Observational Study

Medicine

OPEN

Identification of potential biomarkers for colorectal cancer by clinical database analysis and Kaplan–Meier curves analysis

Chongyang Li, MM[a,b], Ying Gao, MD[b], Chunlei Lu, MD[b], Mingxiao Guo, MD[b,*]

Abstract

This study aimed to explore critical genes as potential biomarkers for the diagnosis and prognosis of colorectal cancer (CRC) for clinical utility. To identify and screen candidate genes involved in CRC carcinogenesis and disease progression, we downloaded microarray datasets GSE89076, GSE73360, and GSE32323 from the GEO database identified differentially expressed genes (DEGs), and performed a functional enrichment analysis. A protein-protein interaction network was constructed, and correlated module analysis was performed using STRING and Cytoscape. The Kaplan–Meier survival curve shows the survival of the hub genes. The expression of cyclin-dependent kinase (CDK1), cyclin B1 (CCNB1), and PCNA in tissues and changes in tumor grade were analyzed. A total of 329 DEGs were identified, including 264 upregulated and 65 downregulated genes. The functions and pathways of DEGs include the mitotic cell cycle, poly(A) RNA binding replication, ATP binding, DNA replication, ribosome biogenesis in eukaryotes, and RNA transport. Forty-seven Hub genes were identified, and biological process analysis showed that these genes were mainly enriched in cell cycle and DNA replication. Patients with mutations in CDK1, PCNA, and CCNB1 had poorer survival rates. CDK1, PCNA, and CCNB1 were significantly overexpressed in the tumor tissues. The expression of CDK1 and CCNB1 gradually decreased with increasing tumor grade. CDK1, CCNB1, and PCNA can be used as potential markers for the diagnosis and prognosis of CRC. These genes are overexpressed in colon cancer tissues and are associated with low survival rates in CRC patients.

Abbreviations: BP = biological process, CCNB1 = cyclin B1, CDK1 = cyclin-dependent kinase, CDK1 = cyclin-dependent kinase 1, CRC = colorectal cancer, DEGs = differentially expressed genes, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, K-M = Kaplan–Meier, MAD2L1 = mitotic arrest deficient 2 like 1, MCM7 = minichromosome maintenance complex component 7, MCODE = molecular complex detection, MF = molecular function, PPI = protein-protein interaction, UCSC = University of California, Santa Cruz.

Keywords: biomarkers, CCNB1, CDK1, clinical database, colorectal cancer, PCNA

1. Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer-related deaths globally. It is the most common cancer diagnosis in women and the third most common cancer diagnosis in men. The number of patients aged <50 with CRC (especially rectal cancer and left hemicolectomy) is increasing.

Moreover, related epidemiological studies have found that the incidence in males is closely related to the increase in age.[1] Increasing evidence has also shown that the progression of CRC is strongly associated with genetic mutations. Chen et al found that epidermal growth factor receptor and its ligands amphiregulin and epiregulin play a central role in the development of CRC and the expression of amphiregulin and epiregulin is independent of CRC patients. As a prognostic marker, ligand mRNA expression in primary tumors correlates with survival in patients with metastatic CRC.[2] Studies have found that abnormal expression of TGFBI is regulated by DNA methylation and that the expression has clinical significance in the diagnosis of CRC.[3] TGFBI promotes tumor growth and angiogenesis.[4] Given the high incidence and mortality of CRC, identifying new biomarkers to reveal its pathogenesis, predict clinical outcomes, and enable personalized therapies for patients is of great importance and desperately needed.

With people's understanding of cancer and the development of science and technology, microarray technology and clinical database analysis have gradually become widely used to screen genetic changes in related genomes,[5] which helps us identify

This study was supported by the National Natural Science Foundation of China (Grant no.81500688) and Shandong Provincial Natural Science Foundation, China (ZR2021MH362, ZR2015HL033).

The authors have no conflicts of interest to disclose.

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

[a] Second Clinical Medical College, Binzhou Medical University, Yantai, China, [b] Department of General Surgery Center, Linyi People's Hospital, Linyi, China.

** Correspondence: Mingxiao Guo, Department of General Surgery Center, Linyi People's Hospital, 27 East Jiefang Road, Linyi 276000, China (e-mail: gmx1211@163.com).*

How to cite this article: Li C, Gao Y, Lu C, Guo M. Identification of potential biomarkers for colorectal cancer by clinical database analysis and Kaplan–Meier curves analysis. Medicine 2023;102:6(e32877).

Received: 3 November 2022 / Received in final form: 11 January 2023 / Accepted: 18 January 2023

http://dx.doi.org/10.1097/MD.0000000000032877

differentially expressed genes (DEGs) and possible molecular functional pathways involved in CRC carcinogenesis and progression. Therefore, this study used the clinical database and the Kaplan–Meier (K-M) survival curve to identify biomarkers highly related to CRC, hoping to be used in the diagnosis and prognosis of CRC.

2. Methods

2.1. Ethical approval

The data selected in this study are given to the online database, so ethical approval and patient consent are not required.

2.2. Data set selection and VENN drawing

The GEO database is a high-throughput gene expression data repository for public functional genomics.[6] We screened 3 gene expression datasets from the GEO database,[7–9] including GSE89076, GSE73360, and GSE32323. The GSE89076 dataset contains 40 CRC tissue samples and 40 standard tissue samples; GSE73360 includes 55 and 22 non-tumor samples; and the GSE32323 dataset contains 22 CRC samples and 22 non-tumor samples.

Grouping by GEO2R in the GEO database, divided into cancer and normal tissue groups and analyzed and saved the analysis data in an Excel (Microsoft Corporation) table. The required differential genes were screened according to an adjusted P value of < .01 and an absolute value of logFC > 1.

Through the VENN diagram creation website, the retrieved differential genes that met the criteria were imported into the website. A VENN diagram was drawn to identify the differential genes common to the 3 groups.

2.3. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis of differential genes

DAVID Functional Annotation Bioinformatics Microarray Analysis (version 6.8) (http://david.ncifcrf.gov) can provide gene and protein annotation information.[10] The KEGG is a database for genome deciphering that includes complete and partially sequenced genome sequences with available information on metabolism, signaling, cell cycle, and membrane transport.[11] GO is a primary bioinformatics tool.[12] Gene function is divided into 3 parts: cellular components, molecular function, and biological processes (BPs). Using the GO database, we obtained annotation information of the selected target genes in cellular component, molecular function, and BP. Bioinformatics Microarray Analysis was performed using the DAVID online database to analyze the functions of DEGs. Statistical significance was set at $P < .05$.

