Comprehensive analysis of long non-coding RNA and mRNA expression profile in rectal cancer

De-Zhong Wang[1], Guan-Yang Chen[2], Yi-Feng Li[3], Neng-Wei Zhang[1]

[1]Department of Gastroenterology, Liver and Gallbladder Surgery, Peking University Ninth School of Clinical Medicine, Beijing 100038, China;
[2]The Clinical Center of Gene and Cell Engineering, Beijing Shijitan Hospital, Peking University Ninth School of Clinical Medicine, Beijing 100038, China;
[3]Department of General Surgery, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China.

Abstract

Background: Rectal cancer (RC) is a malignant tumor that seriously threatens human health. Long non-coding RNAs (lncRNAs) play a vital role in tumor regulation. Nevertheless, their exact expression features and functions remain obscure, and therefore was the aim of the current study.

Methods: We utilized the Affymetrix human GeneChip to screen differentially expressed profiles of lncRNAs and mRNAs from the cancer tissues and matched paracancer tissues of 6 RC patients. Gene Ontology (GO) and pathway enrichment analyses identified crucial functions and pathways of the aberrantly expressed mRNAs. We used quantitative real-time polymerase chain reaction to verify the significant expression differences of 11 candidate lncRNAs between the cancer and paracancer tissues. LncRNA-mRNA coexpression networks were built by calculating the Pearson correlation value to identify significant correlation pairs. Online bioinformatics tools GEPIA2, ONCOMINE, and PROGgeneV2 were used to mine the expression and prognosis of three crucial mRNAs and six verified lncRNAs. Competing endogenous RNA networks were constructed by predicting microRNA response elements and calculating free energy.

Results: We found 1658 differentially expressed lncRNAs (778 up-regulated and 880 down-regulated) and 1783 aberrantly expressed mRNAs (909 up-regulated and 874 down-regulated). GO and pathway enrichment analyses revealed the vital functions of the differentially expressed mRNAs, including cell proliferation, cell migration, angiogenesis, and cellular response to zinc ion. The canonical signaling pathways mainly included the interleukin-17, cell cycle, Wnt, and mineral absorption signaling pathways. Six lncRNAs including AC017002.2 ($P = 0.039$), cancer susceptibility 19 (CASC19) ($P = 0.021$), LINC00152 ($P = 0.013$), NONHSAT058834 ($P = 0.007$), NONHSAT007692 ($P = 0.045$), and ENST00000415991.1 ($P = 0.045$) showed significant differences in expression levels between the cancer tissue and paracancer tissue groups. AC017002.2, NONHSAT058834, NONHSAT007692, and ENST00000415991.1 have not yet been reported in RC. The crucial mRNAs myelocytomatosis viral oncogene (MYC), transforming growth factor beta induced (TGFBI), and solute carrier family 7 member 5 (SLC7A5) were selected. AC017002.2 and LINC00152 were positively correlated with MYC, TGFBI, and cytochrome P450 family 2 sub-family B member 6 (All $r > 0.900$, $P < 0.050$). NONHSAT058834 was positively associated with MYC ($r = 0.930$, $P < 0.001$), and CASC19 was positively correlated with SLC7A5 ($r = 0.922$, $P < 0.001$).

Conclusion: This study offers convincing evidence of differentially expressed lncRNAs and mRNAs as potential biomarkers in RC.

Keywords: lncRNA; mRNA; Expression profile; Rectal cancer

Introduction

Colorectal cancer (CRC) is a common malignant tumor that seriously threatens human health. Worldwide, CRC ranks third in terms of incidence and second in terms of mortality.[1] Among all tumor types in China in 2018, the age-standardized incidence of CRC ranked second, and the age-standardized death rate ranked fifth.[2] As a type of CRC, rectal cancer (RC) is also a challenge. Although we have made significant progress in the diagnosis and treatment of RC, the effective treatment of this refractory disease still requires exploration at the molecular pathophysiological level.

Recently, gene sequencing has become an effective approach to explore the molecular mechanisms underlying the initiation, progression, and prognosis of tumors. Long non-coding RNAs (lncRNAs) are RNAs whose transcripts

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000000753

De-Zhong Wang and Guan-Yang Chen contributed equally to this work.

Correspondence to: Prof. Neng-Wei Zhang, Department of Gastroenterology, Liver and Gallbladder Surgery, Peking University Ninth School of Clinical Medicine, Beijing 100038, China
E-Mail: zhangnw1@sohu.com

Chinese Medical Journal 2020;133(11)

Received: 26-12-2019 **Edited by:** Qiang Shi

are longer than 200 nucleotides in length and lack the ability to encode proteins due to the lack of specific open reading frames.[3] At present, many studies have shown that the dysregulation of lncRNAs plays a critical role in regulating biological processes (BPs) in CRC, including tumorigenesis progression and metastasis. For example, lncRNA SET binding factor 2 antisense RNA 1, as a miR-619-5p sponge, regulates the proliferation, migration, and invasion of CRC.[4] The lncRNAs, colon cancer associated transcript 1 and colon cancer associated transcript 2, are associated with tumor recurrence and poor prognosis, and the expression of these two lncRNAs may provide actionable biomarkers for predicting tumor recurrence or prognosis in CRC patients.[5]

However, the effects of aberrantly expressed lncRNAs on the molecular pathologic mechanism of RC are still unclear. Moreover, few studies have examined the regulation of aberrantly expressed lncRNAs on the BPs of RC. Therefore, we described the profile of differentially expressed lncRNAs in RC, analyzing and predicting the correlated functions and mechanisms of some vital lncRNAs and mRNAs.

Methods

Ethical approval

All patients provided written informed consent. This study was approved by the Medical Ethics Committee of Beijing Shijitan Hospital, Capital Medical University (Scientific Research Ethics approval No. 2019-50) and was conducted in strict compliance with the *Declaration of Helsinki*.

Patients and tissue specimens

We selected 16 patients who were pathologically diagnosed with RC and underwent surgical treatment between October 2018 and March 2019 at the Beijing Shijitan Hospital, Capital Medical University. Sixteen pairs of specimens of cancer tissues and paracancer tissues were obtained from these patients [Table 1]. Six of them were used for microarray analysis, and ten of them were used for quantitative real-time polymerase chain reaction (qRT-PCR). In the operating room, dissected cancer tissues and paracancer tissues approximately 2 to 5 cm away from the edge of the tumor were collected, immediately frozen in liquid nitrogen and stored at −80°C. All patients had not received radiotherapy or chemotherapy before surgery.

