CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer

Objective: Colorectal cancer (CRC) is considered to be the most prevalent malignant tumors that contribute to high cancer-related mortality. However, the signaling pathways involved in CRC and CRC-driven genes are largely unknown. We seek to discover a novel biomarker in CRC.

Materials and methods: All clinical CRC samples (n = 33) were from Xiangya Hospital. We first selected CCNA2 by integrated bioinformatics analysis of four GSE databases. Next, the expression of CCNA2 in tissues and cell lines was verified by quantitative real-time PCR. The effects of CCNA2 on cell growth, proliferation, cell cycle, and apoptosis were examined by in vitro assays.

Results: We identified 498 shared DEGs (294 upregulated and 204 downregulated), and the top ten hub genes were selected by integrated analysis. These hub genes were significantly overexpressed in CRC samples and were positively correlated. Our data revealed that the expression of CCNA2 in CRC tissues is higher than that in normal tissues. The CCNA2 knockdown could significantly suppress CRC cell growth by impairing cell cycle progression and inducing cell apoptosis.

Conclusion: CCNA2, as a novel oncogenic gene, plays a role in regulating cancer cell growth and apoptosis. It could be used as a new biomarker for diagnosis and therapy in CRC.

Keywords: CCNA2, colorectal cancer, bioinformatics analysis, proliferation, cell cycle, apoptosis

Introduction
Colorectal cancer (CRC) is not only the most common malignant tumor in the digestive system but also the major cause of cancer-related death in the world.1 According to the annual report of Cancer Facts & Figures 2017, there are 0.14 million new cases of CRC patients in the US in 2017, accounting for 43.625% of gastrointestinal cancers.2 In addition, CRC caused ~50,000 deaths in 2017, accounting for 31.87% of death caused by gastrointestinal cancers.2 The Global Cancer Statistic showed that there are 1.36 million new CRC patients each year around the world. CRC malignant tumors have the third highest occurrence in the world, ranking third in males and second in females. Also, CRC has caused ~0.69 million deaths, and the mortality rates are still increasing in developed countries.3–6 It is worth noting that while the 5-year survival rates of CRC patients are still rising in some populations and countries, the morbidity and mortality rates remain very high due to cancer recurrence and metastasis.7–10 A large number of studies have reported that the screening and early detection of CRC are deemed to be crucial for improving survival rate.3,11 So far, the main clinical screening methods for CRC involve endoscopic screening, particularly colonoscopy.12–15 Colonoscopy comprises shortcomings such as the poor patient compliance, the influence of
family history, inconvenience, and high cost and risk. Therefore, it is emergent to find an effective early diagnostic strategy to detect relapse and monitor CRC progression.

Due to the high heterogeneity of CRC, researchers are exploring biomarkers that could indicate the specific cancer subtype and prognosis of CRC in recent decades. Until now, several biomarkers have been discovered in a wide range of cancers. Yes-associated protein (YAP), for example, is a newly found downstream effector molecule of Hippo, and the Hippo pathway can function in tumor suppression. Many studies have shown that YAP is an oncogenic gene, which is highly expressed in various tumors and can regulate tumor progression. Nuclear YAP can promote cancer cell proliferation, apoptosis, metastasis, and maintenance of stemness by interacting with other transcription factors and affecting tumor progression. Therefore, YAP has the prospect of being a new tumor biomarker or therapeutic target for CRC. During the past few decades, the high-throughput sequencing technology has been widely used in the life sciences. The widespread use of high-throughput sequencing has rendered a large amount of core slice data stored in public databases, which can provide valuable clues for novel researches after being integrated and re-analyzed. In this analysis, we chose GSE21815, GSE32323, GSE44076, and GSE74602 from Gene Expression Omnibus ([GEO], https://www.ncbi.nlm.nih.gov/geo), and used the GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) online tool to detect differentially expressed genes (DEGs). Gene ontology (GO) and pathway enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) is an online bioinformatics program that provides tools for functionally explaining biological processes and a large quantity of genes or proteins. GO analysis is a universally useful method for gene annotation and gene products. The KEGG is a collection of databases for associating related gene sets with the pathways they are involved in. The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) is an online bioinformatics program that provides tools for functionally explaining biological processes and a large quantity of genes or proteins. GO analysis is a universally useful method for gene annotation and gene products. The KEGG is a collection of databases for associating related gene sets with the pathways they are involved in.

Materials and methods

Data processing of DEGs

GEO2R was used to analyze DEGs between CRC tissues and noncancerous tissues. GEO2R is an interactive online tool which can compare several groups of samples in GEO sequences and analyze nearly any GEO series with a gene symbol. The Benjamin and Hochberg (BH) false discovery rate method of default was corrected for the occurrence of false-positive results according to adjusted P-values. Adjusted P-values <0.05 and |logFC| ≥1.5 were set as the cutoff standards. We extracted 8,843, 2,606, 2,678, and 1,674 DEGs from the expression profile data sets GSE21815, GSE32323, GSE44076, and GSE74602.

GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of DEGs

GO analysis is a universally useful method for gene annotation and gene products. The KEGG is a collection of databases for associating related gene sets with the pathways they are involved in. The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) is an online bioinformatics program that provides tools for functionally explaining biological processes and a large quantity of genes or proteins. P values <0.05 and P BH values <0.05 were set as the cutoff standards. We chose DAVID to analyze the core biological processes, cellular components, molecular functions, and pathways among these DEGs.

PPI network and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) was employed to identify DEG-encoded proteins and their PPI networks. We first mapped the DEGs into a protein–protein interaction (PPI) network for DEGs and selected ten core genes that have a high degree of connectivity. At the same time, the Molecular Complex Detection (MCODE) app in Cytoscape was used to analyze the PPI network modules. In addition, cutoff =2, node score cutoff =0.2, k-core =2, and max depth =100 was set as cutoff standards. The top three
modules were selected. KEGG pathway analysis of genes in different modules was performed by DAVID.

Comparison of the hub gene expression levels and their relationship
GEPIA is a recently developed interactive web server used to explore the large data sets from the The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) projects. It provides a common task, such as analysis of DEGs based on the comparison of tumor and normal samples. In this study, GEPIA was used to analyze the relative expression levels of hub genes in CRC tissues and nontumorous tissues. Boxplots were used to compare the gene expressions of hub genes in CRC tissues and nontumorous tissues. The correlation was used to visualize the relationships among these ten genes.

Validation based on clinical samples from CRC
To further verify the data from GEO, we conducted quantitative real-time PCR (qRT-PCR) to quantify the expression level of CCNA2 in clinical CRC patient samples (n=33) from Xiangya Hospital (Central South University, Changsha, China). Written informed consents were obtained from all patients. This study was approved by the Institute Research Ethics Committee of Xiangya Hospital.

Cell lines and cell transfection
All cell lines, including normal cell line NCM460, and CRC cell lines HT-29, HCT116, SW620, and SW480 were purchased from the American Type Culture Collection. The cells were cultured in RPMI-1640, supplemented with 10% FBS (Thermo Fisher Scientific, Waltham, MA, USA). All cells were maintained in a humidified incubator with 5% CO2 at 37°C. A total of 5 x 10^5 cells/mL were plated about 24 hours before transfection. Once cells were at 40%–60% confluence, in each well of a six-well plate, cells were transfected with 50 nM of siRNA/NC (RiboBio, Guangzhou, China) using Lipofectamin 2000 (Thermo Fisher Scientific) at indicated concentrations according to the manufacturer’s instructions. Six hours later, the culture medium was replaced with fresh medium containing 10% FBS. The cells were harvested after 24 hours of transfection for the following assays.

siRNA sequences are as follows
Si-h-CCNA2_001: forward, 5’-GGUGUGAAGUCAUCUAUGAU dTdT-3’; reverse, 3’-dTdT UAUCUAGUAGUUCACAGC-5’.

RNA extraction, reverse transcription (RT)-PCR, and qRT-PCR
Total RNAs from cells were extracted using TRIzol reagent (Thermo Fisher Scientific). Reverse-transcribed complementary DNA was synthesized using the GoScript Reverse Transcription System (Promega Corporation, Fishburg, WI, USA). The RT-PCR conditions were 42°C for 15 minutes, 70°C for 15 minutes, and 4°C hold. After the dilution of cDNA with Nuclease-free water by 1:4, qRT-PCR was performed by the Applied Biosystems 7500 Real-Time PCR System and the GoTaq qPCR Master Mix (Promega Corporation, A6001). The mixes were predenatured at 95°C for 10 minutes, followed by 40 cycles of denaturation at 95°C for 15 seconds, and 60°C for 1 minute. The results were normalized to GAPDH expression. The relative expression level of CCNA2 was calculated by the 2^−ΔΔCt method.

The primers used for qRT-PCR were as follows: CCNA2-forward, 5’-GGTACCTACGCGCTTTTCTC-3’; and reverse, 5’-GGGCATGCGCCGCTCTATTT-3’; GAPDH forward, 5’-GCACCGTCAAGGCTGAGAAC-3’; and reverse, 5’-TGGTGAAGACGCCAGTGGA-3’.

Cell proliferation assay
For cell proliferation analysis, 5,000 living cells seeded on 96-well plates were transfected with siCCNA2 001, 002 or an siRNA control. After 48 hours, cell growth was determined at 0, 1, and 2 days by Cell Counting Kit-8 ([CCK-8]; Dojindo Chemical Laboratories, Kumamoto, Japan). Viable cells were measured with CCK-8 reagent 10 µL/well, incubated at 37°C for 2 hours, and then the optical density values were measured at a wavelength of 450 nm using a Bio-Rad iMark™ Microplate Absorbance Reader (Bio-Rad Laboratories Inc., Hercules, CA, USA).