2.4. Protein-protein interaction (PPI) network construction, module analysis, and Hub genes analysis

The PPI network retrieved differential gene interactions by STRING (version.11.5) (STRING: Functional Protein Association Networks) (https://string-db.org/). The STRING online database allows analysis of protein interactions through module analysis and network construction, and STRING can also provide possible correlations for disease development.[13] This helps us to explore and study the mechanism of the event and the action of the conditions. Cytoscape (version 3.9.0) (https://cytoscape.org) is a bioinformatics software platform,[14] and the Cytoscape plugin molecular complex detection (MCODE) can be used to cluster a given network based on the topology to find densely connected regions.[15] The PPI network was constructed using Cytoscape, and the most critical modules in the PPI network were identified using MCODE (Fig. 1B). The selection criteria were as follows: MCODE score > 10, degree cutoff = 2, node score cutoff = 0.2, maximum depth = 100, and k score = 2. The genes in this module were subjected to KEGG and GO analysis using DAVID.

The MCODE plugin selected the core genes in Cytoscape, and the selected hub genes had a degree of ≥10. Gene networks and co-expressed genes were analyzed using the Pathway Commons Networks Visualizer. Through the Biological Networks Gene Oncology plug-in in Cytoscape, essential genes' BPs and related functions were visualized and analyzed.[16] Hierarchical clustering of hub genes was performed using UCSC Xena (UCSC: University of California, Santa Cruz, https://xenabrowser.net/).

2.5. Analysis of survival rate and expression difference of Hub genes

Gene taxonomy was constructed using UCSC Xena. Overall survival and disease-free survival analysis of central genes were performed using K-M curves in cBioPortal (http://www.cbioportal.org/).

Based on the Gene Expression Profiling Interactive Analysis online database, the differences in gene expression of related genes between CRC samples and standard samples were determined. The mRNA expression levels of genes related to tumor grade were analyzed.[17] The screening criteria were |Log2FC| Cutoff: 1, P value cutoff: 0.01, Jitter Size: 0.4, Match TCGA standard, and GTEx data. The results are shown as box and violin plots.

3. Results

3.1. Identification of DEGs

We screened 3 datasets related to colon cancer from the GEO database, identified and screened the common DEGs in 3 datasets after unified processing of the relevant data in the dataset, and determined GSE89076 after normalization of the dataset with 7398 DEGs in GSE73360, 8098 DEGs in GSE73360, and 3213 DEGs in GSE32323. The overlap between the 3 datasets contained 329 genes, as shown in the VENN plot (Fig. 1A).

3.2. Functional enrichment analysis of DEGs

Functional and pathway enrichment analyses were performed using DAVID to analyze the biological classification of DEGs. GO analysis showed that the upregulated DEGs were mainly enriched in carbonate dehydratase activity, and oxidoreductase activity, inorganic anion transmembrane transporter activity, oxidoreductase activity, acting on the CH-OH group of donors, while the downregulated DEGs were mainly enriched in RNA binding, poly(A) RNA binding, and ATP binding (Table 1). KEGG pathway analysis revealed that upregulated DEGs were primarily enriched in nitrogen metabolism, proximal tubule bicarbonate reclamation, and retinol metabolism, while the downregulated DEGs were mainly enriched in cell cycle, DNA replication, ribosome biogenesis in eukaryotes, and RNA transport (Table 1).

3.3. PPI network analysis

The PPI network of DEGs contained 329 differential genes, including 264 upregulated and 65 downregulated genes (Fig. 1C). The most significant module from the PPI network contained 47 nodes and 989 edges (Fig. 1B). Functional analysis of the most significant module using DAVID showed that the genes in this module were mainly enriched in ATP binding, adenyl ribonucleotide binding, adenyl nucleotide binding, chromatin binding, cell cycle, DNA replication, progesterone-mediated oocyte maturation, and oocyte meiosis (Table 2).

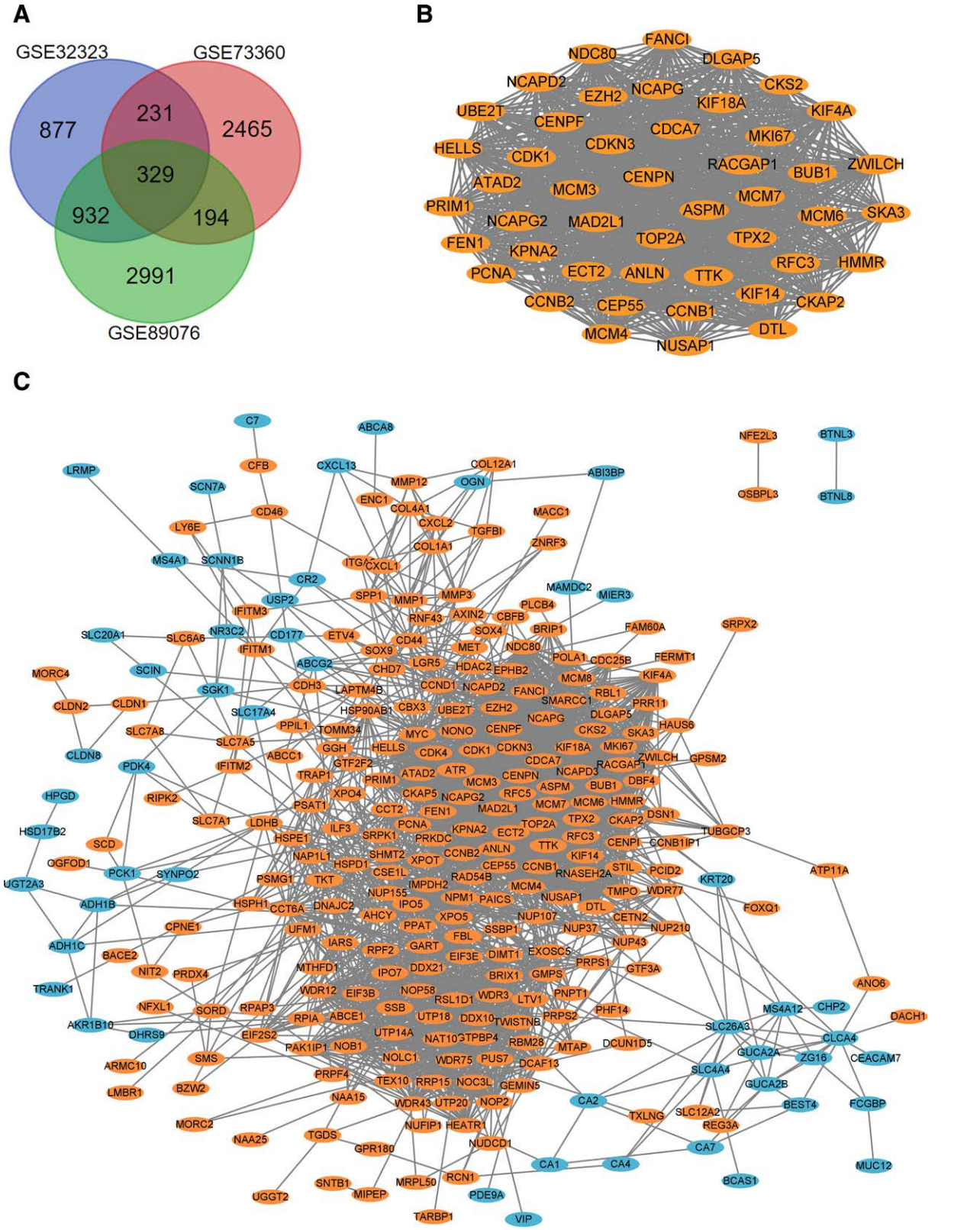

Figure 1. Venn diagram, PPI network of DEGs, and the most significant module. (A) DEGs were selected with a fold change >2 and *P* value <.01 among the mRNA expression profiling sets GSE32323, GSE73360 and GSE89076. The 3 datasets showed an overlap of 329 genes. (B) The most significant module. (C) The PPI network of DEGs was constructed using Cytoscape. Upregulated genes are marked in light red; downregulated genes are kept in light blue. DEGs = differentially expressed genes, PPI = protein-protein interaction.

