Predicting MicroRNA Target Genes and Identifying Hub Genes in IIA Stage Colon Cancer Patients Using Bioinformatics Analysis

Zhiyong Dong,1 Wei Lin,2 Stacy A. Kujawa,3 Shike Wu,4 and Cunchuan Wang1

1Department of Gastrointestinal Surgery, The First Affiliated Hospital of Jinan University, No. 601, Huangpu Ave, Guangzhou, Guangdong, 510630, China
2Department of Gastrointestinal Surgery, The Affiliated Hospital of Putian University, No. 999, Dongzhen Road, Putian, 351100, China
3Robert H Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, 303 E. Superior St., Chicago, IL, 60611, USA
4Department of Gastrointestinal and Anal Surgery, Ruikang Hospital, Guangxi Traditional Chinese Medical University, No. 10, Huadong Road, Nanning 530001, China

Correspondence should be addressed to Cunchuan Wang; twcc2015@163.com

Received 4 October 2018; Accepted 26 December 2018; Published 7 February 2019

Academic Editor: Giandomenico Roviello

Copyright © 2019 Zhiyong Dong 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.

Background. Colon cancer is a heterogeneous disease, differing in clinical symptoms, epigenetics, and prognosis for each individual patient. Identifying the core genes is important for early diagnosis and it provides a more precise method for treating colon cancer. Materials and Methods. In this study, we wanted to pinpoint these core genes so we obtained GSE101502 microRNA profiles from the GEO database, which resulted in 17 differential expressed microRNAs that were identified by GEO2R analysis. Then, 875 upregulated and 2920 downregulated target genes were predicted by FunRich. GO and KEGG pathway were used to do enrich analysis. Results. GO analysis indicated that upregulated genes were significantly enriched in the regulation of cell communication and signaling and in nervous system development, while the downregulated genes were significantly enriched in nervous system development and regulation of transcription from the RNA polymerase II promoter. KEGG pathway analysis suggested that the upregulated genes were enriched in axon guidance, MAPK signaling pathway, and endocytosis, while the downregulated genes existed in pathways in cancer, focal adhesion, and PI3K-Akt signaling pathway. The top four molecules including 82 hub genes were identified from the PPI network and involved in endocytosis, spliceosome, TGF-beta signaling pathway, and lysosome. Finally, NUDT21, GNB1, CLINT1, and COLIA2 core gene were selected due to their correlation with the prognosis of IIA stage colon cancer. Conclusion. this study suggested that NUDT21, GNB1, CLINT1, and COLIA2 might be the core genes for colon cancer that play an important role in the development and prognosis of IIA stage colon cancer.

1. Introduction

Colon cancer is the second most commonly diagnosed cancer and the fourth leading cause of cancer death worldwide. It has been estimated that there were 1,360,600 new cases and 693,900 deaths of colon and rectum cancer worldwide in 2012 [1]. The American Cancer Society estimated that the incidence of colon cancer (71%) is higher than rectum cancer (29%) [2]. Colon cancer is a heterogeneous disease, differing in clinical symptoms, gene mutation or alteration, epigenetics, prognosis, and the response to therapy [3]. It is reported that multiple genes and pathways play a role in the occurrence and development of colon cancer [4]. Moreover, colon cancer is a global burden due to the rising healthcare costs to manage the disease.

MicroRNA (miRNA) is a small endogenous, noncoding RNA molecule, which is composed of approximately 21-25 nucleotides. These small miRNAs usually target one or more mRNA, regulating gene expression through translation level inhibition or breaking target mRNAs [5]. miRNAs characterize an innovative epigenetic mechanism that controls gene expression in several pathological conditions within
2. Materials and Methods

2.1. Database and MicroRNA Selection. GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/) is a public genomics database, including gene array, RNA-seq, DNA-seq, and ChIP-seq based data [16]. "Colon cancer" AND "microRNA" AND "Homo sapiens" keywords were used to search related gene expression profiles by GEO datasets. The GSE101502 profile included three IIA stage colon cancer tissues and three normal colon mucosa tissues.

2.2. Identifying Differentially Expressed MicroRNA. GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) is an online statistics tool that allows user to compare different groups of samples to identify differential microRNA across experimental conditions. We performed a t-test to identify differential microRNA. logFC ≥ 2 cutoff and P value < 0.05 were considered to have a statistically significant difference whereas logFC ≥ 2 was upregulated microRNA and logFC ≤ -2 was downregulated [16].

2.3. Predicting Target Genes. FunRich (functional enrichment) is an analysis tool used for functional enrichment and protein–protein interaction network analysis for genes or proteins. The microRNA enrichment function in FunRich could be used to perform microRNA enrichment analysis, to predict targets of microRNAs, or to find microRNAs through given target genes. Functional analysis of differentially expressed microRNA target genes was conducted to predict target genes with FunRich [17].

