Research Article

Identification of Common Genes Refers to Colorectal Carcinogenesis with Paired Cancer and Noncancer Samples

Lihua Zhang, Yonglong Yang, Lin Cheng, Yu Cheng, Hong-Hao Zhou, and Zhi-Rong Tan

1Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha 410008, China
2Institute of Clinical Pharmacology, Central South University, Hunan Key Laboratory of Pharmacogenetics, Changsha 410078, China
3Haikou People’s Hospital and Affiliated Haikou Hospital of Xiangya Medical School, Central South University, Haikou, Hainan 570311, China
4State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, 54 South Xianlie Road, Guangzhou, Guangdong, China

Correspondence should be addressed to Hong-Hao Zhou; hhzhou2003@163.com and Zhi-Rong Tan; zrtan800820@163.com

Received 21 June 2017; Accepted 16 October 2017; Published 30 January 2018

Copyright © 2018 Lihua Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Colorectal cancer is a malignant tumor which harmed human beings’ health. The aim of this study was to explore common biomarkers associated with colorectal carcinogenesis in paired cancer and noncancer samples. At first, fifty-nine pairs of colorectal cancer and noncancer samples from three gene expression datasets were collected and analyzed. Then, 181 upregulation and 282 downregulation common differential expression genes (DEGs) were found. Next, functional annotation was performed in the DAVID database with the DEGs. Finally, real-time polymerase chain reaction (PCR) assay was conducted to verify the analyses in sixteen colorectal cancer and individual-matched adjacent mucosa samples. Real-time PCR showed that $\text{MCM2}$, $\text{RNASEH2A}$, and $\text{TOP2A}$ were upregulated in colorectal cancer compared with adjacent mucosa samples ($\text{MCM2}$, $P < 0.001$; $\text{RNASEH2A}$, $P < 0.001$; $\text{TOP2A}$, $P = 0.001$). These suggested that 463 DEGs might contribute to colorectal carcinogenesis.

1. Introduction

Colorectal cancer is one of the most commonly diagnosed cancers worldwide, with approximately 1.4 million cases and 693,900 deaths in 2012 [1]. Some risk factors may lead to colorectal cancer such as family history [2, 3], inflammatory bowel disease [4], and smoking [5, 6]. In the molecular level, colorectal cancer is a heterogeneous disease. Mutations, epigenetic changes, and expression differences of multiple genes are well-known colorectal cancer contributors. However, the underlying molecular mechanism of colorectal carcinogenesis has not been fully understood yet.

More than twenty years ago, complementary DNA microarray has been used to analyze the gene expression patterns in human cancer [7]. With the development of technology, more and more expression profile platforms and researches on colorectal cancer have been released in recent years. Currently, a large number of expression profile datasets were uploaded and shared publically on several public databases, and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo) and ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) were included. The public shared data of the databases could benefit the researchers greatly in finding interesting research targets or verify their ideas easily.

In this study, to further understand the mechanism of colorectal carcinogenesis from the gene expression level, we searched the GEO database for paired colorectal cancer and noncancer samples. Fifty-nine paired samples from GSE21510, GSE23878, and GSE32323 were selected in total. The DEGs were screened from three datasets, and 463 common DEGs (181 upregulation and 282 downregulation genes) were detected and sent to the DAVID database (https://david.ncifcrf.gov) for functional annotation. Three
upregulated genes were chosen to verify the analysis results in sixteen pairs of colorectal cancer and adjacent mucosa tissues. Real-time PCR results showed that MCM2, RNA-SEH2A, and TOP2A were associated with colorectal carcinogenesis.

2. Materials and Methods

2.1. Microarray Data. We searched the NCBI-GEO database with the following keywords: Human Genome U133 Plus 2.0, colorectal cancer, and normal (Human Genome U133 Plus 2.0 Array is a platform of Affymetrix). Fifty-nine paired colorectal cancer and noncancerous samples were selected from three gene expression datasets (twenty-three pairs in GSE21510 from Japan, seventeen pairs in GSE23878 from Saudi Arabia, and nineteen pairs in GSE32323 from Japan). The cancer and noncancerous samples were divided into cancer and noncancer groups, and the noncancer group was used as control.