Total RNA extraction and microarray analysis

Total RNA was extracted from the specimens using TRIzol reagent (Life Technologies, CA, USA) according to the manufacturer's instructions and then further purified with the RNeasy Mini Kit and RNase-Free DNase Set (Qiagen, Valencia, CA, USA). The purity and concentration of the RNA were detected with NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). According to the normal Affymetrix protocol, a total of 150 ng RNA was reverse transcribed to complementary DNA (cDNA). The fragmented cDNA was hybridized to the Affymetrix GeneChip Human Clariom® D Array. Then, the GeneChip was cleaned, stained, and scanned. Finally, Affymetrix analysis settings were used to assess the GeneChip data. The values shown are log2 robust multi-array average (RMA) signal intensity. $P < 0.05$ and fold change >1.2 were considered as the criteria to identify the differential expression of lncRNAs and mRNAs. The GeneChip data analyzed in this article are publicly available at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database under accession number GSE139814.

Gene Ontology (GO) enrichment analysis

GO is a structured and controlled vocabulary of terms. The terms are sub-divided into three non-overlapping ontologies, including molecular function, BP, and cellular

Table 1: Clinical and pathological data of 16 patients with rectal cancer.

Sample	Sex	Age (years)	Tumor location	TNM stage	Hepatic metastases
1	Female	82	Rectum	T4aN0M1a	Yes
2	Male	45	Rectum	T3N0M0	No
3	Male	66	Rectum	T4aN0M1a	Yes
4	Male	73	Rectum	T1N0M0	No
5	Male	72	Rectum	T3N0M0	No
6	Female	69	Rectum	T2N1M0	No
7	Male	71	Rectum	T2N0M0	No
8	Male	74	Rectum	T2N0M0	No
9	Male	66	Rectum	T2N0M0	No
10	Male	67	Rectum	T3N0M0	No
11	Male	49	Rectum	T2N0M0	No
12	Male	50	Rectum	T4aN0M0	No
13	Male	87	Rectum	T4aN2bM1a	Yes
14	Female	66	Rectum	T4aN0M1a	Yes
15	Male	61	Rectum	T2N0M0	No
16	Male	63	Rectum	T4aN2aM1a	Yes

T: Tumor; N: Node; M: Metastasis.

component, which are now widely used for annotating genes, gene products, and sequences.[6] The primary functions of the differentially expressed lncRNAs and mRNAs were analyzed by GO analysis. The differential expression of genes can be organized into hierarchical categories, which uncover the gene regulatory network based on BPs and molecular functions.

Pathway enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database of biological systems that integrates genomic, chemical, and systemic functional information.[7] At present, for the identification of significant pathways from large-scale genetic experimental data, such as GeneChip or next-generation sequencing data, which is considered an effective method to identify downstream targets, pathway enrichment is assessed by the hypergeometric distribution.

Real-time PCR

The standards we adopted to select candidate lncRNAs included the following: (1) The expression values of the lncRNAs were >1. (2) The fold changes in these genes in the experimental group were >2 compared with those in the control group. (3) The sizes of the lncRNAs were between 300 and 1200 base pairs. (4) The lncRNAs were intergenic. (5) They originated from the lncRNAWiki[8] and National Center for Biotechnology Information (NCBI) Reference Sequence Database (RefSeq).

For qRT-PCR, total RNA was isolated from cancer tissues and paracancer tissues using TRIzol reagent (Life Technologies) and then transcribed to cDNA with a reverse transcription kit (CoWin Biosciences, Beijing, China). The qRT-PCR was verified by SYBR Green reagent (CoWin Biosciences). Expression levels were measured using the $2^{-\Delta\Delta Ct}$ method, with the cycle threshold (Ct) values normalized. Glyceraldehyde-3-phosphate dehydrogenase and actin beta (verification stage) were used as internal controls.

Coexpression network (lncRNA-mRNA)

To evaluate the coexpression of genes, the Pearson correlation coefficient was calculated by the R function correlation test to obtain the coexpression coefficient of differential gene expression between the two groups. Combining the differences in lncRNA expression profile data and mRNA expression data, lncRNA-mRNA expression relationship networks with gene expression correlation coefficients were constructed. The relationship between lncRNAs and mRNAs might be clearly demonstrated in the expression network. In this way, the gene regulatory relationships of lncRNAs could be further indicated, and their possible functions could be explored. Generally, we chose 0.9 as the threshold of the correlation coefficient.

GEPIA2, ONCOMINE, and PROGgeneV2

Gene expression profiling interactive analysis 2 (GEPIA2)[9] is a updated version of GEPIA for analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) projects, using a standard processing pipeline. GEPIA2 was used to mine the expression and prognosis of AC017002.2, cancer susceptibility 19 (CASC19), and LINC00152 in RC.

ONCOMINE,[10] a cancer microarray database and web-based data-mining platform aimed at facilitating discovery from genome-wide expression analyses. ONCOMINE was used to mine the expression of myelocytomatosis viral oncogene (MYC), transforming growth factor beta induced (TGFBI), and solute carrier family 7 member 5 (SLC7A5) in CRC.

PROGgeneV2,[11] a tool that can be used to study prognostic implications of genes in various cancers. PROGgeneV2 was used to mine the prognosis of MYC, TGFBI, and SLC7A5 in CRC.

The competing endogenous RNA (ceRNA) network

LncRNAs can bind microRNAs (miRNAs) and act as ceRNAs; therefore, modulating the levels of the mRNAs targeted by the sponged miRNAs.[12] Specifically, lncRNAs manage miRNAs by sponging miRNA response elements (MREs) and act as modulators of miRNAs. By predicting MREs and calculating free energy, the competition correlation between lncRNAs and mRNAs can be assessed. In the field of tumors, these complicated lncRNA-miRNA-mRNA networks are broadly involved in regulating tumorigenesis.

Statistical analysis

The experimental results are shown as the mean ± standard deviation, and Statistical Product and Service Solutions (SPSS) 22.0 software (SPSS Inc., Chicago, IL, USA) was used for statistical analysis. In GO enrichment analysis, statistical analysis was conducted by two-sided Fisher exact test. In KEGG analysis, the significant pathways were evaluated by Fisher exact test. The significant differences in the qRT-PCR results were evaluated by Student's *t* test. In coexpression network, the Pearson correlation coefficient was calculated by the R function correlation test. A P value <0.05 was considered statistically significant.