2.4. GO and KEGG Pathway Analysis of DEGs. GO (Gene Ontology) analysis is a common advantage method for annotating genes and classifying characteristic biological attributes for high-throughput genome and transcriptome data. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database used in conducting searches regarding genomes, biological pathways, diseases, drugs, and chemical substances. DAVID (Database for Annotation, Visualization and Integrated Discovery, https://david.ncifcrf.gov/) is an online bioinformatics tool that is utilized to provide the functional understanding of large lists of genes. P < 0.05 was set as the cutoff criterion. We conducted key biological processes (BP), molecular functions (MF), cellular components (CC), and pathways among those DEGs by DAVID [18, 19].

2.5. PPI Network and Modules Analysis. STRING (the Retrieval of Interacting Genes, https://string-db.org/) is a web tool created to evaluate PPI (protein–protein interaction) networks information. To detect the potential relationship among those DEGs, we used Cytoscape software and a confidence score of ≥ 0.4 was set as the cutoff criterion. MCODE (Molecular Complex Detection) app in Cytoscape was utilized to display modules of PPI network with node score cutoff = 0.2, k-core = 2, max. depth from seed = 100, and degree cutoff = 2. Then, the top four molecules were mapped into STRING [20, 21].
2.6. Comparison of the Hub Genes Expression Level. The GEPIA (http://gepia.cancer-pku.cn/index.html) is an interactive online tool for analyzing the RNA-seq expression data of 9,736 tumor samples and 8,587 normal samples from the TCGA (the Cancer Genome Atlas dataset, found by NCI and NHGRI, multidimensional maps of important genomic changes in 33 types of cancer) and the GTEx projects (Genotype-Tissue Expression projects, launched by NIH, is a tissue bank and resource for biological research), with a standard processing pipeline. It offers customizable functions such as tumor and normal tissue gene differential expression analysis, and we can determine the expression of hub genes in colon cancer tissues and normal colon mucosa tissues. Survival analysis is then performed to show the high expression and low expression hub genes relationship of colon cancer and overall survival. P<0.05 was considered as significantly different. The boxplot was conducted to visualize the association between cancer and normal tissue [22].

3. Results

3.1. MicroRNA Data. The gene expression profiles for GSE101502, “microRNA expression profiling in human colon cancer”, were obtained from GEO datasets (https://www.ncbi.nlm.nih.gov/geo/). GSE101502, which was based on the GPL21439 platform (miRCURY LNA microRNA Array, 7th generation hsa, mmu, and rno [miRBase 21; probe ID version]), was submitted by Huang et al. on Jul 18th, 2017. The GSE101502 dataset contained three male patients’ tissues comprised of six samples including three IIA stage colon cancer tissues and three normal colon mucosa tissues. Table 1 showed the characteristics of tissues’ information from GSE101502.
| Group          | Accession | Patient No. | Disease state       | Stage      | Status after 1 years | Age | Sex | Tissue                      |
|---------------|-----------|-------------|---------------------|------------|----------------------|-----|-----|-----------------------------|
| Cancer group  | GSM2705118| Patient 1   | Primary colon cancer| TNM:IIA    | alive                | 62  | male | Cancer tissue               |
|               | GSM2705119| Patient 2   | Primary colon cancer| TNM:IIA    | alive                | 62  | male | Cancer tissue               |
|               | GSM2705120| Patient 3   | Primary colon cancer| TNM:IIA    | alive                | 64  | male | Adjacent normal mucosa      |
| Normal group  | GSM2705121| Patient 4   | Primary colon cancer| TNM:IIA    | alive                | 62  | male | Adjacent normal mucosa      |
|               | GSM2705122| Patient 5   | Primary colon cancer| TNM:IIA    | alive                | 62  | male | Adjacent normal mucosa      |
|               | GSM2705123| Patient 6   | Primary colon cancer| TNM:IIA    | alive                | 64  | male | Adjacent normal mucosa      |
Table 2: Identification of differently expressed miRNA.

| miRNA Name     | adj.P.Val | P.Value  | t         | B         | logFC         |
|----------------|-----------|----------|-----------|-----------|---------------|
| hsa-miR-5195-3p| 0.863     | 0.00755  | 3.480604  | -4.53     | 2.143968      |
| hsa-miR-548aw  | 0.863     | 0.04462  | 2.481081  | -4.58     | 2.063515      |
| hsa-miR-5681a  | 0.863     | 0.02794  | -2.816049 | -4.57     | -2.112315     |
| hsa-miR-561-3p | 0.863     | 0.04632  | -2.559382 | -4.58     | -2.141209     |
| hsa-miR-4777-3p| 0.863     | 0.01428  | -3.314158 | -4.56     | -2.160746     |
| hsa-miR-500a-3p| 0.863     | 0.03301  | -2.541129 | -4.55     | -2.226463     |
| hsa-miR-29c-3p | 0.863     | 0.03073  | -2.585747 | -4.55     | -2.232961     |
| hsa-miR-200a-3p| 0.863     | 0.03114  | -2.577578 | -4.55     | -2.279405     |
| hsa-miR-34c-3p | 0.863     | 0.01521  | -3.028324 | -4.54     | -2.305722     |
| hsa-miR-378d   | 0.863     | 0.03592  | -2.488453 | -4.56     | -2.354464     |
| hsa-miR-142-3p | 0.863     | 0.03736  | -2.463817 | -4.56     | -2.488274     |
| hsa-miR-4524b-3p| 0.863   | 0.02313  | -3.116108 | -4.58     | -2.565208     |
| hsa-miR-3653-3p| 0.863     | 0.01347  | -3.105707 | -4.54     | -2.832038     |
| hsa-miR-320c   | 0.863     | 0.00898  | -3.367253 | -4.54     | -2.926313     |
| hsa-miR-375    | 0.863     | 0.00836  | -3.431574 | -4.54     | -3.17402      |
| hsa-miR-4539   | 0.863     | 0.02251  | -2.789029 | -4.55     | -3.278709     |
| hsa-miR-215-5p | 0.863     | 0.00106  | -3.293314 | -4.54     | -3.955771     |