Colony formation assay
Forty-eight hours after SW480 cells were infected with siRNA, about 500 cells were seeded on each well of a six-well plate. The cells were allowed to incubate at 37°C for 10 days. Cells were fixed, stained with crystal violet, and photographed. ImageJ (1.48 u; National Institutes of Health, Bethesda, MA, USA) was used to count the number of clones per well.
Cell cycle analysis
Forty-eight hours after siRNA interference, SW480 cells were harvested, centrifuged, and resuspended in 1× PBS. The cells were fixed in 70% ethanol overnight. On the second day, after being washed by 1× PBS solution and centrifuged, cells were resuspended in 1× PBS solution and incubated with RNase A at 37°C for 30 minutes. Finally, cells were then stained with propidium iodide and analyzed by FACSCalibur system (BD Biosciences, San Jose, CA, USA).

Apoptosis analysis
SW480 cells were infected with siRNA for 48 hours and harvested and centrifuged. Then the supernatant was removed and resuspended in 1× PBS solution. This procedure was repeated thrice in 1×10^6 cells per unit well, then stained by Annexin V/FITC and PI kit. After staining, the cells were analyzed in a FACS Calibur system (BD Biosciences).

Statistical analysis
All experiments were performed at least three times and each independent test was carried out in triplicate for each condition under the protocol and according to the manufacturer’s instructions. All statistical analyses were performed using PASW Statistics 19.0 (IBM, Chicago, IL, USA) or GraphPad Prism 6 software (GraphPad Software, Inc., La Jolla, CA, USA).

Results
Identification of DEGs
Gene expression profiles from GSE21815, GSE32323, GSE44076, and GSE74602 of CRC and normal or adjacent mucous tissues were used in this study. The microarray data of GSE21815 included 131 CRC tissues and nine normal colon tissues. The GSE32323 data included 17 paired CRC tumors and adjacent noncancerous tissues. The GSE44076 data included 98 paired normal adjacent mucosa and tumor samples, and the GSE74602 data had 30 paired normal and tumor colorectal samples. The GEO2R online analysis tool was used to detect DEGs; by using an adjusted P-value <0.05 and |logFC| ≥1.0 as the cutoff criteria, we extracted 8843, 2606, 2678, and 1674 DEGs from the expression profile datasets GSE21815, GSE32323, GSE44076, and GSE74602, respectively. A total of 498 consistently expressed genes were identified from the four profile data sets, in which 294 were upregulated genes and 204 were downregulated genes in CRC tissues, compared to normal colon tissues (Figure 1A and B).

GO function and KEGG pathway enrichment analyses in CRC
To further understand the selected DEGs, GO function and KEGG pathway enrichment analyses were applied by using DAVID. After all the DEG data were input into DAVID software, it turned out that they can be classified into three major functional groups: biological process, cellular component, and molecular function (Table 1). As shown in Table 1, in the biological process group, upregulated genes were mainly enriched in mitotic cell cycle processes, mitotic cell cycle, cell cycle processes, mitotic nuclear division, and cell cycle, and the downregulated genes were mainly enriched in ion homeostasis, inorganic ion homeostasis, chemical homeostasis, cation homeostasis, and cellular cation homeostasis. For GO cell component analysis, the upregulated DEGs were enriched in nucleoplasm, chromosome, chromosomal part, chromosomal region, and condensed chromosome, and

Figure 1 Identification of synchronized upregulated genes (294 DEGs, A) and downregulated genes (204 DEGs, B) from the four cohort profile data sets (GSE21815, GSE32323, GSE44076, and GSE74602) using Funrichnew software.

Notes: The different color areas represent different data sets. The cross-sectional areas mean the synchronized and changed DEGs.
the downregulated DEGs were enriched in the extracellular region, apical part of the cell, extracellular vesicles, and extracellular organelles. In addition, molecular function analysis showed that the upregulated DEGs were significantly enriched in ATP binding, adenylyl nucleotide binding, adenylyl ribonucleotide binding, purine ribonucleoside triphosphate binding, and purine ribonucleoside binding, while the downregulated DEGs were enriched in carbonate dehydratase activity, oxidoreductase activity, acting on CH-OH group of donors, carboxylic ester hydrolase activity, hormone activity, and hydro-lyase activity.

Table 2 lists the most significantly enriched KEGG pathways for the upregulated and downregulated DEGs. The upregulated DEGs were enriched in cell cycle, DNA replication, and mismatch repair (MMR), while the downregulated DEGs were enriched in mineral absorption, aldosterone-regulated sodium reabsorption, pancreatic secretion, nitrogen metabolism, and proximal tubule bicarbonate reclamation.