Results

Overview of the lncRNA and mRNA profiles

We used hierarchical clustering analysis. In total, we found 1658 differentially expressed lncRNAs (778 up-regulated and 880 down-regulated) and 1783 aberrantly expressed mRNAs (909 up-regulated and 874 down-regulated) (fold change >1.2, $P < 0.050$). A heat map showed that the aberrantly expressed mRNAs [Figure 1A] and lncRNAs [Figure 1B] obviously distinguished cancer tissues from paracancer tissues.

GO analysis

GO analysis showed that the up-regulated mRNAs were related to functions mainly including cell proliferation, lung development, wound healing, chromosome segrega-

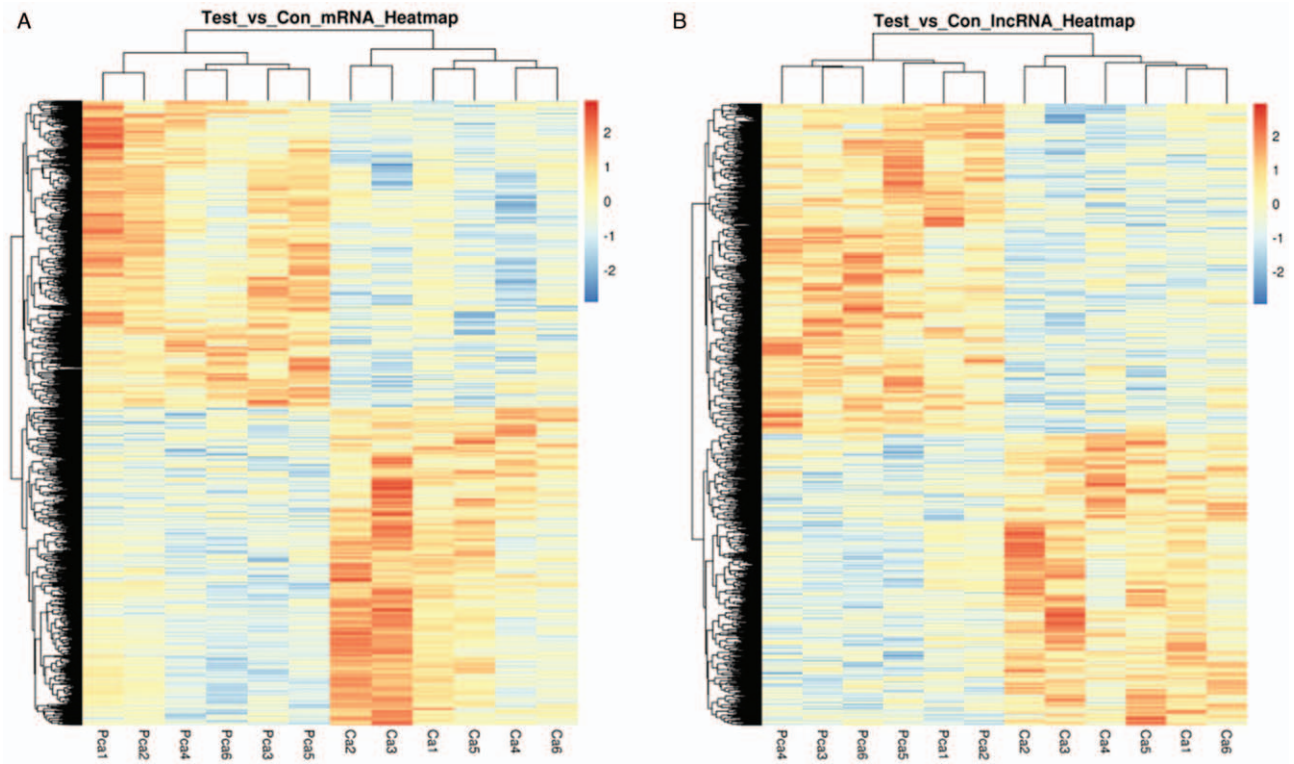

Figure 1: Comparison of the aberrant expression of mRNAs (A) and lncRNAs (B) in rectal cancer. A total of 1658 differentially expressed lncRNAs were found (fold change >1.2; $P < 0.05$), of which 778 were up-regulated and 880 were down-regulated. A total of 1783 aberrantly expressed mRNAs were found (fold change >1.2; $P < 0.05$), of which 909 were up-regulated and 874 were down-regulated. Differential mRNAs and lncRNAs of cancer tissues and paracancer tissues in a hierarchical clustering heatmap. Orange represents up-regulated genes, and blue represents down-regulated genes. Ca: Cancer; Con: Control; lncRNA: Long non-coding RNA; Pca: Paracancer.

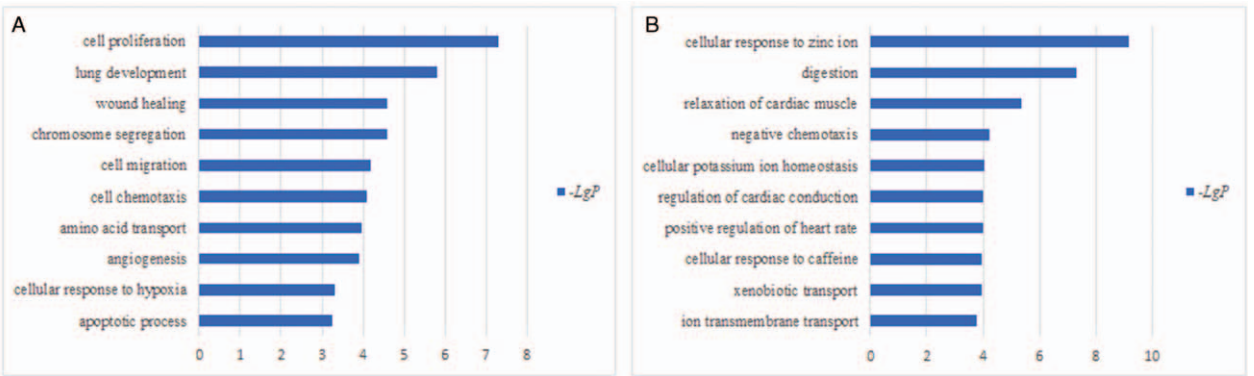

Figure 2: GO analysis of aberrantly expressed mRNAs. The GO analysis results of the up-regulated mRNAs (A) and down-regulated mRNAs (B). The statistical significance shown on the *X*-axis is indicated by −LgP. A larger –LgP represents a smaller *P* value. The *Y*-axis indicates the category of GO biological processes. GO: Gene ontology.

tion, cell migration, cell chemotaxis, amino acid transport, angiogenesis, cellular response to hypoxia, and apoptotic process [Figure 2A]. In contrast, the GO terms of the down-regulated mRNAs included cellular response to zinc ion, digestion, relaxation of cardiac muscle, cellular potassium ion homeostasis, and ion transmembrane transport [Figure 2B].