3.2. Identification of Differentially Expressed MicroRNA. The six samples were divided into two groups (cancer and normal tissue group), and the differentially expressed miRNA analysis was conducted by GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE101502). P values <0.05, |LogFC| > 2 were considered as differentially expressed microRNA. LogFC > 2 was upregulated microRNA, LogFC < -2 was down-regulated. Table 2 showed the identification of differentially expressed miRNA in the two groups.

3.3. Prediction of Target Genes. miRNA enrichment was used to predict potential target genes from differentially expressed miRNA. The up- and downregulated microRNA were inputted into the FunRich software tool, respectively. There were 875 up- and 2920 downregulated target genes found.

3.4. GO Function and KEGG Pathway Enrichment Analysis. All target genes were imported into the online analysis tool, DAVID, to identify potential GO categories and KEGG pathways. GO analysis results revealed that upregulated target genes were expressively enriched in biological processes (BP), including regulation of cell communication, regulation of signaling, and nervous system development; in molecular function (MF) including receptor signaling protein activity, transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding, and enzyme binding; and in cell component (CC) including cell junction, cell leading edge, and adherens junction (Table 3). The downregulated target genes were expressively enriched in BP, including nervous system development, regulation of transcription from RNA polymerase II promoter, and positive regulation of RNA metabolic process; in MF including RNA polymerase II transcription factor activity, sequence-specific DNA binding, regulatory region nucleic acid binding, and regulatory region DNA binding; in CC, including nucleoplasm, neuron projection, and neuron part (Table 3). KEGG pathway analysis showed that the upregulated target genes were enriched in axon guidance, MAPK signaling pathway, endocytosis, proteoglycans in cancer, and the FoxO signaling pathway, while the downregulated target genes were enriched in pathways in cancer, focal adhesion, PI3K-Akt signaling pathway, small cell lung cancer, and signaling pathways regulating pluripotency of stem cells. Table 4 shows the most significantly enriched pathways of the upregulated target genes and downregulated target genes were performed by KEGG analysis.

3.5. Module Screening and Hub Gene Selecting from the Protein-Protein Interaction (PPI) Network. All target genes were imported into the STRING database to conduct the PPI network. A combined score of > 0.4 of the nodes was considered as significance (Figure 1). Then, the results of the PPI network were exported as .txt and imported to Cytoscape software which was analyzed using plug-ins MCODE. Finally, the top four significant modules were selected and considered as hub genes. The 82 hub genes are illustrated in Figures 2(a), 2(c), 2(e), and 2(g). The functional annotations of those genes were analyzed by DAVID. Enrichment analysis indicated that the genes in modules 1 through 4 were mainly associated with endocytosis, spliceosome, TGF-beta signaling pathway, and lysosome (Figures 2(b), 2(d), 2(f), and 2(h)).

3.6. Survival Plots and Expression Level of Hub Genes. We used survival analysis by GEPIA (Gene Expression Profiling Interactive Analysis) to detect the overall survival of 82 hub genes between the high and low expression groups. It was found that high expressions of NUDT21 (HR=0.57, P=0.023) (Figure 1(a)), GNB1 (HR=0.028, P=0.026) (Figure 1(c)), and CLINT1 (HR=0.6, P=0.043) (Figure 1(d)) were associated
### Table 3: GO analysis of differentially expressed genes associated with IIA stage colon cancer.