### Table 1 GO analysis of DEGs associated with CRC

<table>
<thead>
<tr>
<th>Expression</th>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>P-value</th>
<th>P_{BH}-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated</td>
<td>GOTERM_BP_FAT</td>
<td>GO:1903047, mitotic cell cycle process</td>
<td>73</td>
<td>24.91</td>
<td>1.19E–29</td>
<td>4.82E–26</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0000278, mitotic cell cycle</td>
<td>73</td>
<td>24.91</td>
<td>2.23E–27</td>
<td>4.52E–24</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0007067, mitotic nuclear division</td>
<td>45</td>
<td>15.35</td>
<td>5.39E–22</td>
<td>5.45E–19</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0007049, cell cycle</td>
<td>87</td>
<td>29.69</td>
<td>8.40E–22</td>
<td>6.80E–19</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0005654, nucleoplasm</td>
<td>114</td>
<td>38.90</td>
<td>6.25E–16</td>
<td>3.12E–13</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0005694, chromosome</td>
<td>55</td>
<td>18.77</td>
<td>3.81E–14</td>
<td>9.93E–12</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0044427, chromosomal part</td>
<td>49</td>
<td>16.73</td>
<td>9.59E–13</td>
<td>1.56E–10</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0096687, chromosomal region</td>
<td>30</td>
<td>10.23</td>
<td>1.11E–09</td>
<td>8.97E–07</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0000793, condensed chromosome</td>
<td>21</td>
<td>7.67</td>
<td>2.01E–09</td>
<td>1.89E–07</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0005524, ATP binding</td>
<td>55</td>
<td>18.77</td>
<td>2.59E–08</td>
<td>1.75E–05</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0030554, adenyl nucleotide binding</td>
<td>56</td>
<td>19.11</td>
<td>2.87E–08</td>
<td>9.67E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0032559, adenyl ribonucleotide binding</td>
<td>55</td>
<td>18.77</td>
<td>5.76E–08</td>
<td>1.30E–05</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0035639, purine ribonucleoside</td>
<td>57</td>
<td>19.45</td>
<td>3.87E–06</td>
<td>6.52E–04</td>
</tr>
<tr>
<td>Downregulated</td>
<td>GOTERM_ME_FAT</td>
<td>GO:0032550, purine ribonucleoside binding</td>
<td>57</td>
<td>19.45</td>
<td>4.49E–06</td>
<td>6.05E–04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0050801, ion homeostasis</td>
<td>30</td>
<td>14.70</td>
<td>9.58E–10</td>
<td>3.56E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0098771, inorganic ion homeostasis</td>
<td>29</td>
<td>14.21</td>
<td>1.10E–09</td>
<td>2.04E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0048878, chemical homeostasis</td>
<td>37</td>
<td>18.13</td>
<td>1.26E–09</td>
<td>1.57E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0055080, cation homeostasis</td>
<td>28</td>
<td>13.72</td>
<td>3.00E–09</td>
<td>2.79E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_FAT</td>
<td>GO:0030003, cellular cation homeostasis</td>
<td>26</td>
<td>12.74</td>
<td>5.71E–09</td>
<td>4.24E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0044421, extracellular region part</td>
<td>81</td>
<td>39.70</td>
<td>1.07E–09</td>
<td>3.33E–07</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0005576, extracellular region</td>
<td>87</td>
<td>42.64</td>
<td>3.08E–08</td>
<td>4.81E–06</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0045177, apical part of cell</td>
<td>17</td>
<td>8.33</td>
<td>3.30E–06</td>
<td>3.44E–04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:1903561, extracellular vesicle</td>
<td>57</td>
<td>27.94</td>
<td>5.99E–06</td>
<td>4.67E–04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_FAT</td>
<td>GO:0043230, extracellular organelle</td>
<td>57</td>
<td>27.94</td>
<td>6.05E–06</td>
<td>3.78E–04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0004089, carbonate dehydratase activity</td>
<td>5</td>
<td>2.45</td>
<td>1.45E–05</td>
<td>0.00901538</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0016614, oxidoreductase activity, acting on</td>
<td>8</td>
<td>3.92</td>
<td>9.39E–04</td>
<td>0.25479095</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>CH-OH group of donors</td>
<td>5</td>
<td>2.45</td>
<td>0.001359036</td>
<td>0.24672519</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:0052889, carboxylic ester hydrolase activity</td>
<td>8</td>
<td>3.92</td>
<td>0.00252258</td>
<td>0.32608492</td>
</tr>
<tr>
<td></td>
<td>GOTERM_ME_FAT</td>
<td>GO:000179, hormone activity</td>
<td>7</td>
<td>2.43</td>
<td>0.002801112</td>
<td>0.29575964</td>
</tr>
</tbody>
</table>

Notes: P-values were calculated by Fisher’s exact test. P_{BH} values were adjusted by BH method.Abbreviations: BH, Benjamini and Hochberg; BP, biological process; CC, cell component; CRC, colorectal cancer; DEGs, differentially expressed genes; GO, gene ontology; MF, molecular function.