Table 1

GO and KEGG pathway enrichment analysis of DEGs in CRC samples.

Term	Description	Count in gene set	*P* value
Downregulated			
GO:0003723	RNA binding	65	3.01E-13
GO:0044822	Poly(A) RNA binding	52	4.13E-12
GO:0005524	ATP binding	52	1.12E-08
GO:0030554	Adenyl nucleotide binding	53	1.14E-08
GO:0032559	Adenyl ribonucleotide binding	52	2.45E-08
GO:0003676	Nucleic acid binding	97	3.06E-07
GO:0000166	Nucleotide binding	65	9.71E-07
GO:1901265	Nucleoside phosphate binding	65	9.86E-07
GO:0036094	Small molecule binding	68	1.33E-06
GO:1901363	Heterocyclic compound binding	125	1.47E-06
GO:0097159	Organic cyclic compound binding	126	1.73E-06
GO:0001882	Nucleoside binding	54	1.85E-06
GO:0035639	Purine ribonucleoside triphosphate binding	53	3.09E-06
GO:0017076	Purine nucleotide binding	54	3.34E-06
GO:0032550	Purine ribonucleoside binding	53	3.57E-06
GO:0001883	Purine nucleoside binding	53	3.74E-06
GO:0032549	Ribonucleoside binding	53	3.74E-06
GO:0032555	Purine ribonucleotide binding	53	6.05E-06
GO:0003682	Chromatin binding	23	6.44E-06
GO:0032553	Ribonucleotide binding	53	7.72E-06
GO:0004386	Helicase activity	12	2.58E-05
GO:0097367	Carbohydrate derivative binding	57	4.55E-05
GO:0008536	Ran GTPase binding	6	7.30E-05
GO:0008094	DNA-dependent ATPase activity	8	2.00E-04
GO:0044877	Macromolecular complex binding	37	2.18E-04
GO:0016887	ATPase activity	18	3.09E-04
hsa04110	Cell cycle	20	4.95E-13
hsa03030	DNA replication	12	1.87E-11
hsa03008	Ribosome biogenesis in eukaryotes	12	3.99E-07
hsa03013	RNA transport	11	0.001092
hsa00230	Purine metabolism	11	0.001301
Upregulated			
GO:0004089	Carbonate dehydratase activity	4	1.13E-05
GO:0016616	Oxidoreductase activity	6	3.66E-05
GO:0015103	Inorganic anion transmembrane transporter activity	6	6.51E-05
GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	6	7.77E-05
GO:0004022	Alcohol dehydrogenase (NAD) activity	3	2.83E-04
GO:0008509	Anion transmembrane transporter activity	6	4.96E-04
GO:0016836	Hydro-lyase activity	4	6.27E-04
GO:0015081	Sodium ion transmembrane transporter activity	5	9.87E-04
GO:0022804	Active transmembrane transporter activity	7	0.001124
GO:0016835	Carbon-oxygen lyase activity	4	0.001683
hsa00910	Nitrogen metabolism	4	3.95E-05
hsa04964	Proximal tubule bicarbonate reclamation	4	1.01E-04
hsa00830	Retinol metabolism	4	0.002131

ATP = adenosine triphosphate, CRC = colorectal cancer, DEGs = differentially expressed genes, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.

3.4. Hub gene selection and analysis

We identified 20 genes with degrees ≥ 10 as hub genes. The gene symbols, full names, and functions of these hub genes are shown in Table 3. We analyzed the network of hub genes and their co-expressed genes using the cBioPortal online platform. The BPs of hub genes are shown in Figure 2B. The BPs of hub genes are mainly in the cell cycle, mitotic cell cycle, M phase of the mitotic cell cycle, nuclear division, and DNA replication. We constructed hierarchical clusters of hub genes using UCSC and distinguished hub genes into colon cancer samples and normal tissues (Fig. 2C). We analyzed the hub genes' overall survival and disease-free survival using K-M curves.

We found that CRC patients with alterations in ANLN, cyclin B1 (CCNB1), cyclin-dependent kinase (CDK1), CDKN3, CEP55, FEN1, HELLS, HMMR, KIF4A, minichromosome maintenance complex component 7 (MCM7), and TOP2A had poor overall survival (Fig. 3A). Alterations in CCNB1, CDK1, CEP55, HMMR, KIF14, NCAPD2, and RFC3 resulted in worse disease-free survival (Fig. 3B). In addition, among these key genes, CDK1, PCNA, and CCNB1 showed a higher degree of note; therefore, we speculate that these 3 genes may play a vital role in the occurrence and development of CRC patients. We noted a decrease in overall survival and disease-free survival in patients with CRC associated with genomic alterations in CDK1. The results were statistically significant (overall survival, $P = .0492$; disease-free survival, $P < .001$). The overall survival of patients with CRC associated with CCNB1 alterations was not statistically significant, but disease-free survival was statistically significant (overall survival, $P = .0505$; disease-free survival, $P < .001$). Therefore, changes in CCNB1 levels were significantly associated with poor disease-free survival. On the other hand, the change in PCNA was not significantly correlated with the overall survival and disease-free survival of CRC patients, and the observed results were not statistically significant (overall survival $P = .674$, disease-free survival $P = .424$).

Table 2

GO and KEGG pathway enrichment analysis of DEGs in the most significant module.

Pathway ID	Pathway description	Count in gene set	FDR
GO:0005524	ATP binding	16	1.37E-04
GO:0032559	Adenyl ribonucleotide binding	16	1.37E-04
GO:0030554	Adenyl nucleotide binding	16	1.37E-04
GO:0003682	Chromatin binding	10	1.59E-04
GO:0016887	ATPase activity	9	3.16E-04
GO:0017111	Nucleoside-triphosphatase activity	11	3.16E-04
GO:0035639	Purine ribonucleoside triphosphate binding	16	3.16E-04
GO:0032550	Purine ribonucleoside binding	16	3.16E-04
GO:0001883	Purine nucleoside binding	16	3.16E-04
GO:0032549	Ribonucleoside binding	16	3.16E-04
GO:0001882	Nucleoside binding	16	3.16E-04
GO:0032555	Purine ribonucleotide binding	16	3.16E-04
GO:0016462	Pyrophosphatase activity	11	3.16E-04
GO:0016818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	11	3.16E-04
GO:0016817	Hydrolase activity, acting on acid anhydrides	11	3.16E-04
GO:0017076	Purine nucleotide binding	16	3.16E-04
GO:0032553	Ribonucleotide binding	16	3.16E-04
GO:0097367	Carbohydrate derivative binding	17	4.90E-04
GO:0008094	DNA-dependent ATPase activity	5	4.90E-04
hsa04110	Cell cycle	11	4.45E-12
hsa03030	DNA replication	8	3.07E-11
hsa04914	Progesterone-mediated oocyte maturation	5	5.87E-04
hsa04114	Oocyte meiosis	5	0.001135553

ATP = adenosine triphosphate, DEGs = differentially expressed genes, FDR = false discovery rate, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.