| Regulation | Category | Term                                                                 | Count | %     | P-Value | Fold Enrichment | FDR    |
|------------|----------|----------------------------------------------------------------------|-------|-------|---------|-----------------|--------|
| up         | GOTERM\_BP\_FAT | GO:0001646~regulation of cell communication                         | 234   | 24.92 | 2.36E-12 | 1.526744778    | 4.63E-09|
| up         | GOTERM\_BP\_FAT | GO:0023051~regulation of signaling                                    | 236   | 25.13 | 4.03E-12 | 1.514715803    | 7.91E-09|
| up         | GOTERM\_BP\_FAT | GO:0007399~nervous system development                                 | 180   | 19.17 | 8.89E-12 | 1.634444496    | 1.75E-08|
| up         | GOTERM\_BP\_FAT | GO:00010604~positive regulation of macromolecule metabolic process  | 220   | 23.43 | 2.72E-11 | 1.517826065    | 5.34E-08|
| up         | GOTERM\_BP\_FAT | GO:007167~enzyme linked receptor protein signaling pathway            | 98    | 10.44 | 2.73E-11 | 2.02415217     | 5.37E-08|
| up         | GOTERM\_MF\_FAT | GO:0005057~receptor signaling protein activity                        | 27    | 2.88  | 2.69E-08 | 3.580476019    | 4.39E-05|
| up         | GOTERM\_MF\_FAT | GO:0000982~transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding | 41    | 4.37  | 1.00E-06 | 2.33015106     | 0.001631|
| up         | GOTERM\_MF\_FAT | GO:0009899~enzyme binding                                              | 136   | 14.48 | 1.81E-06 | 1.481913682    | 0.002953|
| up         | GOTERM\_MF\_FAT | GO:0000092~cytoskeletal protein binding                               | 75    | 7.98  | 3.96E-06 | 1.728165319    | 0.006455|
| up         | GOTERM\_MF\_FAT | GO:0019904~protein domain specific binding                             | 59    | 6.28  | 4.69E-06 | 1.873173393    | 0.00764 |
| up         | GOTERM\_CC\_FAT | GO:0030054~cell junction                                              | 123   | 13.09 | 2.20E-09 | 1.71673403     | 3.35E-06|
| up         | GOTERM\_CC\_FAT | GO:0031252~cell leading edge                                           | 44    | 4.68  | 2.05E-07 | 2.378506707    | 3.12E-04|
| up         | GOTERM\_CC\_FAT | GO:0005912~adherens junction                                          | 69    | 7.34  | 3.11E-07 | 1.907961809    | 4.74E-04|
| up         | GOTERM\_CC\_FAT | GO:0070161~anchoring junction                                         | 69    | 7.34  | 7.57E-07 | 1.862342469    | 0.001153|
| down       | GOTERM\_BP\_FAT | GO:0007399~nervous system development                                 | 546   | 18.66 | 3.19E-36 | 1.619387       | 6.50E-33|
| down       | GOTERM\_BP\_FAT | GO:0006357~regulation of transcription from RNA polymerase II promoter | 468   | 16.16 | 4.34E-32 | 1.64145        | 8.83E-29|
| down       | GOTERM\_BP\_FAT | GO:0051254~positive regulation of RNA metabolic process               | 375   | 12.95 | 7.66E-29 | 1.705661       | 1.56E-25|
| down       | GOTERM\_BP\_FAT | GO:0006366~transcription from RNA polymerase II promoter              | 453   | 15.65 | 1.98E-28 | 1.601934       | 4.03E-25|
| down       | GOTERM\_BP\_FAT | GO:1903508~positive regulation of nucleic acid-templated transcription | 360   | 12.43 | 2.90E-28 | 1.717546       | 5.91E-25|
| down       | GOTERM\_MF\_FAT | GO:000988~RNA polymerase II transcription factor activity, sequence-specific DNA binding | 188   | 6.49  | 3.54E-17 | 1.799249       | 6.12E-14|
| regulation | Category     | Term                                                                 | Count | %       | P-Value   | Fold Enrichment | FDR     |
|------------|--------------|----------------------------------------------------------------------|-------|---------|-----------|-----------------|---------|
| down       | G:0001067    | regulatory region nucleic acid binding                               | 225   | 7.772021 | 7.28E-16  | 1.663386        | 1.34E-12|
| down       | G:0000975    | regulatory region DNA binding                                        | 224   | 7.737478 | 1.35E-15  | 1.657928        | 2.31E-12|
| down       | G:0044212    | transcription regulatory region DNA binding                          | 223   | 7.702936 | 1.78E-15  | 1.656331        | 3.08E-12|
| down       | G:0043565    | sequence-specific DNA binding                                        | 267   | 9.222798 | 3.59E-15  | 1.566314        | 6.15E-12|
| down       | G:0005654    | nucleoplasm                                                          | 632   | 21.83074 | 2.75E-18  | 1.331786        | 4.42E-15|
| down       | G:0043005    | neuron projection                                                    | 233   | 8.048359 | 5.31E-12  | 1.522783        | 8.54E-09|
| down       | G:0097458    | neuron part                                                          | 297   | 10.23907 | 8.60E-11  | 1.406649        | 1.38E-07|
| down       | G:0098644    | complex of collagen trimers                                          | 18    | 0.621762 | 9.04E-10  | 4.940876        | 1.45E-06|
| Regulated | Category      | Term                           | Count | %      | P-Value  | Fold Enrichment | FDR       |
|-----------|---------------|--------------------------------|-------|--------|-----------|-----------------|-----------|
| Up        | KEGG:PATHWAY  | hsa04360:Axon guidance         | 21    | 2.24   | 7.30E-06 | 3.173885        | 0.009462  |
| Up        | KEGG:PATHWAY  | hsa04010:MAPK signaling pathway| 31    | 3.30   | 1.94E-05 | 2.333442        | 0.025073  |
| Up        | KEGG:PATHWAY  | hsa04144:Endocytosis           | 31    | 3.30   | 2.43E-05 | 2.306309        | 0.031476  |
| Up        | KEGG:PATHWAY  | hsa05205:Proteoglycans in cancer| 26    | 2.77   | 3.53E-05 | 2.495278        | 0.045696  |
| Up        | KEGG:PATHWAY  | hsa04068:FoxO signaling pathway| 20    | 2.13   | 5.61E-05 | 2.864842        | 0.072721  |
| Down      | KEGG:PATHWAY  | hsa05200:Pathways in cancer    | 108   | 3.73   | 1.58E-10 | 1.810230026     | 2.10E-07  |
| Down      | KEGG:PATHWAY  | hsa04510:Focal adhesion         | 66    | 2.28   | 1.44E-09 | 2.110470443     | 1.90E-06  |
| Down      | KEGG:PATHWAY  | hsa04151:PI3K-Akt signaling pathway| 91    | 3.14   | 4.76E-08 | 1.737500173     | 6.31E-05  |
| Down      | KEGG:PATHWAY  | hsa05222:Small cell lung cancer| 32    | 1.11   | 9.87E-07 | 2.479896821     | 0.001306893|
| Down      | KEGG:PATHWAY  | hsa04550:Signaling pathways regulating pluripotency of stem cells | 44    | 1.52   | 2.04E-06 | 2.070271006    | 0.002694908|
with better overall survival for colon cancer patients. However, a high expression of COL1A2 (HR 1.8, P = 0.017) (Figure 1(b)) was associated with worse overall survival for colon cancer patients (Figure 3), and there was no statistical significance in the other 78 hub genes. Taken together, NUDT21, GNB1, CLINT1, and COL1A2 were considered as core genes with a close relationship to colon cancer. Then, we used GEPIA analysis to explore the core genes’ expression level between colon cancer and normal tissue (Figures 4(a), 4(b), 4(c), and 4(d)).