Hub genes and module screening from PPI network and modular analyses

Using the STRING online database and Cytoscape software, a total of 363 DEGs from the 498 synchronously altered DEGs were filtered into the DEG PPI network complex, and 135 of the 498 DEGs did not fall into the DEG PPI network. Among the 363 DEGs, we constructed a PPI network of the top ten hub genes that have increased connectivity (Figure 2), which comprises TOP2A, CDK1, PCNA, MAD2L1, RFC4, CCNA2, BUB1, PAICS, AURKA, and CHEK1. These ten hub genes are listed from the largest to the smallest degree. In order to detect significant modules in this PPI network, we used the MCODE plug-in, by which three modules with
Table 2 KEGG pathway analysis of DEGs associated with CRC

<table>
<thead>
<tr>
<th>Expression Term Count</th>
<th>P-value</th>
<th>Downregulated DEGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell cycle</td>
<td>19</td>
<td>7.79E-12</td>
</tr>
<tr>
<td>DNA replication</td>
<td>12</td>
<td>2.32E-11</td>
</tr>
<tr>
<td>MMR</td>
<td>7</td>
<td>2.80E-06</td>
</tr>
<tr>
<td>Nucleotide excision repair</td>
<td>6</td>
<td>0.001658</td>
</tr>
<tr>
<td>Progesterone-mediated oocyte maturation</td>
<td>7</td>
<td>0.005332</td>
</tr>
<tr>
<td>Mineral absorption</td>
<td>11</td>
<td>4.04E-07</td>
</tr>
<tr>
<td>Aldosterone-regulated sodium reabsorption</td>
<td>8</td>
<td>7.98E-07</td>
</tr>
<tr>
<td>Pancreatic secretion</td>
<td>10</td>
<td>4.45E-05</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>5</td>
<td>6.96E-05</td>
</tr>
<tr>
<td>Proximal tubule</td>
<td>5</td>
<td>2.43E-04</td>
</tr>
</tbody>
</table>

**Expression Term Count**

- **Downregulated DEGs**: Cell cycle, DNA replication, MMR, Nucleotide excision repair, Progesterone-mediated oocyte maturation, Mineral absorption, Aldosterone-regulated sodium reabsorption, Pancreatic secretion, Nitrogen metabolism, Proximal tubule bicarbonate reclamation.

**Notes**: P-values were calculated by Fisher’s exact test. P<sub>adj</sub> values were adjusted by BH method.

**Abbreviations**: CRC, colorectal cancer; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; MMR, mismatch repair.

The highest degree of importance were selected (Figure 3). KEGG pathway enrichment analysis showed that Module 1 was mainly associated with cell cycle, DNA replication, oocyte meiosis, and nucleotide excision repair, Module 2 was mostly linked with ribosome biogenesis in eukaryotes, and Module 3 was primarily connected with chemokine signaling pathway and tumor necrosis factor signaling pathway. Meanwhile, we input ten hub genes into the DAVID software. GO analysis showed that these genes were particularly enriched in the mitotic cell cycle checkpoint, negative regulation of mitotic cell cycle, cell cycle checkpoint, negative regulation of cell cycle, and regulation of mitotic cell cycle (Table S1). KEGG analysis showed that these genes were enriched in cell cycle, progesterone-mediated oocyte maturation, and oocyte meiosis (Table S2).

### The expression level and correlation analyses of the ten hub genes in GEPIA

GEPIA is an interactive online server for exploring the large data sets from the TCGA and the GTEx projects. To confirm the reliability of the ten identified hub genes from the four data sets, we used GEPIA to verify the correlation between them, and they were obviously positively correlated with each other in CRC (Figure 4A). GEPIA was also used to figure out the expression levels of the ten genes in CRC. Figure 4B shows that these genes were all significantly overexpressed in colon cancer (COAD) and rectal cancer (READ) samples compared to the normal samples.

### Expression patterns of CCNA2 in CRC

To identify the expression level of CCNA2 in CRC, we performed qRT-PCR to confirm the expression of CCNA2 in 33 paired clinical samples, in which the mean expression level of CCNA2 was notably higher in CRC tissues than that in normal tissues (Figure 5A). Next, we measured the expression of CCNA2 in various cell lines, including the normal cell line NCM460 and CRC cell lines HT-29, HCT116, SW620, and SW480. The expression of CCNA2 was higher in tumor cells than that in normal cells (Figure 5B), which is similar to the
results from the four datasets in GEO, GEPIA, suggesting that our results on these genes are reliable.

Knockdown of CCNA2 suppressed cell growth by impairing cell cycle progression and inducing cell apoptosis

To determine whether CCNA2 could be a therapeutic target in CRC, we inactivated CCNA2 by using siRNA in SW480 cell. We found that CCNA2 knockdown (Figure 5C), as compared to control knockdown, significantly inhibited cell proliferation (Figure 5D) and reduced cell numbers in SW480 cells (Figure 5E and F), which indicated that CCNA2 might promote cell proliferation. To examine how CCNA2 affects cell growth, the cell cycle phase distribution and apoptosis were analyzed by flow cytometric analysis. Knockdown of CCNA2 resulted in a decrease of cell percentage in the G1 phase, and an increase of cell percentage in the S phase and G2 phase (Figure 5G and H), which indicated that CCNA2 knockdown prevented cell passage from the G2 phase into M phase. Therefore, CCNA2 was shown to promote G2/M...
Figure 4 (A) The correlation analysis of the ten hub genes. (B) Expression levels of the ten hub genes in COaD and ReaD compared to the normal samples.

Notes: R is the Pearson correlation coefficient. *P<0.05.