3.5. Changes in the expression of relevant hub genes in different clinical states

We analyzed the expression of these 3 genes in tumor and normal tissues (Fig. 4) and the changes in tumor grade (Fig. 5) using the Gene Expression Profiling Interactive Analysis online database. The expression levels of CDK1, PCNA, and CCNB1 in the tumor tissues were significantly higher than those in the normal tissues. The changes in CCNB1 expression in different grades of colon cancer tissues were statistically significant (P = .000223). Changes in CDK1 expression in colon cancer tissues of various grades were also statistically significant (P = .0218). The changes in PCNA expression in different grades of colon cancer tissues were statistically insignificant (P = .0618).

4. Discussion

CRC is a common malignancy worldwide, and its mortality rate has increased in recent years. However, the molecular mechanisms underlying CRC development remain poorly understood, and CRC has an insidious onset and is difficult to detect early, which may be one of the reasons for its poor prognosis. Therefore, there is an urgent need to identify potential biomarkers for efficient diagnosis and treatment. In addition, screening methods for colon cancer tumor markers have been introduced one after another, among which, new analytical methods for fecal proteins and related DNA markers based on fecal occult blood tests have been developed with good results.[18] Microarray technology has enabled the exploration of genetic alterations in CRC, and the combined application of microarray technology and clinical databases is a critical approach to analyzing gene expression and potential biomarkers, improving our understanding of the underlying molecular mechanisms of complex diseases, and has proven to be a helpful method for identifying novel biomarkers in other diseases.

In this study, we analyzed 3 colon cancer mRNA microarray datasets, resulting in the identification of 329 DEGs, which included 264 upregulated genes and 65 downregulated genes, mainly enriched in oxidoreductase activity, nitrogen metabolism, bicarbonate recycling, retinol metabolism, and downregulated genes mainly enriched in ATPase activity, cell cycle, and DNA replication. Additionally, studies have shown that dysregulation of mitosis plays a crucial role in the occurrence and development of tumors. Mitotic dysregulation and the cell cycle play an essential role in cancer progression. Kohoutova et al also demonstrated that apoptosis, cell cycle, and mitosis were significantly dysregulated during CRC tumor progression,[19] which is also in line with the findings of this study.

We selected 20 DEGs with a degree ≥ 10 as hub genes. Among these hub genes, CDK1, CCNB1, and PCNA had the highest expression levels. As a member of the family of cyclin-dependent kinases (CDKs), cyclin-dependent kinase 1 (CDK1) is a key driver of cell-cycle transition.[20] Numerous studies have shown that CDK1 disorder can accelerate tumor growth and induce spontaneous proliferation of cancer cells, especially in colorectal, lung, and liver cancers. Zhu et al also demonstrated that high expression of CDK1 may lead to poor prognosis in CRC patients. In addition, inhibition of CDK1 expression may improve the sensitivity of pentafluorouracil in the treatment of CRC.[21] Because cell proliferation has very high requirements for protein synthesis, Haneke et al found that CDK1 is required for translating 59TOP mRNA, which is closely related to the phosphorylation of 59TOP mRNA-binding protein.[22] In other tumor tissues, such as hepatocellular carcinoma, the expression of CDK1, CCNB1, and CCNB2 is positively correlated with the infiltration of $CD4^+$T cells, $CD8^+$T cells, neutrophils, macrophages, and dendritic cells. These genes are also significantly associated with the expression of various immune markers in hepatocellular carcinoma.[23,24]

We evaluated the relationship of CDK1 expression with overall survival and disease-free survival using the cBioportal online database and found that CDK1 alterations were significantly associated with worsening overall survival and disease-free survival. We found that the expression of CDK1 was considerably higher in colon cancer tumor tissues than in normal tissues. It has also been confirmed that high expression of CDK1 is associated with the development of colon cancer.[25] In addition to CRC, some studies have demonstrated that the expression of CDK1 in lung adenocarcinoma, squamous cell lung cancer, and gangrenous carcinoma is higher than that in normal lung tissue.

Table 3

Functional roles of 26 hub genes with degree ≥10.