### 4. Discussion

Pathogenesis of colon cancer is association with gene mutation, epigenetics, and the CpG island methylator phenotype [23]. In order to diagnose this disease early to precisely and effectively treat colon cancer, understanding the molecular mechanism is imperative. Microarray and high-throughput next generation sequencing have been widely utilized in order to predict the potential therapeutic targets gene of colon cancer as both techniques could provide expression levels for thousands of genes. miRNA regulates the progression of the tumor by regulating these target genes, and some miRNAs have been identified as being involved in several types of cancer [24, 25]. Therefore, it is of great significance to study the expression profile of miRNAs and predict the target genes in colon cancer. In this study, we extracted the data from GSE101502 and identified two upregulated and 15 downregulated differential expressed microRNAs between colon cancer tissue and adjacent normal mucosa tissue using bioinformatics analysis [26, 27]. And we found that NUDT21, GNB1, CLINT1, and COL1A2 might be the potential core genes that play an important role in the development and prognosis in IIA stage colon cancer.

| Term                        | Count | P-Value  | Fold Enrichment | FDR   | Genes                                      |
|-----------------------------|-------|----------|-----------------|-------|--------------------------------------------|
| Endocytosis                 | 9     | 1.03E-08 | 16.10573123     | 9.27E-06 | EPS15, AP2B1, DAB2, TFRC, RAB5C, ARRB1, NEDD4L, ARPC5, EPN1 |
| Ubiquitin mediated proteolysis | 3     | 0.031312 | 9.987132353     | 24.84759 | CUL5, UBA6, NEDD4L                        |
| Spliceosome                 | 5     | 4.61E-06 | 32.07720588     | 0.00145213 | AQR, DDX46, CRNK1L, U2AF2, DHX15          |
| mRNA surveillance pathway   | 2     | 0.0896571| 18.76344086     | 25.5986930 | CSTF3, NUDT21                             |
| TGF-beta signaling pathway  | 9     | 4.18E-09 | 21.4207         | 4.43E-06 | INHBB, ACVR2A, ACVR1B, SMAD5, SMAD4, SMAD3, SMAD2, SMAD1, TGF2 |
| Signaling pathways regulating pluripotency of stem cells | 8     | 1.04E-06 | 13.76783        | 0.001104 | INHBB, ACVR2A, ACVR1B, SMAD5, SMAD4, SMAD3, SMAD2, SMAD1 |
| Cell cycle                  | 6     | 2.08E-04 | 10.40592        | 0.220111 | CCND2, SMAD4, SMAD3, CDK6, SMAD2, TGF2     |
| Lysosome                    | 4     | 4.48E-06 | 60.18254        | 4.48E-04 | AP1S1, AP4E1, AP1G1, AP3S1                 |

Figure 2: Top 4 modules from the PPI network. (a, c, e, g) modules 1 to 4; (b, d, f, h) the enriched pathway of modules 1 to 4.
Figure 3: Prognostic values of the four genes (NUDT21, GNB1, CLINT1, and COL1A2) in colon cancer patients. Overall survival analysis over time with high vs. low NUDT2 expression (a), low vs. high COL1A2 expression (b), high vs. low GNB1 expression (c), and low vs. high CLINT1 expression (d). The P value was determined by log-rank test between risk groups.