Pathway analysis

The KEGG database was used to determine the main related pathways of the aberrantly expressed genes. Regarding the related pathways of the up-regulated genes, there were ten relatively significant pathways [Figure 3A]. Among them, the interleukin (IL)-17 signaling pathway had the highest significance and the lowest false discovery rate (FDR) ($P < 0.001$, FDR = 0.001). In addition, some classic signaling pathways related to tumor BPs are also shown in the figure, including the cell cycle signaling pathway, Hippo signaling pathway, and Wnt signaling pathway. In contrast, the down-regulated gene pathways were mainly related to mineral absorption, bile secretion, the peroxisome proliferator activated receptor signaling pathway, insulin secretion, pancreatic secretion, and the

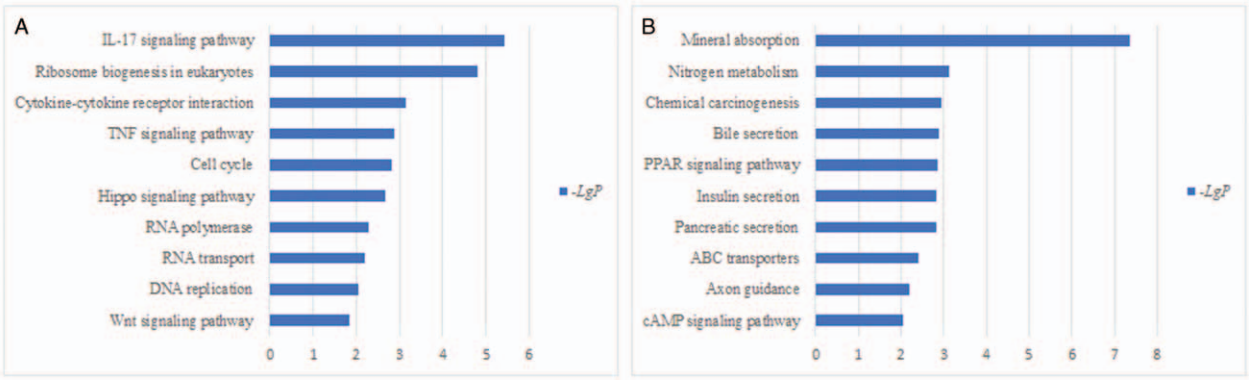

Figure 3: Pathway analysis of differentially expressed mRNAs. The pathway analysis results of the up-regulated mRNAs (A) and down-regulated mRNAs (B). The *Y*-axis represents the significantly modulated pathways, and the *X*-axis represents −LgP. A larger −LgP indicates a smaller *P* value. ABC: ATP binding cassette; cAMP: 3′-5′-Cyclic adenosine monophosphate; IL: Interleukin; PPAR: Peroxisome proliferator activated receptor; TNF: Tumor necrosis factor.

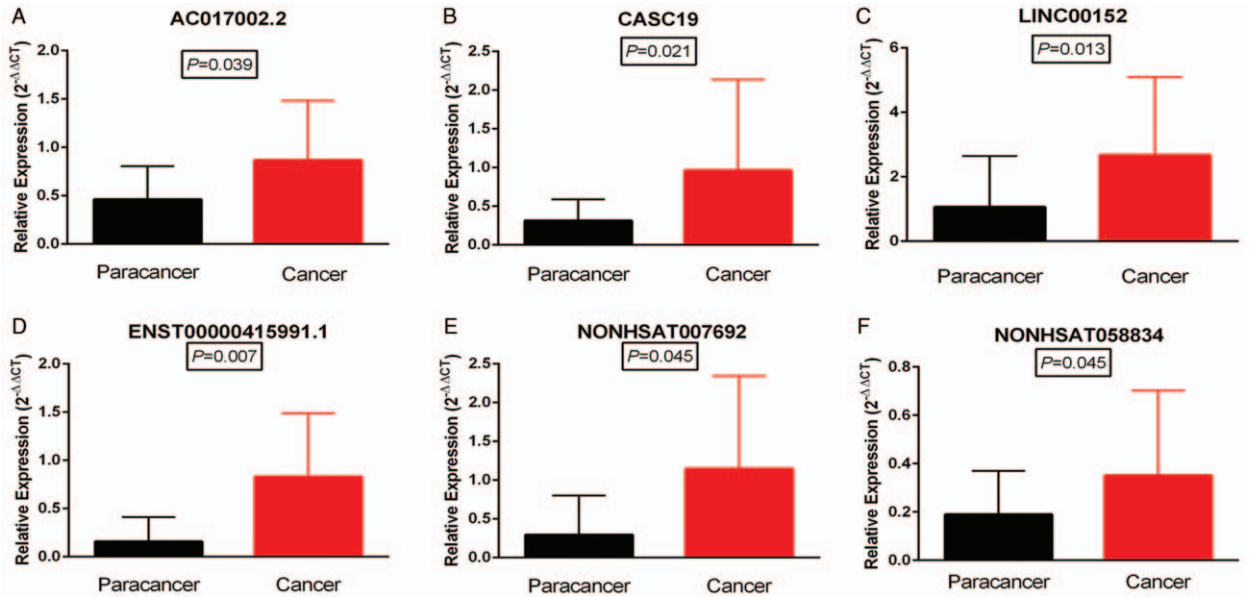

Figure 4: The expression of six lncRNAs verified by quantitative real-time polymerase chain reaction in rectal cancer tissues and matched paracancer tissues. (A) The expression of AC017002.2 in paracancer tissues and RC tissues (B) The expression of CASC19 in paracancer tissues and RC tissues. (C) The expression of LINC00152 in paracancer tissues and RC tissues. (D) The expression of ENST00000415991.1 in paracancer tissues and RC tissues. (E) The expression of NONHSAT007692 in paracancer tissues and RC tissues. (F) The expression of NONHSAT058834 in paracancer tissues and RC tissues. CASC19: Cancer Susceptibility 19; lncRNA: Long non-coding RNA; RC: Rectal cancer.