Abbreviations: COaD, colon adenocarcinoma; ReaD, rectum adenocarcinoma.

Figure 5 CCNA2 knockdown suppressed colon cancer cell proliferation by impairing cell cycle progression and inducing apoptosis.

Notes: (A) Expression level of CCNA2 gene in 33 paired CRC tissues (n=3; ***P<0.001; two-tailed t-test). (B) Expression level of CCNA2 gene in colon normal cell line NCM460 and CRC cell lines HT-29, HCT116, SW620, and SW480 (n=3; **P<0.01, ***P<0.001, ****P<0.0001; two-tailed t-test). (C) Expression level of CCNA2 were examined after siRNA transfection in SW480 cells (n=3; *P<0.05, **P<0.01, ***P<0.001; two-tailed t-test). (D) The cell proliferation rates were analyzed by CCK-8 assay. All values were mean±SD (n=3; ***P<0.001; two-tailed t-test). (E) Graph illustrating quantified values (n=3; ****P<0.0001; two-tailed t-test). (F) Colony formation assays were performed. (G,H) Distribution of cells in three cell cycle phases was examined by flow cytometry assay, and the graph shows quantification for each phase. (I) For measurement of apoptotic cells, cells were stained with both AV and PI, and analyzed by an image flow assay. (J) Graph illustrating the quantification of apoptotic cells, AV+ means early apoptosis, AV+/PI means late apoptosis (n=3; **P<0.01, ***P<0.001, ****P<0.0001; two-tailed t-test).

Abbreviations: AV, Annexin V FITC; CRC, colorectal cancer; CCK-8, cell counting kit-8; PI, propidium iodide; NC, negative control.
phase transition. Apoptosis assay results indicated that the
apoptotic cells significantly increased in SW480 cells with
si-CCNA2 transfection (Figure 5I and J). These data indicate
that CCNA2 knockdown could impair cell cycle progression
and induce cell apoptosis.

Discussion

Even with a progressive decrease in morbidity in the past
few years, CRC is still the fourth leading cause of cancer-
related death worldwide. The development and progression
of CRC is a dynamic process, and the expression levels of
some molecules differ across the different stages of CRC. Under
this circumstance, the difficulty is increasing in early
screening and diagnosis. Hence, it is essential to find sensi-
tive and precise biomarkers of CRC.

In this analysis, GSE21815, GSE32323, GSE44076, and
GSE74602 were selected from the GEO database. A total of
498 consistently expressed genes were identified from these
four profile data sets, in which 294 were upregulated genes
and 204 were downregulated genes in CRC tissues, compared
to normal colon tissues. In order to have a deeper understand-
ing of these selected DEGs, we performed the GO function
and KEGG pathway analyses of these DEGs.

GO analysis showed that the upregulated DEGs were
particularly enriched in mitotic cell cycle processes, mitotic
cell cycle, nucleoplasm, chromosome, ATP binding, and
adenyl nucleotide binding, while the downregulated DEGs
were involved in ion homeostasis, inorganic ion homeostasis,
extracellular regions, carbonate dehydratase activity, and
oxidoreductase activity. In addition, the KEGG pathways for
the upregulated DEGs included the cell cycle, DNA replica-
tion, and MMR, while the pathways of downregulated DEGs
showed that CRC was associated with cell cycle related
tumor progression. MSI accounts for 10%–20% of CRCs and
is the second most important model of colorectal carcino-
genesis. The most common MMR genes with germline
mutations are MSH2, MLH1, PMS2, and MSH6. Ten DEGs with high connectivity were selected as hub
genes for PPI network analysis. Through analyzing the corre-
lation and expression level in GEPIA, we determined that hub
genes were obviously positively correlated and significantly
overexpressed in CRC samples.

We searched PubMed literatures for associations among
the ten hub genes in CRC. RFC4 and TOP2A are related to
tumor progression and poor survival outcome through regu-
lation on cell proliferation and the cell cycle in CRC. In
addition, TOP2A was revealed to be a vital enzyme in DNA
replication and a molecular target of topo II inhibitors which
is a class of anticancer drugs. Overexpression of TOP2A is
associated with resistance to chemotherapy by inhibition of
apoptosis in CRC. CDK1 influences apoptosis by interact-
ing with the iASPP/p53 apoptosis pathway or combined
targeting with a MEK/ERK inhibitor in BRAFV600E
CRCs. Moreover, CDK1 is involved in an iron-regulated
signaling axis to regulate cell cycle and promote colorectal
tumorigenesis. The spindle proteins AURKA, BUB1, and
MAD2L1 are important components of the spindle assembly
checkpoint, which has been frequently established to be an
important mechanism that drives aneuploidy and carcinogen-
ess in CRC. PAICS, a de novo purine biosynthetic gene,
participates in purine metabolism and metabolic pathways.
Using integrated bioinformatics analysis, we identified ten
hub genes, including nine genes that have been documented
for their clinical significance and mechanism in CRC. As a
responsive gene, CCNA2 was chosen for further investigation
because knockdown of CCNA2 (Figure 5C) showed signifi-
cant inhibition of proliferation (Figure 5D) and reduction in
colony formation (Figure 5E and F).