In this study, the GO analysis showed that these potentially upregulated genes were mainly enriched in the regulation of cell communication, receptor signaling protein activity, and cell junction. Potential downregulated genes were involved in nervous system development, RNA polymerase II transcription factor activity, sequence-specific DNA binding, and nucleoplasm. Pinto et al. [28] indicated that there is a complicated cell communication in response to ionizing radiation revealed by primary human macrophage-cancer cell culture. Kim et al. [29] reported that IFITM1 expression was positively correlated with galectin-3 via receptor signaling protein activity in human colon cancer cells. The cell junctions might lead to cancer due to the differences in cell junctions for colorectal cancer [30]. Moreover, nervous system development also plays a key role in colorectal cancer metastasis [31, 32]. RNA polymerase II transcription factor contains sequence-specific DNA binding, transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins [33, 34]. Between the nucleoplasm and cytoplasm called perinuclear, the signal transmission becomes abnormal by the perinucleus in malignant cell transformation [35]. All those studies indicated that the molecular functions of those up- and downregulated genes are related to colon cancer.

Moreover, the KEGG pathways for upregulated genes were enriched in axon guidance, MAPK signaling pathway, endocytosis, proteoglycans in cancer, and FoxO signaling pathway. Downregulated genes were involved in the pathways in cancer, focal adhesion, PI3K-Akt signaling pathway, small cell lung cancer, and regulation of signaling pathways in the pluripotency of stem cells. The axon guidance indicates that netrin I and Slits are causally involved in human
The ERK MAPK (extracellular-signal-regulated kinases) is one of the subfamilies of MAPK (mitogen-activated protein kinases), and it has been found that overexpression and activation of ERK MAPK play a role in the progression of colorectal cancer [37]. Through PIP2 mediated vinculin activation, PIPKIγ might positively regulate focal adhesion dynamics and colon cancer cell invasion [38]. The PI3K/AKT pathway plays an important role in the prognostic and predictive values in colorectal cancer [39]. Evidence suggests that endocytosis, proteoglycans in cancer, FoxO signaling pathway, and regulation of signaling pathways in the pluripotency of stem cells are all associated with colorectal cancer [40–42].

Finally, NUDT21, COL1A2, GNBI, and CLINT1 closely related to the overall survival of colon cancer were selected as core genes. COL1A2 is collagen type I alpha 2 chain, the fibrillar collagen detected in most connective tissues. This observation suggests that patients with a high expression of COL1A2 have a worse prognosis. Pekow et al. indicated that downregulating miR-4728-3p reduces ulcerative colitis associated colon cancers, and miR-4728-3p is a regulator of COL1A2 [43]. NUDT21 is Nudix hydrolase 21, belonging to the Nudix family of hydrolases. GNBI is G protein subunit beta 1. Wazir et al. researched on 136 human breast cancer tissues and 31 normal tissues, undertook reverse transcription and quantitative polymerase chain reaction, and suggested...
that GNB1 plays an important character in the mTOR-related antiapoptosis pathway and might potentially be targeted in breast cancer [44]. CLINT1 is Clathrin interactor 1. Ajiro et al. [45] indicated that SRSF3 regulates a lot of genes including CLINT1 affecting gene expression to keep cell homeostasis. Moreover, further deeply investigated molecular mechanism of NUDT21, COL1A2, GNB1, CLINT1, and colon cancer is necessary; it is also the limitation of this study.

5. Conclusion

In conclusion, this study showed that NUDT21, GNB1, CLINT1, and COL1A2 might be the potential core genes that play an important role in the development and prognosis in IIA stage colon cancer. After discovering this, we have come to the conclusion that a series of experiments and further deeply investigated molecular mechanism of those four core genes should be designed to confirm the results of this study.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

All the authors have no conflicts of interest.

Authors’ Contributions

Zhiyong Dong and Wei Lin contributed equally to this work.

Acknowledgments

We would like to thank Jun Bai (Feinberg School of Medicine, Northwestern University) for the help in reviewing and editing this paper.

References

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, and J. Lortet-Tieulent, “Global cancer statistics, 2012,” CA: A Cancer Journal for Clinicians, vol. 65, no. 2, pp. 87–108, 2015.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer Statistics, 2017,” CA: A Cancer Journal for Clinicians, vol. 67, no. 1, pp. 7–30, 2017.