Gene symbol	Full name	Function
CDK1	Cyclin dependent kinase 1	This protein is a catalytic subunit of the highly conserved protein kinase complex known as M-phase promoting factor (MPF), which is essential for G1/S and G2/M phase transitions of eukaryotic cell cycle.
PCNA	Proliferating cell nuclear antigen	The protein encoded by this gene is found in the nucleus and is a cofactor of DNA polymerase delta. The encoded protein acts as a homotrimer and helps increase the processivity of leading strand synthesis during DNA replication. In response to DNA damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway.
CCNB1	Cyclin B1	The protein encoded by this gene is a regulatory protein involved in mitosis. The gene product complexes with p34(cdc2) to form the maturation-promoting factor (MPF).
MAD2L1	Mitotic arrest deficient 2 like 1	MAD2L1 is a component of the mitotic spindle assembly checkpoint that prevents the onset of anaphase until all chromosomes are properly aligned at the metaphase plate.
MCM7	Minichromosome maintenance complex component 7	The protein encoded by this gene is one of the highly conserved mini-chromosome maintenance proteins (MCM) that are essential for the initiation of eukaryotic genome replication. The hexameric protein complex formed by the MCM proteins is a key component of the pre-replication complex (pre_RC) and may be involved in the formation of replication forks and in the recruitment of other DNA replication related proteins.
MCM3	Minichromosome maintenance complex component 3	The acetylation of this protein inhibits the initiation of DNA replication and cell cycle progression. Several transcript variants encoding different isoforms have been found for this gene.
MCM4	Minichromosome maintenance complex component 4	The protein encoded by this gene is one of the highly conserved mini-chromosome maintenance proteins (MCM) that are essential for the initiation of eukaryotic genome replication. The hexameric protein complex formed by MCM proteins is a key component of the pre-replication complex (pre_RC) and may be involved in the formation of replication forks and in the recruitment of other DNA replication related proteins.
AURKB	Aurora kinase B	This gene encodes a member of the aurora kinase subfamily of serine/threonine kinases. These kinases participate in the regulation of segregation of chromosomes during mitosis and meiosis through association with microtubules.
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	This gene encodes a serine/threonine-protein kinase that play a central role in mitosis. The encoded protein functions in part by phosphorylating members of the mitotic checkpoint complex and activating the spindle checkpoint. This protein also plays a role in inhibiting the activation of the anaphase promoting complex/cyclosome. This protein may also function in the DNA damage response.
CCNB2	Cyclin B2	Cyclin B2 is a member of the cyclin family, specifically the B-type cyclins. The B-type cyclins, B1 and B2, associate with p34cdc2 and are essential components of the cell cycle regulatory machinery. B1 and B2 differ in their subcellular localization.
NDC80	NDC80 kinetochore complex component	This gene encodes a component of the NDC80 kinetochore complex. The encoded protein functions to organize and stabilize microtubule-kinetochore interactions and is required for proper chromosome segregation.
KIF18A	Kinesin family member 18A	KIF18A is a member of the kinesin superfamily of microtubule-associated molecular motors (see MIM 148760) that use hydrolysis of ATP to produce force and movement along microtubules.
FEN1	Flap structure-specific endonuclease 1	The protein encoded by this gene removes 5' overhanging flaps in DNA repair and processes the 5' ends of Okazaki fragments in lagging strand DNA synthesis.
UBC	Ubiquitin C	This gene represents a ubiquitin gene, ubiquitin C. The encoded protein is a polyubiquitin precursor. Ubiquitination has been associated with protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways.
TOP2A	DNA topoisomerase II alpha	This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This nuclear enzyme is involved in processes such as chromosome condensation, chromatid separation, and the relief of torsional stress that occurs during DNA transcription and replication.
PRIM1	DNA primase subunit 1	It is a heterodimer of a small subunit and a large subunit, synthesizes small RNA primers for the Okazaki fragments made during discontinuous DNA replication. The protein encoded by this gene is the small, 49 kDa primase subunit.
UBE2T	Ubiquitin conjugating enzyme E2 T	The protein encoded by this gene catalyzes the covalent attachment of ubiquitin to protein substrates. Defects in this gene have been associated with Fanconi anemia of complementation group T. Two transcript variants encoding different isoforms have been found for this gene.
CDK2	Cyclin dependent kinase 2	This gene encodes a member of a family of serine/threonine protein kinases that participate in cell cycle regulation. The encoded protein is the catalytic subunit of the cyclin-dependent protein kinase complex, which regulates progression through the cell cycle.
PLK1	Polo like kinase 1	It is highly expressed during mitosis and elevated levels are found in many different types of cancer. Depletion of this protein in cancer cells dramatically inhibited cell proliferation and induced apoptosis; hence, it is a target for cancer therapy.
CENPN	Centromere protein N	It is bound to kinetochores during S phase and G2 and recruits other proteins to the centromere. Pseudogenes of this gene are located on chromosome 2.

CCNB1 = cyclin B1, CDK1 = cyclin-dependent kinase, MAD2L1 = mitotic arrest deficient 2 like 1, MCM7 = minichromosome maintenance complex component 7.

High expression of CDK1 is also closely related to poor overall survival in patients with lung cancer.[26]

In the present study, CCNB1 was closely associated with CDK1 and had a high degree in hub genes. CCNB1 is a regulatory protein involved in mitosis. Accumulating evidence indicates that CCNB1 plays a crucial role in regulating and forming the CDK1 complex and in promoting cell cycle progression from the G2 phase to mitosis. Fang et al also demonstrated that overexpression of CCNB1 promoted cell proliferation and tumor growth in human CRC.[27] CCNB1 is highly expressed in various human cancers. CCNB1 has also been confirmed as a key gene in the development of cervical cancer. Downregulation