CCNA2, which is normally silenced in the postnatal
mammalian myocardium, can induce cardiac repair in small
animal models of myocardial infarction. It is a functional
target of retinoblastoma-mediated cell cycle arrest. After
a thorough search in PubMed, there was no clear evidence indi-
cating the relationship between CCNA2 and CRC. CCCNA2
was highly expressed in colon cancer according to biologi-
cal information. Moreover, like CCNA2, MAD2L1 have a
high correlation that is obviously positively correlated, and
their Pearson correlation coefficient was 0.88. To examine
the expression level of CCNA2, we found that CCNA2 has a
higher expression in the CRC clinical samples and cell
lines. We further confirmed that knockdown of CCNA2

Analysis of the three selected modules of the PPI network
showed that CRC was associated with cell cycle related
pathway, DNA replication related pathway, and MMR related
pathway. Overall, they were all associated with cell growth,
cell invasion, cell proliferation, the cell cycle, and microsat-
eellite instability (MSI), all of which play critical roles in the
process of tumorigenesis. MSI tumors usually arise because
of a genetic defect in the MMR genes which is one of the
main DNA-repair systems. MMR is primarily involved in
the post-replication elimination of base–base mismatches
and insertion deletion loops (IDLs) that arise as a result of
DNA polymerase slippage during DNA synthesis. MSI
colorectal tumors accumulate mutations at microsatellite
sequences in coding regions of genes implicated in tumor
progression. MSI accounts for 10%–20% of CRCs and
is the second most important model of colorectal carcino-
genesis. The most common MMR genes with germline
mutations are MSH2, MLH1, PMS2, and MSH6. Ten DEGs with high connectivity were selected as hub
genes for PPI network analysis. Through analyzing the corre-
lation and expression level in GEPIA, we determined that hub
genes were obviously positively correlated and significantly
overexpressed in CRC samples.

We searched PubMed literatures for associations among
the ten hub genes in CRC. RFC4 and TOP2A are related to
tumor progression and poor survival outcome through regu-
lation on cell proliferation and the cell cycle in CRC. In
addition, TOP2A was revealed to be a vital enzyme in DNA
replication and a molecular target of topo II inhibitors which
is a class of anticancer drugs. Overexpression of TOP2A is
associated with resistance to chemotherapy by inhibition of
apoptosis in CRC. CDK1 influences apoptosis by interact-
ing with the iASPP/p53 apoptosis pathway or combined
targeting with a MEK/ERK inhibitor in BRAFV600E
CRCs. Moreover, CDK1 is involved in an iron-regulated
signaling axis to regulate cell cycle and promote colorectal
tumorigenesis. The spindle proteins AURKA, BUB1, and
MAD2L1 are important components of the spindle assembly
checkpoint, which has been frequently established to be an
important mechanism that drives aneuploidy and carcinogen-
ess in CRC. PAICS, a de novo purine biosynthetic gene,
participates in purine metabolism and metabolic pathways.
Using integrated bioinformatics analysis, we identified ten
hub genes, including nine genes that have been documented
for their clinical significance and mechanism in CRC. As a
responsive gene, CCNA2 was chosen for further investigation
because knockdown of CCNA2 (Figure 5C) showed signifi-
cant inhibition of proliferation (Figure 5D) and reduction in
colony formation (Figure 5E and F).

CCNA2, which is normally silenced in the postnatal
mammalian myocardium, can induce cardiac repair in small
animal models of myocardial infarction. It is a functional
target of retinoblastoma-mediated cell cycle arrest. After
a thorough search in PubMed, there was no clear evidence indi-
cating the relationship between CCNA2 and CRC. CCCNA2
was highly expressed in colon cancer according to biologi-
cal information. Moreover, like CCNA2, MAD2L1 have a
high correlation that is obviously positively correlated, and
their Pearson correlation coefficient was 0.88. To examine
the expression level of CCNA2, we found that CCNA2 has a
higher expression in the CRC clinical samples and cell
lines. We further confirmed that knockdown of CCNA2
could significantly suppress CRC cell growth by impairing cell cycle progression and inducing cell apoptosis. CCNA2 has the potential to be a new biomarker for diagnosis and CRC therapy.

In summary, using four cohorts profile data sets and multiple bioinformatics analyses, our present work identified ten hub genes as DEGs. These DEGs are significantly enriched in several pathways that are mainly associated with the cell cycle, DNA replication, and progesterone-mediated oocyte maturation in CRC, and they might play key roles in the development and progression of CRC. CCNA2 shows higher expression level in CRC, involving in colon cancer cell growth and cell cycle progression, which could be used as a new biomarker, and has significant meanings for clinical treatment.

Conclusion
In this study, using four cohorts profile data sets and multiple bioinformatics analyses, we identified ten hub genes which are significantly enriched in the cell cycle, DNA replication, and progesterone-mediated oocyte maturation pathways in CRC. Moreover, the expression level of CCNA2 was significantly increased in CRC and knockdown of CCNA2 suppressed colon cancer cell growth by impairing cell cycle and apoptosis progression. Our findings also establish that CCNA2 could be a new biomarker for diagnosis and guide the combination therapy for CRC.