[3] M. C. Merlano, C. Granetto, E. Fea, V. Ricci, and O. Garrone, “Heterogeneity of colon cancer: from bench to bedside,” ESMO Open, vol. 2, no. 3, Article ID e000218, 2017.

[4] L.-A. MacFarlane and P. R. Murphy, “MicroRNA: biogenesis, function and role in cancer,” Current Genomics, vol. 11, no. 7, pp. 537–561, 2010.

[5] T. A. Farazi, S. A. Juranek, and T. Tuschi, “The growing catalog of small RNAs and their association with distinct Argonauta/Piwi family members,” Development, vol. 135, no. 7, pp. 1201–1214, 2008.

[6] A. Nakamura, T. Osonoi, and Y. Terauchi, “Relationship between urinary sodium excretion and pioglitazone-induced edema,” Journal of Diabetes Investigation, vol. 1, no. 5, pp. 208–211, 2010.

[7] L. Ruan, Y. Xie, F. Liu, and X. Chen, “Serum mir-124 and mir-134 associated with ovarian cancer: MiRNA microarray data analysis for a pilot study,” European Journal of Obstetrics & Gynecology and Reproductive Biology, vol. 222, pp. 31–38, 2018.

[8] Y. Zhang, X. Guo, L. Xiong et al., “Comprehensive analysis of microRNA-regulated protein interaction network reveals the tumor suppressive role of microRNA-149 in human hepatocellular carcinoma via targeting AKT-mTOR pathway,” Molecular Cancer, vol. 13, no. 1, article 253, 2014.

[9] Y. Ke, W. Zhao, J. Xiong, and R. Cao, “miR-149 Inhibits Non-Small-Cell Lung Cancer Cells EMT by Targeting FOXMi,” Biochemistry Research International, vol. 2013, Article ID 506731, 8 pages, 2013.

[10] H. Yang, L. Wang, X. Tang, and W. Bai, “miR-203a suppresses cell proliferation by targeting E2F transcription factor 3 in human gastric cancer,” Oncology Letters, vol. 14, no. 6, pp. 7687–7690, 2017.

[11] Y. Liu, X. Chen, R. Cheng et al., “The Jun/miR-22/HuR regulatory axis contributes to tumourigenesis in colorectal cancer,” Molecular Cancer, vol. 17, no. 1, 2018.

[12] J. Lasco, Z. Hemminki, and A. Försti, “Systematic enrichment analysis of gene expression profiling studies identifies consensus pathways implicated in colorectal cancer development,” Journal of Carcinogenesis, vol. 10, no. 1, p. 7, 2011.

[13] M. Rokavec, D. Horst, and H. Hermeking, “Cellular model of colon cancer progression reveals signatures of miRNAs, miRNA, IncRNAs, and epigenetic modifications associated with metastasis,” Cancer Research, vol. 77, no. 8, pp. 1854–1867, 2017.

[14] Y. Lu, X. Zhao, Q. Liu et al., “IncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/B-catenin signaling,” Nature Medicine, vol. 23, no. 11, pp. 1331–1341, 2017.

[15] E. Li, P. Ji, N. Ouyang et al., “Differential expression of miRNAs in colon cancer between African and Caucasian Americans: Implications for cancer racial health disparities,” International Journal of Oncology, vol. 45, no. 2, pp. 587–594, 2014.

[16] T. Barrett, S. E. Wilhite, P. Ledoux et al., “NCBI GEO: archive for functional genomics data sets—update,” Nucleic Acids Research, vol. 41, no. 1, pp. D991–D995, 2013.

[17] M. Pathan, S. Keerthikumar, D. Chisanga et al., “A novel community driven software for functional enrichment analysis of extracellular vesicles data,” Journal of Extracellular Vesicles (JEV), vol. 6, no. 1, Article ID 1321455, 2017.

[18] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources,” Nature Protocols, vol. 4, no. 1, pp. 44–57, 2009.

[19] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” Nucleic Acids Research, vol. 37, no. 1, pp. 1–13, 2009.

[20] D. Szklarczyk, A. Franceschini, S. Wyder et al., “STRING v10: protein-protein interaction networks, integrated over the tree of life,” Nucleic Acids Research, vol. 43, pp. D447–D452, 2015.

[21] P. Shannon, A. Markiel, O. Ozier et al., “Cytoscape: a software Environment for integrated models of biomolecular interaction networks,” Genome Research, vol. 13, no. 11, pp. 2498–2504, 2003.

[22] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, “GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses,” Nucleic Acids Research, vol. 45, no. 1, pp. W98–W102, 2017.
[23] M. Toyota, N. Ahuja, M. Ohe-Toyota, J. G. Herman, S. B. Baylin, and J. J. Issa, "CpG island methylator phenotype in colorectal cancer," Proceedings of the National Academy of Sciences of the United States of America, vol. 96, no. 15, pp. 8681–8686, 1999.

[24] A. Lujambio and S. W. Lowe, "The microcosmos of cancer," Nature, vol. 482, no. 7385, pp. 347–355, 2012.