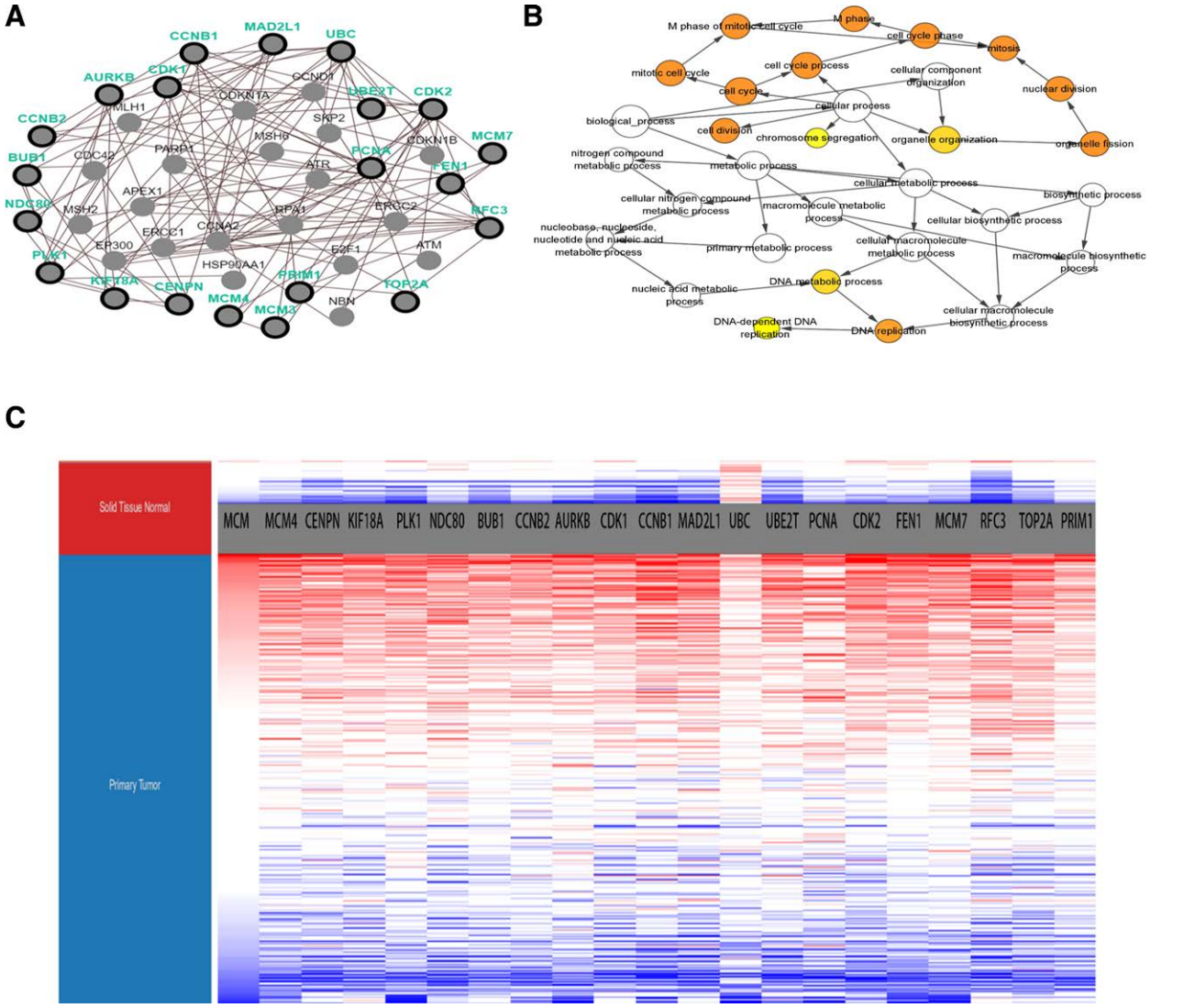

Figure 2. Interaction network and biological process analysis of the hub genes. (A) Hub genes and their co-expression genes were analyzed using Pathway Commons Networks Visualizer (PCViz). Nodes with bold black outlines represent hub genes. Nodes with gray represent the co-expression genes. (B) The biological process analysis of hub genes was constructed using BiNGO. The color depth of nodes refers to the corrected *P* value of ontologies. The size of nodes refers to the number of genes involved in the ontologies. $P < .0001$ was considered statistically significant. (C) The hierarchical clustering of hub genes was constructed using UCSC Xena. The samples under the red bar are non-cancerous, and the samples under the blue bar are CRC samples. The upregulation of genes is marked in red; the downregulation of genes is marked in blue; null is marked in gray. BiNGO = Biological Networks Gene Oncology, CRC = colorectal cancer.

of CCNB1 activates the P53 signaling pathway in cells, which is closely related to cancer development.[28] Alfonso-Pérez et al found that CDK1-CCNB1 is an important component of the spindle that ensures mitotic fidelity.[29] Therefore, alterations in CDK1 and CCNB1 affect mitosis and the cell cycle.

This study also found that PCNA had highly expressed in tumor tissues and markers for identification. PCNA plays important roles in translational synthesis, mismatch repair, and chromatin assembly.[30] In addition, PCNA can interact with DNA polymerase, thereby promoting efficient DNA.[31] It was also noted that PCNA expression was elevated in patients with colon cancer, and its expression level was significantly correlated with tumor infiltration and tumor-node-metastasis stage.[32] In the diagnosis of colon cancer, Cai et al found that the expression level of PCNA was mainly low in stages 0 and I, but high in stages II, III, and IV. These results suggest that PCNA has potential value in diagnosing early CRC.[32] Our analysis of PCNA gene expression in normal and tumor tissues also showed that the expression of PCNA in tumor tissues was higher than that in normal tissues. PCNA had no statistical significance in overall and disease-free survival, and the specific reasons need to be further studied.

In addition, the dysregulation of MCM7 is significantly correlated with tumorigenesis, distant metastasis, recurrence of CRC, and poor clinical prognosis. Studies have also shown that the expression of MCM7 in hepatocellular carcinoma is related to clinicopathological characteristics and patient survival. It can predict patient prognosis after surgical liver cancer resection.[33] Mitotic arrest deficient 2 like 1 (MAD2L1) is a mitotic spindle assembly checkpoint component affecting the correct arrangement of all chromosomes. Another study found that MAD2L1 is highly expressed in gastric cancer, and miR-30a-3p can downregulate the expression level of MAD2L1, thereby inhibiting the proliferation of gastric cancer cells and affecting the cell cycle.[34] CCNB2 is overexpressed in the bladder, lung, and CRCs and is associated with tumor invasion, metastasis, and poor prognosis.[35,36] In addition, we performed hierarchical clustering of the key genes. The results showed that these key

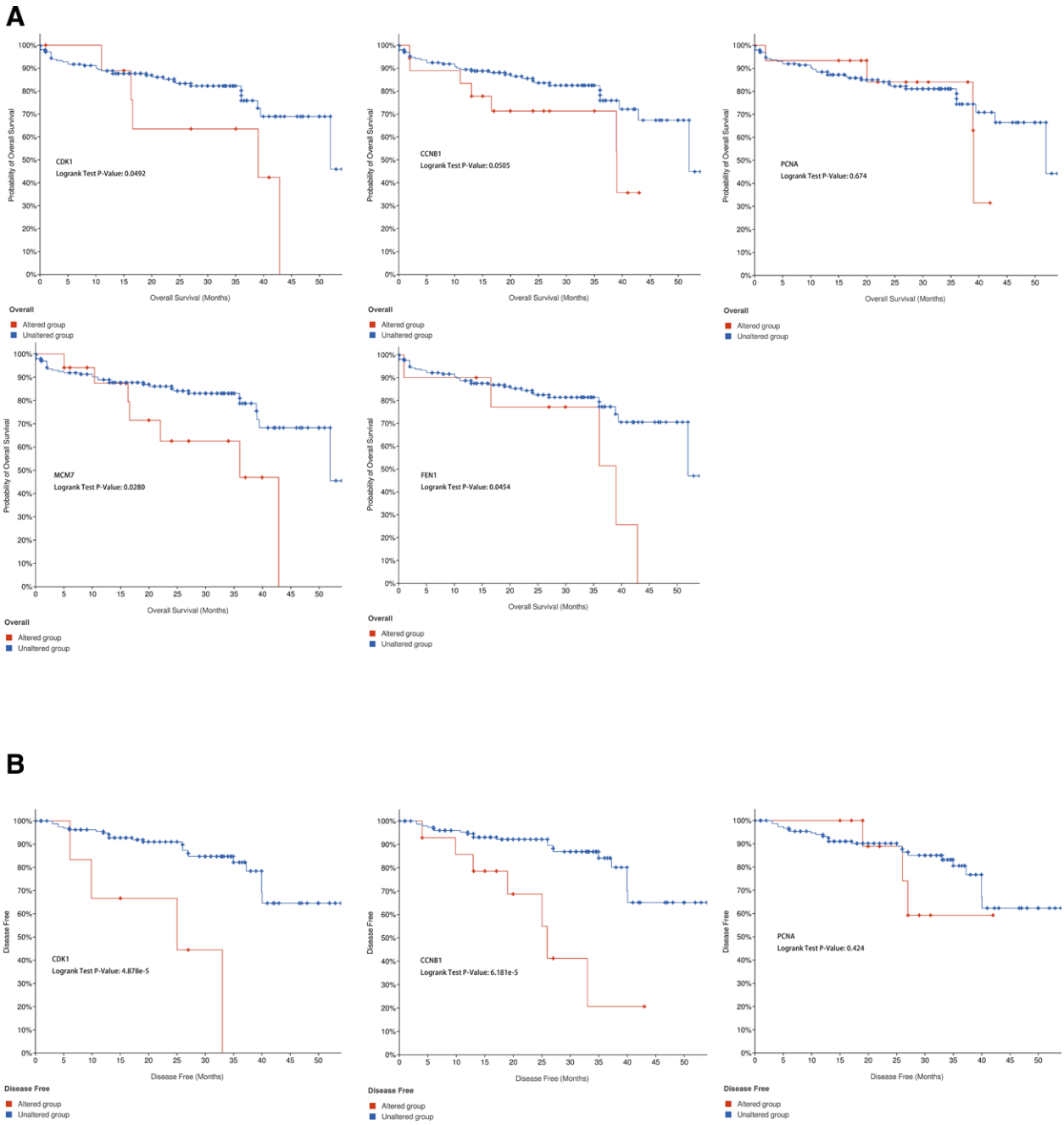

Figure 3. (A) Overall survival and (B) disease-free survival analyses of hub genes were performed using the cBioPortal online platform. $P < .05$ was considered statistically significant. Red cases represent the cases with alterations in query genes. Blue cases represent the cases without alterations in query genes.

genes were expressed at low levels in normal tissue samples and were overexpressed in CRC samples. This clearly distinguishes CRC samples from normal tissue samples and may be a candidate diagnostic biomarker. In addition, alterations in MCM7 and FEN1 were associated with poor overall survival, suggesting that these genes may play an essential role in the carcinogenesis, progression, invasion, and recurrence of CRC.