Acknowledgment
This study was supported by grants from the National Key R&D Program of China, Stem Cell and Translational Research (No. 2016YFA0102000) and the National Natural Science Foundation of China (No. 81572900).

Disclosure
The authors report no conflicts of interest in this work.

References


Supplementary materials

Table S1 GO analysis of ten hub genes associated with CRC

<table>
<thead>
<tr>
<th>Expression</th>
<th>Category</th>
<th>Term Description</th>
<th>Count</th>
<th>%</th>
<th>P-value</th>
<th>PBH value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten hub genes</td>
<td>GOTEERM_BP_FAT</td>
<td>GO:0007093, mitotic cell cycle checkpoint</td>
<td>8</td>
<td>80</td>
<td>2.25E−13</td>
<td>1.81E−10</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_BP_FAT</td>
<td>GO:0045930, negative regulation of mitotic cell cycle</td>
<td>8</td>
<td>80</td>
<td>2.18E−12</td>
<td>8.79E−10</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_BP_FAT</td>
<td>GO:0000075, cell cycle checkpoint</td>
<td>8</td>
<td>80</td>
<td>2.99E−12</td>
<td>8.04E−10</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_BP_FAT</td>
<td>GO:0045786, negative regulation of cell cycle</td>
<td>8</td>
<td>80</td>
<td>4.69E−10</td>
<td>9.45E−08</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_BP_FAT</td>
<td>GO:007346, regulation of mitotic cell cycle</td>
<td>8</td>
<td>80</td>
<td>6.36E−10</td>
<td>1.03E−07</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_CC_FAT</td>
<td>GO:0005894, chromosome</td>
<td>8</td>
<td>80</td>
<td>1.46E−07</td>
<td>1.31E−05</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_CC_FAT</td>
<td>GO:0098687, chromosomal region</td>
<td>6</td>
<td>60</td>
<td>8.67E−07</td>
<td>3.90E−05</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_CC_FAT</td>
<td>GO:0044427, chromosomal part</td>
<td>7</td>
<td>70</td>
<td>2.42E−06</td>
<td>7.27E−05</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_CC_FAT</td>
<td>GO:0000793, condensed chromosome</td>
<td>5</td>
<td>50</td>
<td>5.05E−06</td>
<td>1.14E−04</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_CC_FAT</td>
<td>GO:0000228, nuclear chromosome</td>
<td>6</td>
<td>60</td>
<td>8.87E−06</td>
<td>1.60E−04</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_MF_FAT</td>
<td>GO:0035173, histone kinase activity</td>
<td>3</td>
<td>30</td>
<td>5.11E−05</td>
<td>5.5E−03</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_MF_FAT</td>
<td>GO:0005524, ATP binding</td>
<td>7</td>
<td>70</td>
<td>5.22E−05</td>
<td>2.84E−03</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_MF_FAT</td>
<td>GO:0032559, adenyl ribonucleotide binding</td>
<td>7</td>
<td>70</td>
<td>5.99E−05</td>
<td>2.17E−03</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_MF_FAT</td>
<td>GO:0030554, adenyl nucleotide binding</td>
<td>7</td>
<td>70</td>
<td>6.24E−05</td>
<td>1.70E−03</td>
</tr>
<tr>
<td></td>
<td>GOTEERM_MF_FAT</td>
<td>GO:0035639, purine ribonucleoside triphosphate binding</td>
<td>7</td>
<td>70</td>
<td>1.71E−04</td>
<td>3.73E−03</td>
</tr>
</tbody>
</table>

Abbreviations: CRC, colorectal cancer; GO, gene ontology; BP, biological process; CC, cell component; MF, molecular function.

Table S2 KEGG pathway analysis of ten hub genes associated with CRC

<table>
<thead>
<tr>
<th>Expression</th>
<th>Term Description</th>
<th>Count</th>
<th>P-value</th>
<th>PBH value</th>
<th>Genes included in the pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten hub genes</td>
<td>Cell cycle</td>
<td>6</td>
<td>9.21E−08</td>
<td>1.66E−06</td>
<td>CCNA2, BUB1, CHEK1, PCNA, MAD2L1, CDK1</td>
</tr>
<tr>
<td></td>
<td>Progestosterone-mediated oocyte maturation</td>
<td>4</td>
<td>1.03E−04</td>
<td>9.28E−04</td>
<td>CCNA2, BUB1, MAD2L1, CDK1</td>
</tr>
<tr>
<td></td>
<td>Oocyte meiosis</td>
<td>4</td>
<td>2.02E−04</td>
<td>0.0012106</td>
<td>BUB1, MAD2L1, CDK1, AURKA</td>
</tr>
</tbody>
</table>

Abbreviations: CRC, colorectal cancer; KEGG, Kyoto Encyclopedia of Genes and Genomes.