3′-5′-cyclic adenosine monophosphate signaling pathway [Figure 3B].

qRT-PCR validation

To further validate the results of the microarray analysis, we selected 11 differentially expressed lncRNAs as candidates based on the standards above. Then, ten pairs of patient tissue specimens were used to verify the differential expression of the 11 lncRNAs described above by qRT-PCR. Finally, six differentially expressed lncRNAs including AC017002.2 ($P = 0.039$), CASC19 ($P = 0.021$), LINC00152 ($P = 0.013$), NONHSAT058834 ($P = 0.007$), NONHSAT007692 ($P = 0.045$), and ENST00000415991.1 ($P = 0.045$) showed significant differences in expression levels between the cancer tissues and paracancer tissue groups [Figure 4].

Coexpression network (lncRNA-mRNA)

MYC, SLC7A5, and TGFBI were selected as the crucial mRNAs. They all ranked in the top 20 in fold changes and had a clear correlation with the validated lncRNAs. Coexpression networks of the six qPCR-verified lncRNAs and closely related mRNAs were also constructed [Figure 5]. Among them, AC017002.2 was positively correlated with MYC ($r = 0.917$, $P < 0.001$), TGFBI ($r = 0.916$, $P < 0.001$), and cytochrome P450 family 2 sub-family B member 6 (CYP2B6) ($r = 0.933$, $P < 0.001$). LINC00152 was positively correlated with MYC

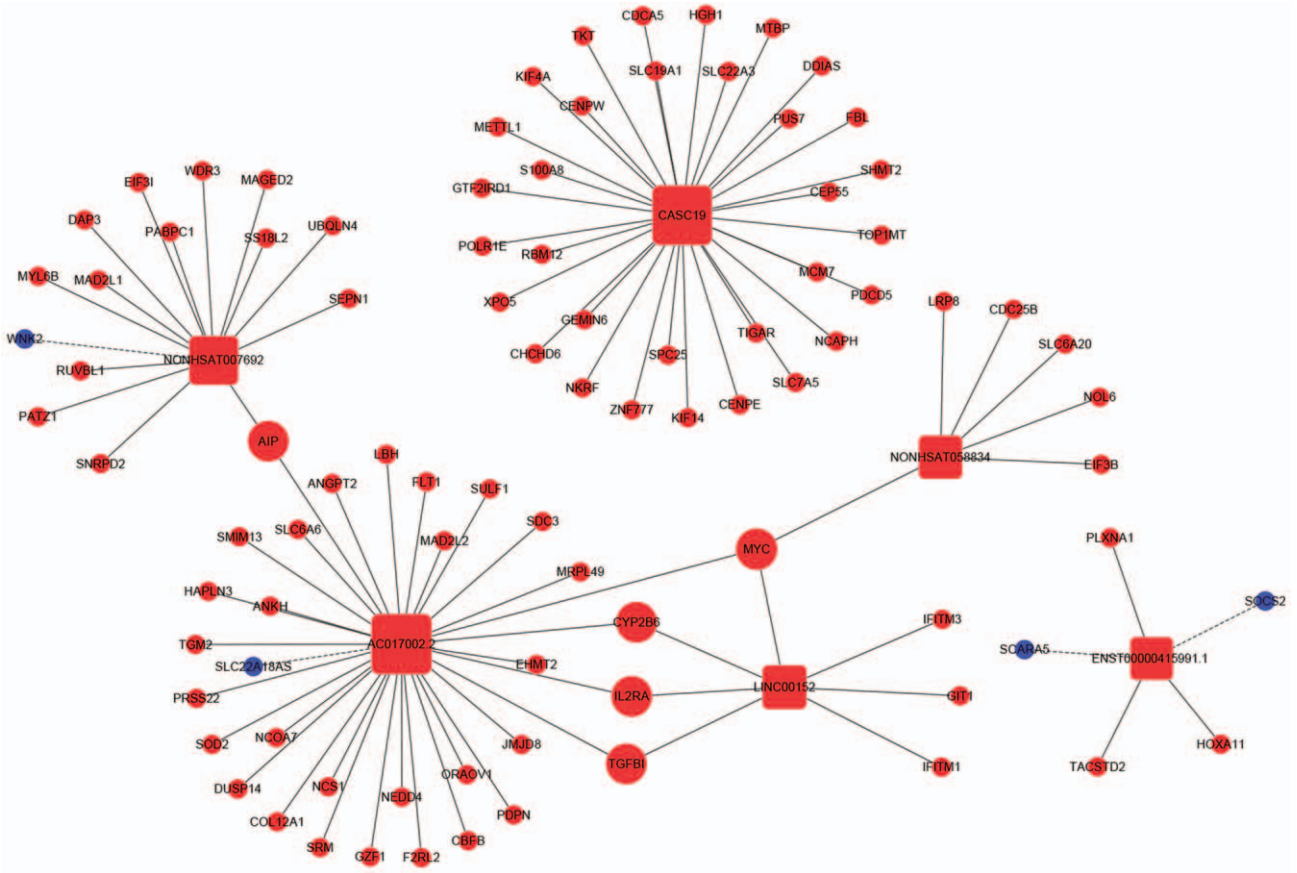

Figure 5: LncRNA-mRNA expression correlation network (co-expression network) analysis of core mRNAs and their correlated lncRNAs. In the network, nodes represent mRNAs, boxes represent lncRNAs, and the size of the node's area represents the value of betweenness centrality. Red indicates up-regulation, and blue indicates down-regulation. The lines between nodes indicate a correlative relationship within the group; solid lines represent positive correlations, and dotted lines represent negative correlations. In the lncRNA target pathway network, circles represent pathways, squares represent lncRNAs, and edges represent the relationships between them. lncRNA: Long noncoding RNA.

($r = 0.931$, $P < 0.001$), TGFBI ($r = 0.913$, $P < 0.001$), and CYP2B6 ($r = 0.917$, $P < 0.001$). NONHSAT058834 was positively associated with MYC ($r = 0.930$, $P < 0.001$). NONHSAT007692 was positively correlated with aryl hydrocarbon receptor interacting protein (AIP) ($r = 0.902$, $P < 0.001$). ENST00000415991.1 was positively correlated with homeobox A11 ($r = 0.958$, $P < 0.001$) and negatively correlated with suppressor of cytokine signaling 2 ($r = -0.904$, $P < 0.001$). The lncRNA CASC19 was positively correlated with SLC7A5 ($r = 0.922$, $P < 0.001$).

Expression and prognosis of the differentially expressed lncRNAs in RC by using GEPIA2

NONHSAT058834, NONHSAT007692, and ENST00000415991.1 were not found in GEPIA2. Compared with that in normal tissues, the expression levels of CASC19 and LINC00152 in RC were significantly increased ($P < 0.05$). For AC017002.2, the difference was not statistically significant.