In conclusion, our study shows that CDK1, CCNB1, and PCNA can be used as potential markers for the diagnosis and prognosis of CRC. These genes are overexpressed in colon cancer tissues and are associated with low survival rates in CRC patients. In addition, more studies are needed to further understand the experimental molecular mechanisms of CDK1, CCNB1, and PCNA.

Author contributions

Conceptualization: Chongyang Li, Mingxiao Guo.
Data curation: Chongyang Li.
Formal analysis: Chongyang Li.
Funding acquisition: Mingxiao Guo.
Investigation: Chongyang Li.
Methodology: Ying Gao.
Resources: Mingxiao Guo.
Software: Mingxiao Guo.
Supervision: Chunlei Lu, Mingxiao Guo.
Validation: Mingxiao Guo.
Visualization: Mingxiao Guo.
Writing – original draft: Chongyang Li, Mingxiao Guo.
Writing – review & editing: Chongyang Li, Mingxiao Guo.

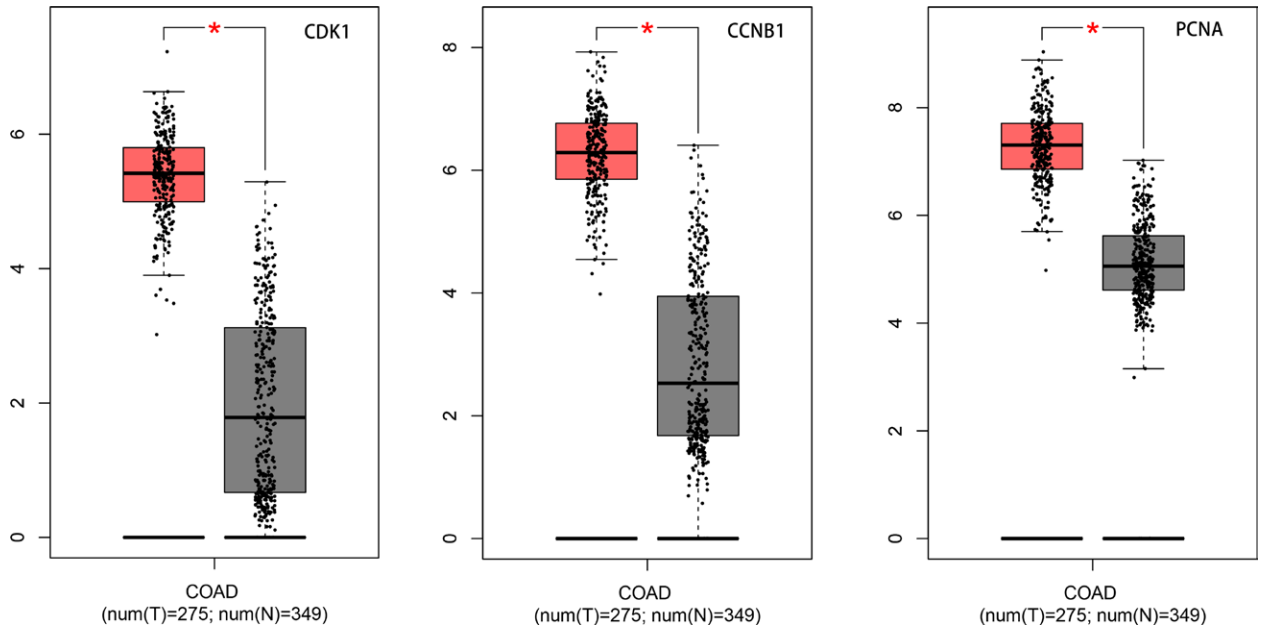

Figure 4. mRNA expression level of CDK1, CCNB1, and PCNA based on TCGA data by GEPIA. The expression levels of CDK1, CCNB1, and PCNA between COAD and the normal samples were consistent with the results of GEO. CCNB1 = cyclin B1, CDK1 = cyclin-dependent kinase, COAD = colon adenocarcinoma, GEO = Gene Expression Omnibus, GEPIA = Gene Expression Profiling Interactive Analysis, TCGA = The Cancer Genome Atlas.

