB1; (B) CCNA2; (3) AURKA; (4) BUB1B. Data are presented as mean ± S.D. P < 0.05 was considered as significant.