The overall survival (OS) of the low-LINC00152 group was better than that of the high-LINC00152 group ($P = 0.180$); however, the difference was not statistically significant. The expression levels of CASC19 and AC017002.2 were not correlated with OS in RC [Figure 6].

Expression and prognosis of crucial mRNAs in CRC by using ONCOMINE and PROGgeneV2

Compared with that in normal tissues, the expression levels of MYC, SLC7A5, and TGFBI in RC were significantly increased ($P < 0.001$) [Figure 7].

The expression levels of MYC, SLC7A5, and TGFBI were correlated with OS in CRC, although the differences were not statistically significant [Figure 8].

The ceRNA network analysis

The ceRNA correlation networks were built to identify whether the differentially expressed lncRNAs and mRNAs were involved in a potential interactive mechanism. Based on the six verified differentially expressed lncRNAs, a lncRNA-miRNA-mRNA network associated with these lncRNAs were established with miRNA-targeted mRNAs and miRNA-targeted lncRNAs [Figure 9]. From this lncRNA-miRNA-mRNA network, we found some mutually regulated ceRNA networks [Table 2].

Discussion

RC is a serious malignant tumor, and its exact pathological molecular mechanisms are still largely unknown. Recently,

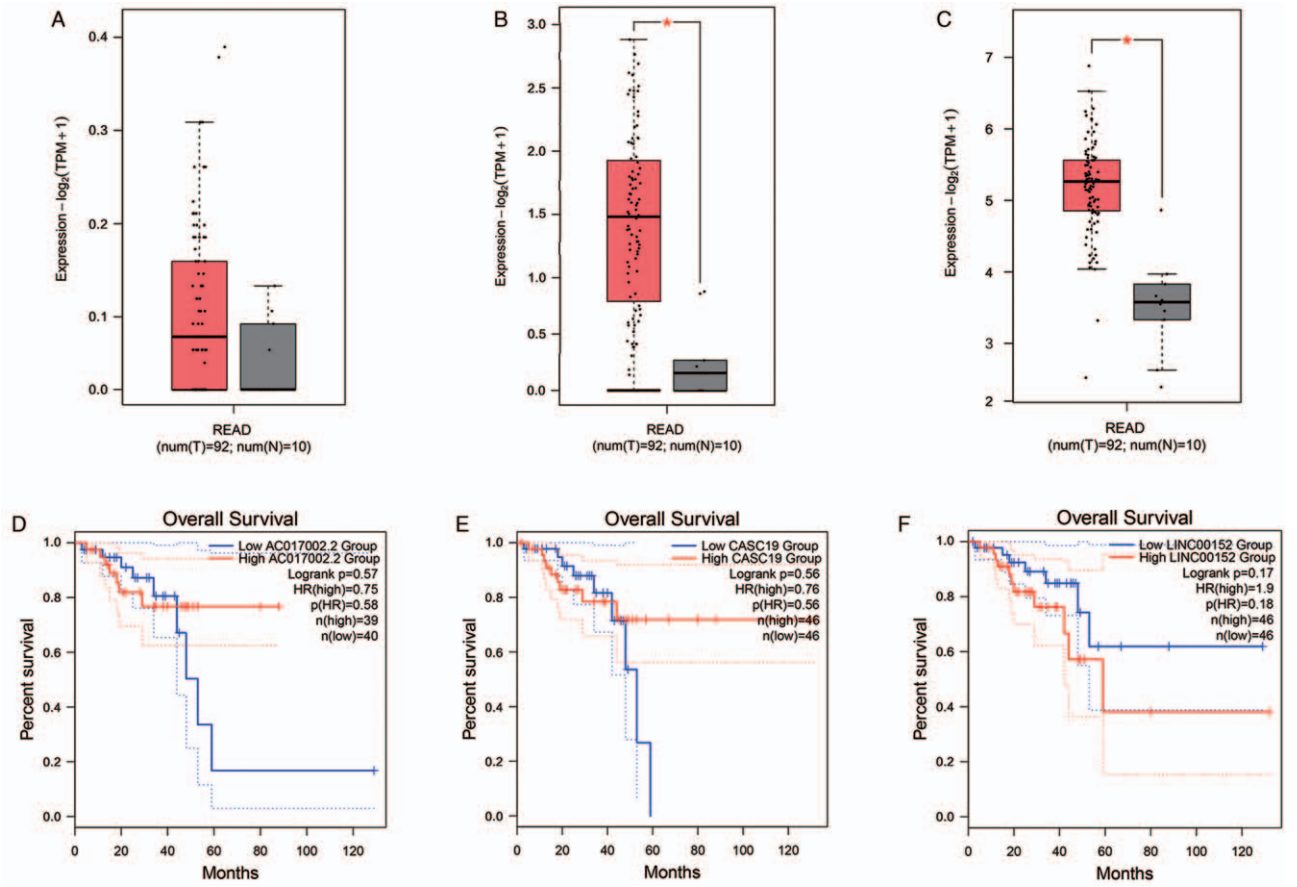

Figure 6: The expression and OS of the differentially expressed lncRNAs in rectal cancer by using GEPIA2. (A) The expression of AC017002.2 in normal and RC tissues. (B) The expression of CASC19 in normal and RC tissues. $^{*}P < 0.05$. (C) The expression of LINC00152 in normal and RC tissues. $^{*}P < 0.05$. (D) The OS of high- and low-AC017002.2 RC patients (HR = 0.750, $P = 0.580$). (E) The OS of high- and low-CASC19 RC patients (HR = 0.760, $P = 0.560$). (F) The OS of high- and low-LINC00152 RC patients (HR = 1.900, $P = 0.180$). CASC19: Cancer susceptibility 19; HR: Hazard ratio; lncRNA: Long non-coding RNA; OS: Overall survival; RC: Rectal cancer.