2.2. Identification of DEGs. The selected samples of GSE21510, GSE23878, and GSE32323 were separately analyzed by R software (https://www.r-project.org). At first, raw data files of the three datasets were downloaded from the GEO database. Then, the Robust Multichip Average algorithm was applied to perform background correction and quantile normalization with the R package Affy [8, 9]. Next, the probability of genes being differentially expressed between cancer and noncancer groups were calculated by the Linear Models for Microarray Data (LIMMA) package. Finally, the DEGs were selected under corrected \( P < 0.05 \) and \[\text{fold change}\] \( \geq 2.0 \) criteria. The volcano plots of differentially expressed genes were also performed by R software [10]. The DEGs from each dataset were intersected to identify common DEGs.

2.3. Gene Ontology and Pathway Enrichment Analysis of DEGs. Gene Ontology (GO, http://www.geneontology.org) was a framework for the model of biology that describes the attributes of gene products. It was classified into three aspects: molecular function, biological process, and cellular component [11]. Pathway analysis was a popular method to analyze microarray data in a more detailed, specific way [12, 13]. DAVID database was a free online bioinformatics resource that aimed to provide functional interpretation of large lists of genes derived from genomic studies [14]. The common DEGs were input in the DAVID database to perform the GO and pathway analysis [15]. \( P < 0.05 \) was considered statistically significant.

2.4. Patients and Tissue Specimens. Tumor tissue samples and individual-matched adjacent mucosa samples were obtained from sixteen patients with colorectal cancer who underwent resection at Xiangya Hospital between 2014 and 2016, and the adjacent mucosa samples were acquired 2–5 cm away from the tumor. The dissected tissue samples were collected in the operating room and stored immediately in liquid nitrogen. This study was approved by the Institutional Review Board of Department of Clinical Pharmacology, Xiangya Hospital, Central South University (registration number: CTXY-150001-2) and by Chinese Clinical Trial Registry (registration number: ChiCTR-DCD-15006289).

2.5. RNA Isolation, Reverse Transcription, and Real-Time PCR. Tissue specimens were grounded and added with TRIzol reagent (Takara). Then the total RNA were isolated, and 1 \( \mu \)g of RNA was reverse-transcribed with PrimeScript 1st Strand complementary DNA Synthesis kit (Takara). Real-time PCR assay was performed on ABI 7500 platform. SYBR Premix Dimer Eraser kit (Takara) was used in 20 \( \mu \)l reaction volume, and the cycling conditions were as follows: an initial 30 s denaturation at 95°C and 45 cycles (5 s at 95°C, 30 s at 55°C, and 34 s at 72°C). PPIA and B2M genes were set as internal controls. MCM2, RNASEH2A, and TOP2A expression level was detected in sixteen pairs of colorectal cancer and adjacent mucosa samples. The primer sequences were shown in Supplementary Table 1.

2.6. Statistical Analysis. Statistical analyses were performed with SPSS version 18.0 (IBM Corporation) and GraphPad Prism 6.0 software (GraphPad Software Inc.). DEGs were determined using t-statistics from the LIMMA Bioconductor package. The real-time PCR data was analyzed with 2\(^{-\Delta\Delta Ct}\) and the significance of the difference between the cancer and noncancer groups was evaluated by two-tailed Student’s t-test. \( P \) values less than 0.05 were considered statistically significant.

3. Results

3.1. Identification of DEGs. Analysis results showed 1623 upregulated and 1179 downregulated DEGs in GSE21510 (Figure 1(a)), 284 upregulated and 627 downregulated DEGs in GSE23878 (Figure 1(b)), and 717 upregulated and 719 downregulated DEGs (Figure 1(c)) in GSE32323. Then, the DEGs of the three datasets were merged. 463 common DEGs were found in total (Supplementary Table 1), consisting of 181 upregulated and 282 downregulated genes (Figures 1(d)–1(e)).