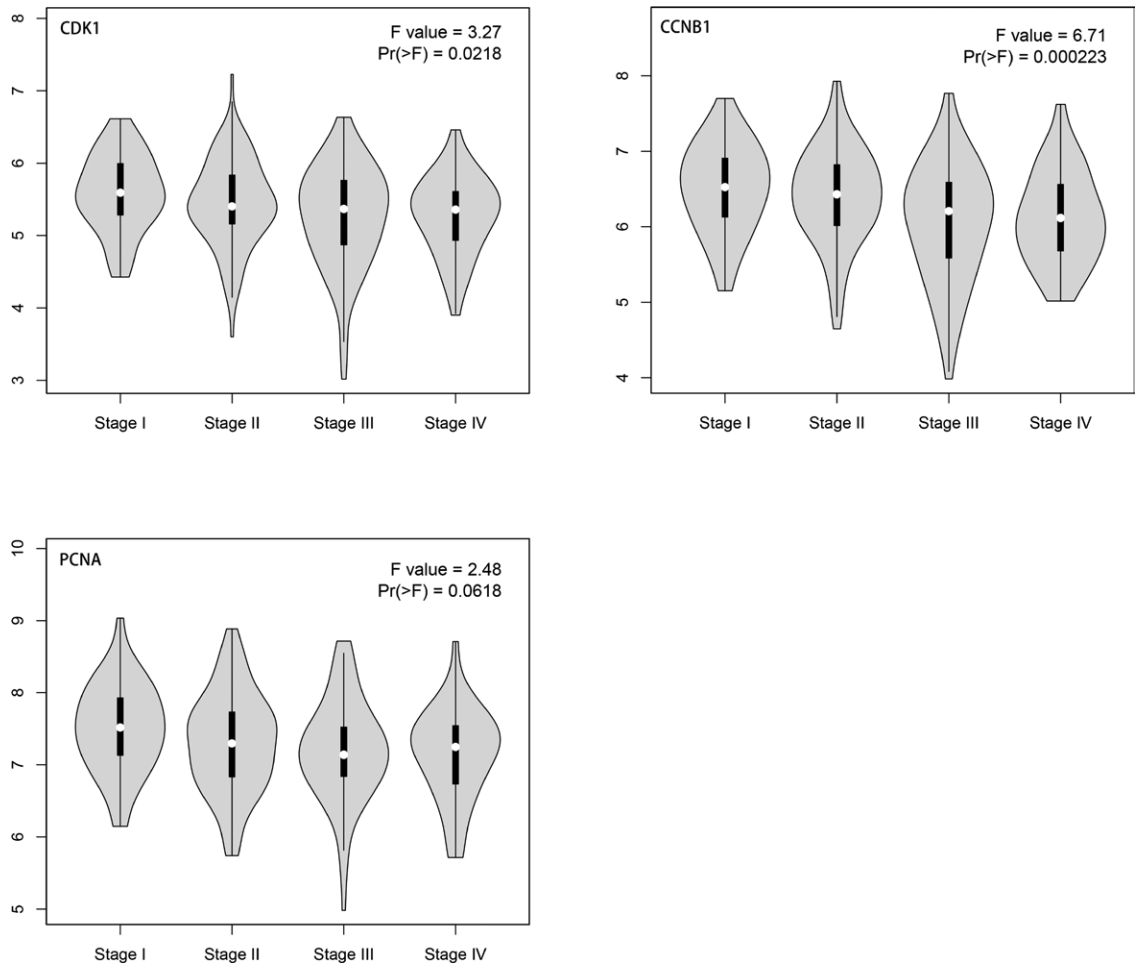

Figure 5. The mRNA expression level of CDK1, CCNB1, and PCNA in tumor grade based on TCGA data by GEPIA. CCNB1 = cyclin B1, CDK1 = cyclin-dependent kinase, GEPIA = Gene Expression Profiling Interactive Analysis, TCGA = The Cancer Genome Atlas.

References

[1] Baidoun F, Elshiwy K, Elkeraie Y, et al. Colorectal cancer epidemiology: recent trends and impact on outcomes. Curr Drug Targets. 2021;22:998–1009.

[2] Jing C, Jin YH, You Z, et al. Prognostic value of amphiregulin and epiregulin mRNA expression in metastatic colorectal cancer patients. Oncotarget. 2016;7:55890–9.

[3] Zhang H, Dong S, Feng J. Epigenetic profiling and mRNA expression reveal candidate genes as biomarkers for colorectal cancer. J Cell Biochem. 2019;120:10767–76.

[4] Chiavarina B, Costanza B, Ronca R, et al. Metastatic colorectal cancer cells maintain the TGFβ program and use TGFBI to fuel angiogenesis. Theranostics. 2021;11:1626–40.

[5] Li L, Lei Q, Zhang S, et al. Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatics analysis. Oncol Rep. 2017;38:2607–18.

[6] Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.

[7] Satoh K, Yachida S, Sugimoto M, et al. Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC. Proc Natl Acad Sci USA. 2017;114:E7697–706.

[8] Condorelli DF, Spampinato G, Valenti G, et al. Positive caricature transcriptomic effects associated with broad genomic aberrations in colorectal cancer. Sci Rep. 2018;8:14826.

[9] Khamas A, Ishikawa T, Shimokawa K, et al. Screening for epigenetically masked genes in colorectal cancer using 5-Aza-2'-deoxycytidine, microarray and gene expression profile. Cancer Genomics Proteomics. 2012;9:67–75.

[10] Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8:R183–R183.

[11] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.

[12] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

[13] Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2012;41:D808–15.

[14] Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011;27:431–2.

[15] Bandettini WP, Kellman P, Mancini C, et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson. 2012;14:83.

[16] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signaling. 2013;6:pl1.

[17] Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45:W98–W102.

[18] Koga Y, Yamazaki N, Matsumura Y. New molecular diagnosis and screening methods for colorectal cancer using fecal protein, DNA and RNA. Expert Rev Mol Diagn. 2014;14:107–20.

[19] Kohoutova D, Pejchal J, Bures J. Mitotic and apoptotic activity in colorectal neoplasia. BMC Gastroenterol. 2018;18:65–65.

[20] Li J, Wang Y, Wang X, et al. CDK1 and CDC20 overexpression in patients with colorectal cancer are associated with poor prognosis: evidence from integrated bioinformatics analysis. World J Surg Oncol. 2020;18:50.

[21] Zhu Y, Li K, Zhang J, et al. Inhibition of CDK1 reverses the resistance of 5-Fu in colorectal cancer. Cancer Manag Res. 2020;12:11271–83.

[22] Haneke K, Schott J, Lindner D, et al. CDK1 couples proliferation with protein synthesis. J Cell Biol. 2020;219:e201906147.

[23] Zou Y, Ruan S, Jin L, et al. CDK1, CCNB1, and CCNB2 are prognostic biomarkers and correlated with immune infiltration in hepatocellular carcinoma. Med Sci Monit. 2020;26:e925289.

[24] Izadi S, Nikkhoo A, Hojjat-Farsangi M, et al. CDK1 in breast cancer: implications for theranostic potential. Anticancer Agents Med Chem. 2020;20:758–67.

[25] Tong Y, Huang Y, Zhang Y, et al. DPP3/CDK1 contributes to the progression of colorectal cancer by regulating cell proliferation, cell apoptosis, and cell migration. Cell Death Dis. 2021;12:529.

[26] Li M, He F, Zhang Z, et al. CDK1 is a potential prognostic biomarker and target for lung cancer. J Int Med Res. 2020;48:030006051989750.

[27] Fang Y, Yu H, Liang X, et al. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol Ther. 2014;15:1268–79.

[28] Zhang H, Zhang X, Li X, et al. Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer. J Cell Physiol. 2018;234:619–31.

[29] Alfonso-Perez T, Hayward D, Holder J, et al. MAD1-dependent recruitment of CDK1-CCNB1 to kinetochores promotes spindle checkpoint signaling. J Cell Biol. 2019;218:1108–17.

[30] Boehm EM, Gildenberg MS, Washington MT. The many roles of PCNA in eukaryotic DNA replication. Enzymes. 2016;39:231–54.

[31] Acharya N, Patel SK, Sahu SR, et al. "PIPs" in DNA polymerase: PCNA interaction affairs. Biochem Soc Trans. 2020;48:2811–22.

[32] Cai F, Li J, Pan X, et al. Increased expression of PCNA-AS1 in colorectal cancer and its clinical association. Clin Lab. 2017;63:1809–14.

[33] Zhou YM, Zhang XF, Cao L, et al. MCM7 expression predicts post-operative prognosis for hepatocellular carcinoma. Liver Int. 2012;32:1505–9.

[34] Wang Y, Wang F, He J, et al. miR-30a-3p targets MAD2L1 and regulates proliferation of gastric cancer cells. Onco Targets Ther. 2019;12:11313–24.

[35] Lei CY, Wang W, Zhu YT, et al. The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer. Urol Oncol. 2016;34:237.e1–237.e10.

[36] Duan R, Du W, Guo W. EZH2: a novel target for cancer treatment. J Hematol Oncol. 2020;13:104.