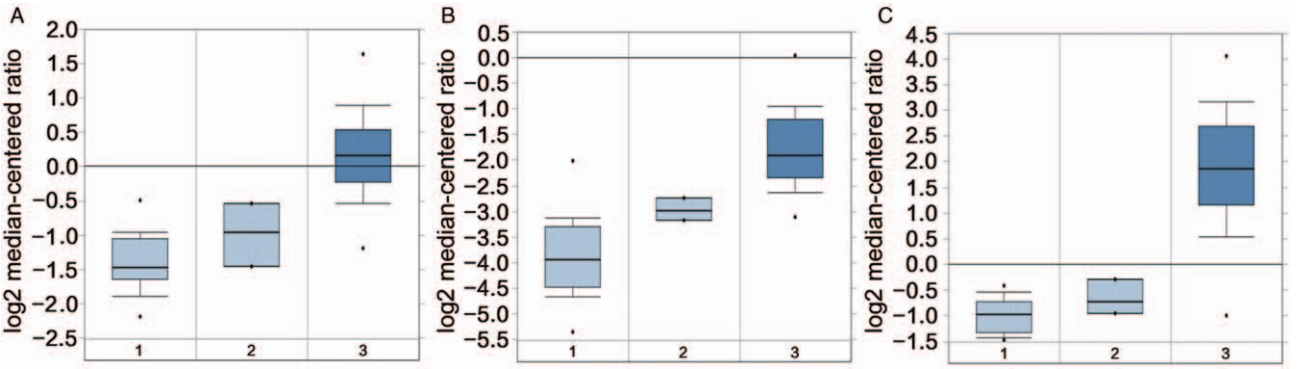

Figure 7: The expression of the crucial mRNAs in colon, rectum, and rectal adenocarcinoma groups using ONCOMINE analysis. (A) The expression of MYC (Group 3 *vs.* Group 2, $P < 0.001$). (B) The expression of SLC7A5 (Group 3 *vs.* Group 2, $P < 0.001$). (C) The expression of TGFBI (Group 3 *vs.* Group 2, $P < 0.001$). Group 1: Colon group ($n = 19$); Group 2: Rectum group ($n = 3$); Group 3: Rectal adenocarcinoma group ($n = 60$). MYC: Myelocytomatosis viral oncogene; RC: Rectal cancer; SLC7A5: Solute carrier family 7 member 5; TGFBI: Transforming growth factor beta induced.

it has been widely reported that lncRNAs play an important role in the occurrence and progression of RC. For instance, Tao *et al*[13] reported that NF-kappaB interacting lncRNA was down-regulated and played a tumor-suppressive role in RC. Tsai *et al*[14] reported that Linc00659 acted as a novel oncogene in CRC by regulating cancer cell growth. Zhang *et al*[15] reported that the expression levels of phosphoglucomutase 5 antisense RNA 1 (PGM5-AS1) were significantly down-regulated in CRC tissues and PGM5-AS1 suppressed CRC cell growth both *in vivo* and *in vitro*. Therefore, we performed a primary analysis of the GeneChip profiles of aberrantly expressed

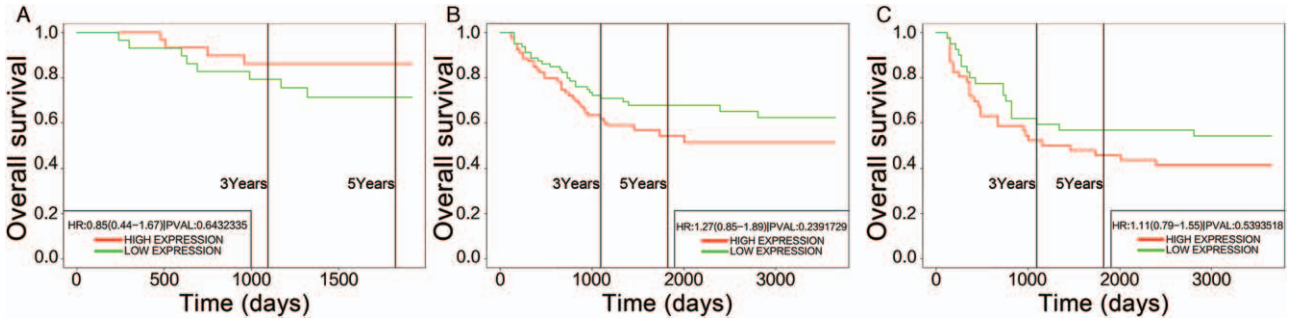

Figure 8: The prognosis of the crucial mRNAs in colorectal cancer by using PROGgeneV2. (A) The OS of high-MYC group CRC patients and low-MYC group CRC patients (HR = 0.850, $P = 0.643$). (B) The OS of high-SLC7A5 group CRC patients and low-SLC7A5 group CRC patients (HR = 1.270, $P = 0.239$). (C) The OS of high-TGFBI group CRC patients and low-TGFBI group CRC patients (HR = 1.110, $P = 0.539$). CRC: Colorectal cancer; HR: Hazard ratio; MYC: Myelocytomatosis viral oncogene; OS: Overall survival; SLC7A5: Solute carrier family 7 member 5; TGFBI: Transforming growth factor beta induced.