3.2. GO Term Enrichment Analysis. The 181 upregulated genes were uploaded to the DAVID database for GO analysis. The results showed that upregulated DEGs were significantly enriched in biological process, including 22 DEGs in cell division \( (P = 1.48\times10^{-11}) \) and 19 in mitotic nuclear division \( (P = 2.06\times10^{-11}) \) GO terms. In the cellular component, nucleoplasm \( (P = 4.95\times10^{-7}) \), spindle microtubule \( (P = 2.49\times10^{-6}) \), and spindle pole \( (P = 6.09\times10^{-6}) \) were the top three GO terms, and there were 51 DEGs enriched in the nucleoplasm. In molecular function, the top three GO terms were frizzled binding \( (P = 3.70\times10^{-4}) \), protein binding \( (P = 4.65\times10^{-4}) \), and microtubule binding \( (P = 8.73\times10^{-4}) \) (Table 1).

The 282 downregulated DEGs were also imported into the DAVID database. GO analysis results showed that the top three terms in biological process were bicarbonate transport \( (P = 2.88\times10^{-6}) \), negative regulation of growth \( (P = 5.88\times10^{-6}) \), and cellular response to zinc ion \( (P = 5.88\times10^{-6}) \). Extracellular exosome \( (P = 4.45\times10^{-8}) \), extracellular space \( (P = 2.88\times10^{-6}) \), and bicarbonate transport.
(P = 8.19E−05) in cellular component were the top three GO terms. In molecular function, carbonate dehydratase activity (P = 3.51E−05), chloride channel activity (P = 1.11 E − 04), and oxidoreductase activity (P = 0.0022) were the top three GO terms (Table 1).

3.3. KEGG Pathway Analysis. After all the DEGs were input into the DAVID database, KEGG pathway analysis results were also acquired. In Table 2, the most significantly enriched pathway of the upregulated and downregulated pathways was set out.

The upregulated DEGs were significantly enriched in the cell cycle (P = 2.90E−07), p53 signaling pathway (P = 0.028), and DNA replication pathways (P = 0.049) (Table 2). The downregulated DEGs were enriched in mineral absorption (P = 8.10E−07), nitrogen metabolism (P = 1.57E−04), bile secretion (P = 0.001), retinol metabolism (P = 0.005), proximal tubule bicarbonate reclamation (P = 0.007), pancreatic secretion (P = 0.020), pentose and glucuronate interconversions (P = 0.023), renin secretion (P = 0.023) pathways, and so forth (Table 2).

3.4. Real-Time PCR Validation of DEGs. To test the DEGs of the analysis, MCM2, RNASEH2A, and TOP2A were chosen to conduct real-time PCR in sixteen pairs of colorectal cancer and adjacent mucosa samples. MCM2, RNASEH2A, and TOP2A expression level in cancer samples was significantly higher than noncancer samples (MCM2, P < 0.001, Figure 2(a); RNASEH2A, P < 0.001, Figure 2(b); TOP2A, P = 0.001, Figure 2(c)).

4. Discussion
Colorectal cancer was a complex disease. To get more information for colorectal cancer occurrence, the gene expression data of fifty-nine paired colorectal cancer and noncancer...
tissues was extracted from the GSE21510, GSE23878, and GSE32323 datasets. The fifty-nine paired samples were from two countries: Japan (40 samples from SE21510 and GSE32323) and Saudi Arabia (19 samples from SE32323). The regional divergence of the samples might contribute to finding the common DEGs from two the ethnic groups.

After analysis, 181 upregulated DEGs and 282 downregulated common DEGs were screened. Pathway analyses showed that the upregulated DEGs were mainly involved in cell cycle pathway (n = 11), p53 signaling pathway (n = 4), and DNA replication pathway (n = 3).

Cell cycle plays a crucial role in tumor evolution and progression, so it acts as an important target for antitumor drugs, such as paclitaxel, vincristine, and Adriamycin. In this study, 3 genes (CDK1, CDK4, and CCNB1) refer to cell cycle control [16, 17]. CCNB1 was also a regulatory protein involved in mitosis [18]. In the p53 signaling pathway, the RRM2 gene was an oncogene that overexpressed in colorectal cancer, with its elevated expression correlated with invasion depth, poorly differentiated type, and tumor node metastasis stage [19]. Transcription factor E2F1 could promote RRM2 expression in colorectal cancer cell lines [20].