[25] Y. W. Kong, D. Ferland-McCollough, T. J. Jackson, and M. Bushell, "microRNAs in cancer management," The Lancet Oncology, vol. 13, no. 6, pp. e249–e258, 2012.

[26] R. Apweiler, A. Bairoch, C. H. Wu et al., "UniProt: the universal protein knowledgebase," Nucleic Acids Research, vol. 32, pp. D115–D119, 2004.

[27] M. Pathan, S. Keerthikumar, C.-S. Ang et al., "FunRich: An open access standalone functional enrichment and interaction network analysis tool," Proteomics, vol. 15, no. 15, pp. 2597–2601, 2015.

[28] A. T. Pinto, M. L. Pinto, S. Velho et al., "Intricate Macrophage-Colorectal Cancer Cell Communication in Response to Radiation," PLoS ONE, vol. 11, no. 8, Article ID e0160891, 2016.

[29] K. Kim, Y. Kwon, C. Cho et al., "Galectin-3-independent Down-regulation of GABABR1 due to Treatment with Korean Herbal Extract HAD-B Reduces Proliferation of Human Colon Cancer Cells," Journal of Pharmacopuncture, vol. 15, no. 3, pp. 19–30, 2012.

[30] M. Shirdel, Differences in Cell Functions of Colorectal Cancer with a Pushing or Infiltrative Growth Pattern: The Role of Gap Junction Protein beta 3 (Dissertation), 2013, http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-85608.

[31] S. Li, Y. Sun, and D. Gao, "Role of the nervous system in cancer metastasis," Oncology Letters, vol. 5, no. 4, pp. 1101–1111, 2013.

[32] S. Mondaca, V. Hornig, P. Munoz-Schuffenegger, F. Acevedo, M. Garrido, and B. Nervi, "Central nervous system metastasis secondary to colorectal cancer: A retrospective cohort study of 20 cases," Ecamcancermedicalscience, vol. 10, no. 705, 2016.

[33] C. S. Parker and J. Topol, "A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity," Cell, vol. 36, no. 2, pp. 357–369, 1984.

[34] P. J. Mitchell and R. Tjian, "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins," Science, vol. 245, no. 4916, pp. 371–378, 1989.

[35] T. E. Shaiken and A. R. Opekun, "Dissecting the cell to nucleus, perinucleus and cytosol," Scientific Reports, vol. 4, article 4923, 2014.

[36] P. Mehlen, C. Delloye-Bourgeois, and A. Chédotal, "Novel roles for Slits and netrins: Axon guidance cues as anticancer targets?" Nature Reviews Cancer, vol. 11, no. 3, pp. 188–197, 2011.

[37] J. Y. Fang and B. C. Richardson, "The MAPK signalling pathways and colorectal cancer," The Lancet Oncology, vol. 6, no. 5, pp. 322–327, 2005.

[38] Z. Wu, X. Li, M. Sunkara, H. Spearman, A. J. Morris, and C. Huang, "Pipki regulates focal adhesion dynamics and colon cancer cell invasion," PLoS ONE, vol. 6, no. 9, Article ID e24775, 2011.

[39] S. A. Danielsen, P. W. Eide, A. Nesbakken, T. Guren, E. Leithe, and R. A. Lothe, "Portrait of the PI3K/AKT pathway in colorectal cancer," Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, vol. 1855, no. 1, pp. 104–121, 2015.

[40] I. M. Mullman and Y. Yarden, "Endocytosis and cancer," Cold Spring Harbor Perspectives in Biology, vol. 5, no. 12, Article ID a016949, 2013.

[41] S. F. De Mattos, P. Villalonga, J. Clardy, and E. W.-F. Lam, "FOXO3a mediates the cytotoxic effects of cisplatin in colon cancer cells," Molecular Cancer Therapeutics, vol. 7, no. 10, pp. 3237–3246, 2008.

[42] D. Dutta, "Signaling pathways dictating pluripotency in embryonic stem cells," The International Journal of Developmental Biology, vol. 57, no. 9-10, pp. 667–675, 2013.

[43] J. Pekow, A. L. Hutchison, K. Meckel et al., "MIR-4728-3p Functions as a Tumor Suppressor in Ulcerative Colitis-associated Colorectal Neoplasia Through Regulation of Focal Adhesion Signaling," Inflammatory Bowel Diseases, vol. 23, no. 8, pp. 1328–1337, 2017.

[44] U. Wazir, W. G. Jiang, A. K. Sharma, and K. Mokbel, "Guanine nucleotide binding protein β 1: A novel transduction protein with a possible role in human Breast cancer," Cancer Genomics & Proteomics, vol. 10, no. 2, pp. 69–74, 2013.

[45] M. Ajiro, R. Jia, Y. Yang, J. Zhu, and Z.-M. Zheng, "A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells," Nucleic Acids Research, vol. 44, no. 4, pp. 1854–1870, 2015.