DNA replication is an important pathway in carcinogenesis, which ranked the third in upregulated DEGs. Three upregulated DEGs refer to DNA replication were MCM2, RFC3, and RNASEH2A. MCM2 was a member of the MCM family (MCM2-7), and all 6 members of this family could form a hexameric protein complex with each other. This complex worked as a DNA helicase to untie the DNA double helix at the initiation stage of DNA synthesis [21]. MCM2 expression was reported to be associated with colorectal cancer stage and prognosis [22] and used to detect colorectal cancer in stool [23].

Eukaryotic RNase H2 was a heterotrimeric enzyme formed by RNASEH2A, RNASEH2B, and RNASEH2C. RNASEH2A was a catalytic subunit that could hydrolyze RNA/DNA hybrid substrate with the structural support from RNASEH2B and RNASEH2C subunits [24]. It had been

<table>
<thead>
<tr>
<th>Expression</th>
<th>Category</th>
<th>Term/gene function</th>
<th>Count</th>
<th>%</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0045926--negative regulation of growth</td>
<td>6</td>
<td>1.60</td>
<td>5.88E-06</td>
<td></td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0006730--one-carbon metabolic process</td>
<td>6</td>
<td>1.60</td>
<td>6.33E-05</td>
<td></td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0007586--digestion</td>
<td>7</td>
<td>1.87</td>
<td>2.88E-04</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0070062--extracellular exosome</td>
<td>75</td>
<td>20.00</td>
<td>4.45E-08</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005615--extracellular space</td>
<td>42</td>
<td>11.20</td>
<td>2.88E-06</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:00031526--brush border membrane</td>
<td>7</td>
<td>1.87</td>
<td>8.19E-05</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0048471--perinuclear region of cytoplasm</td>
<td>19</td>
<td>5.07</td>
<td>0.0035</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0016021--integral component of membrane</td>
<td>93</td>
<td>24.81</td>
<td>0.0067</td>
<td></td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004081--carbonate dehydratase activity</td>
<td>5</td>
<td>1.33</td>
<td>3.51E-03</td>
<td></td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0002524--chloride channel activity</td>
<td>7</td>
<td>1.87</td>
<td>1.11E-04</td>
<td></td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0016491--oxidoreductase activity</td>
<td>10</td>
<td>2.67</td>
<td>2.06E-04</td>
<td></td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008201--heparin binding</td>
<td>8</td>
<td>2.13</td>
<td>9.80E-04</td>
<td></td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0005179--hormone activity</td>
<td>6</td>
<td>1.60</td>
<td>0.0108</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Go analysis of DEGs between paired tumor and nontumor sample.
reported that RNASEH2A showed higher expression level in colorectal cancer [25], and this was validated in our research. TOP2A was a gene that involves copy number variation and chromosomal instability in many cancers [26–29]. In colorectal cancer, protein expression level of TOP2A was related to aggressive tumor phenotype and advanced tumor stage [30]. In our research, we found that TOP2A mRNA expression level was upregulated in colorectal cancer.

Table 2: Pathway analysis of DEGs between paired tumor and nontumor samples.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Pathway</th>
<th>Count</th>
<th>Term/gene function</th>
<th>%</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated</td>
<td>has04110: Cell cycle</td>
<td>11</td>
<td>CCNB1, CDK1, CDC6, MAD2L1, TTK, BUB1B, ORC6, MCM2, PTTG1, CDK4, CDC25B</td>
<td>4.13</td>
<td>2.90E–07</td>
</tr>
<tr>
<td></td>
<td>has04115: p53 signaling pathway</td>
<td>4</td>
<td>CCNB1, CDK1, RRM2, CDK4</td>
<td>1.50</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>has03030: DNA replication</td>
<td>3</td>
<td>RFC3, MCM2, RNASEH2A</td>
<td>1.13</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>has04978: mineral absorption</td>
<td>9</td>
<td>SLC26A3, TRPM6, MT1M, MT2A, CYBRD1, MT1E, MT1H, MT1X, MT1F</td>
<td>2.40</td>
<td>8.10E–07</td>
</tr>
<tr>
<td></td>
<td>has00910: nitrogen metabolism</td>
<td>5</td>
<td>CA12, CA7, CA4, CA2, CA1</td>
<td>1.33</td>
<td>1.57E–04</td>
</tr>
<tr>
<td></td>
<td>has04976: bile secretion</td>
<td>7</td>
<td>AQP8, PRKACB, SLC51B, CA2, SLC51A, SLC4A4, ABCG2</td>
<td>1.87</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>has00830: retinol metabolism</td>
<td>6</td>
<td>ALDH1A1, ADH1C, DHR9, ADH1B, UGT2A3, RETSAT</td>
<td>1.60</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>has04964: proximal tubule</td>
<td>4</td>
<td>CA4, CA2, SLC4A4, PCK1</td>
<td>1.07</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>has04972: pancreatic secretion</td>
<td>6</td>
<td>SLC26A3, CLCA1, CLCA4, PL2G10, CA2, SLC4A4</td>
<td>1.60</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>has00404: pentose and glucuronate interconversions</td>
<td>4</td>
<td>AKR1B10, UGDH, UGT2A3, UGP2</td>
<td>1.07</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>has04924: renin secretion</td>
<td>5</td>
<td>CLCA1, CLCA4, GNAI1, PDE3A, PRKACB</td>
<td>1.33</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>has04960: aldosterone-regulated sodium reabsorption</td>
<td>4</td>
<td>SGK1, NR3C2, HSD11B2, SCNN1B</td>
<td>1.07</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>has04530: tight junction</td>
<td>7</td>
<td>CLDN8, EPB41L3, GNAI1, MYH11, JAM2, CLDN23, MYL9</td>
<td>1.87</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>has05204: chemical carcinogenesis</td>
<td>5</td>
<td>NAT2, ADH1C, ADH1B, SULT1A2, UGT2A3</td>
<td>1.33</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>has04670: leukocyte transendothelial migration</td>
<td>6</td>
<td>CLDN8, GNAI1, JAM2, CXCL12, CLDN23, MYL9</td>
<td>1.60</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>has05030: cocaine addiction</td>
<td>4</td>
<td>GNAI1, MAOA, PRKACB, FOSB</td>
<td>1.07</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Figure 2: Validation of the differentially expressed genes. (a) Validation of mRNA expression of MCM2 expression in sixteen colorectal cancer and individual-matched adjacent mucosa samples. MCM2 expression level in colorectal cancer samples was significantly higher than that in adjacent mucosa samples (P < 0.001). (b) Validation of mRNA expression of RNASEH2A in sixteen colorectal cancer and individual-matched adjacent mucosa samples. RNASEH2A expression level in colorectal cancer samples was significantly higher than that in adjacent mucosa samples (P < 0.001). (c) Validation of mRNA expression of TOP2A in sixteen colorectal cancer and individual-matched adjacent mucosa samples. TOP2A expression level in colorectal cancer samples was significantly higher than that in adjacent mucosa samples (P = 0.001). **P < 0.01 and ***P < 0.001.

The downregulated DEGs are mainly involved in mineral absorption, nitrogen metabolism, bile secretion, retinol metabolism, proximal tubule bicarbonate reclamation, pancreatic secretion and so on, which may signal that the tumor cells lose some function for metabolism of the normal colorectal epithelial cell.

In real-time PCR assay, the commonly used internal genes were GAPDH, β-actin, tubulin, and so forth. However,
in colorectal cancer tissues, GAPDH was not a good internal control gene since it showed higher transcription level than in normal mucosa, nor was β-actin [31, 32]. The combined application of B2M and PPIA were better internal control than others in colorectal cancer [33]. Therefore, B2M and PPIA were used as internal control in our study.

In summary, these gene expression profile analysis results suggested 463 candidate biomarkers for early screening and diagnosis of colorectal cancer. Our study confirmed that MCM2, RNASEH2A, and TOP2A were upregulated in colorectal cancer. The protein expression level and functional studies of these markers were warranted to reveal the molecular mechanism of colorectal cancer development.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant no. 81503563), Hainan Provincial Natural Science Foundation of China (Grant no. 310148), Haikou Key Scientific and Technological Project (Grant no. 2012-074).

Supplementary Materials

Part 1. Primer sequences for real-time PCR. Part 2. 181 upregulated and 282 downregulated common differential expressed genes in colorectal cancer samples. (Supplementary Materials)

References