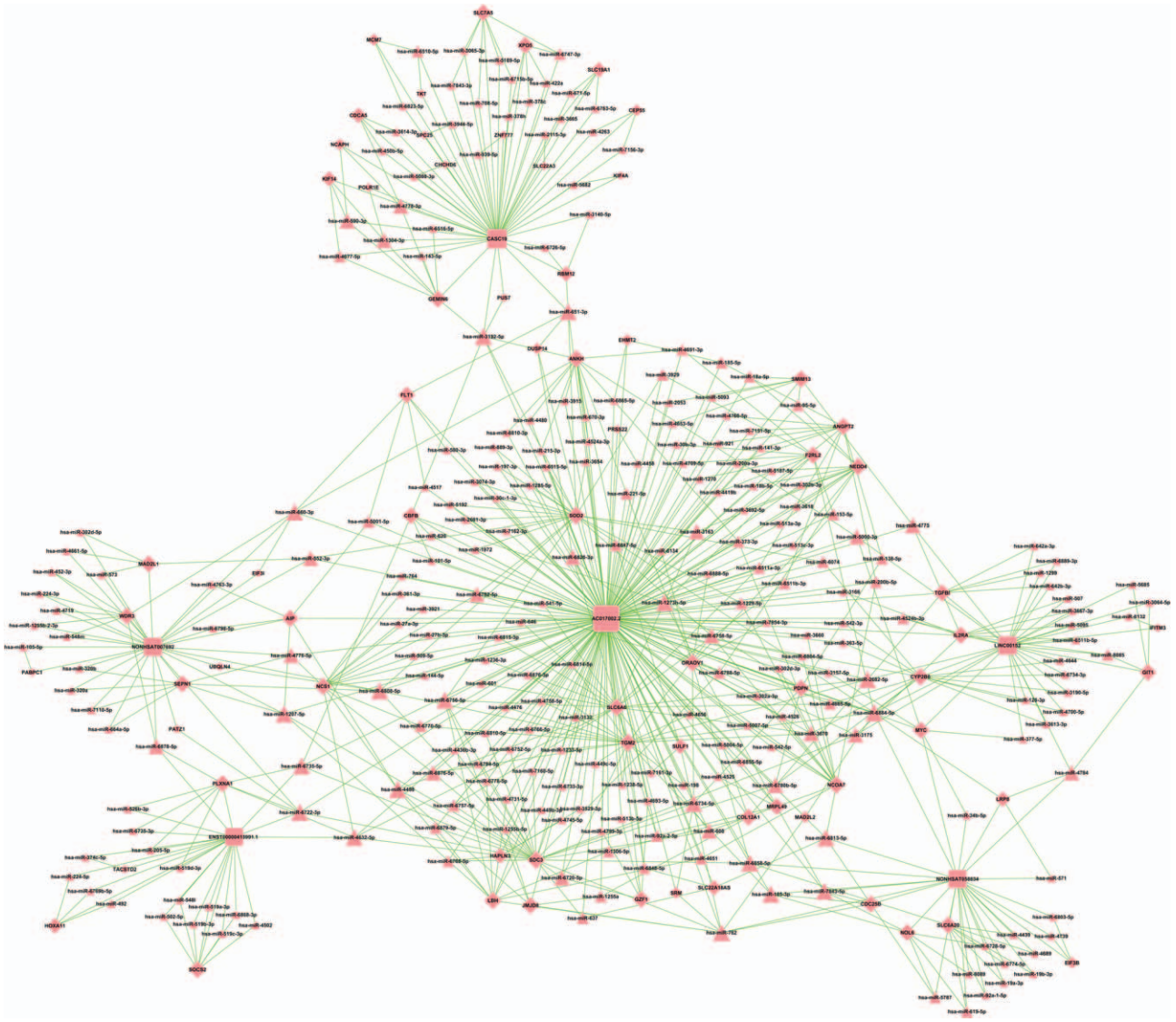

Figure 9: CeRNA network analysis of differentially expressed mRNAs, lncRNAs, and miRNAs. In the network, nodes represent mRNAs, boxes represent lncRNAs, squares represent miRNAs, and the size of the node's area represents the value of betweenness centrality. Red indicates up-regulation, blue indicates down-regulation, and edges indicate target interactions. CeRNA: Competing endogenous RNA; lncRNA: Long non-coding RNA; miRNA: MicroRNA.

Table 2: Crucial lncRNAs, miRNAs, and mRNAs involved in the ceRNA network.

LncRNAs	miRNAs-mRNAs
AC017002.2	miR-5000-3p-MYC, miR-3157-MYC, miR-2682-5p-TGFBI, miR-6884-5p-CYP2B6, miR-3175-IL2RA, miR-5001-5p-AIP
LINC00152	miR-5000-3p-MYC, miR-377-5p-MYC, miR-642b-3p-TGFBI, miR-128-3p-CYP2B6, miR-3175-IL2RA
NONHSAT058834	miR-34b-5p-MYC,miR-6774-5p, miR-6858-5p-SLC6A20
NONHSAT007692	miR-6798-5p, miR-4763-3p, miR-1207-5p-AIP
CASC19	miR-6747-3p, miR-5189-5p, miR-708-5p-SLC7A5
ENST00000415991.1	miR-224-5p, miR-6769b-5p-HOXA11

lncRNA: Long noncoding RNA; miRNA: MicroRNA; ceRNA: Competing endogenous RNA; MYC: Myelocytomatosis viral oncogene; TGFBI: Transforming growth factor beta induced; CYP2B6: Cytochrome P450 family 2 sub-family B member 6; IL2RA: Interleukin 2 receptor sub-unit alpha; AIP: Aryl hydrocarbon receptor interacting protein; HOXA11: Homeobox A11.

lncRNAs and mRNAs in RC. In our study, we identified 1658 lncRNAs and 1783 mRNAs that were differentially expressed in RC tissue specimens.

Then, we performed GO enrichment analyses of these genes. Some vital biological functions were identified, mainly including cell proliferation, cell migration, angiogenesis, and cellular response to zinc ions. Hanahan and Weinberg[16] elucidated that the hallmarks of cancer include sustaining proliferative signaling, activating invasion and metastasis, and inducing angiogenesis. These classic cancer hallmarks are basically consistent with our screening results. Previous studies have also suggested that these classic cancer hallmarks are regulated by lncRNAs in CRC or RC.

In addition, the results of the pathway analysis mainly included the IL-17 signaling pathway, cell cycle signaling pathway, Wnt signaling pathway and mineral absorption signaling pathway. The IL-17 signaling pathway had the highest significance and the lowest FDR in our results. The IL-17 family consists of six members, IL-17A-F, while IL-17B is involved in inflammation and CRC.[17] It was also reported that the overexpression of lncRNA colorectal liver metastasis associated transcript 3 promoted CRC cell proliferation by inhibiting cell cycle arrest and apoptosis.[18] Regarding the Wnt signaling pathway, Yu *et al*[19] reported that lncRNA solute carrier organic anion transporter family member 4A1 antisense RNA 1 facilitated the growth and metastasis of CRC by activating the Wnt/β-catenin signaling pathway. Our research implied similar results as those of the studies above.

Eleven lncRNAs were screened from 1658 lncRNAs and verified by qPCR. After verification, we identified six lncRNAs, including AC017002.2, LINC00152, NONHSAT058834, NONHSAT007692, CASC19, and ENST00000415991.1. LncRNAs were once thought to represent transcriptome noise or garbage sequences. However, it was gradually discovered that lncRNAs are involved in a wide variety of physiological and pathological processes in organisms.[3] Bian *et al*[20] reported that LINC00152 promoted CRC cell proliferation, migration and invasion, and modulated NOTCH1 expression by inhibiting miR-139-5p. Moreover, Wang *et al*[21] reported that overexpressed CASC19 positively regulates cell migration inducing hyaluronidase 1 expression by targeting miR-140-5p and promotes CRC cell invasion, migration, and proliferation. The results of these studies were also consistent with our results. AC017002.2, NONHSAT058834, NONHSAT007692, and ENST00000415991.1 have not yet been reported in RC. Further experiments are needed to explore their phenotypes and functions in RC.

Regulating corresponding mRNAs is one of the main ways that lncRNAs exert their regulatory functions. In our study, AC017002.2, LINC00152, NONHSAT058834, NONHSAT007692, CASC19, ENST00000415991.1 and their correlated mRNAs MYC, TGFBI, SLC7A5 were identified by a lncRNA-mRNA expression correlation network. GEPIA2, ONCOMINE, and PROGgeneV2 were used to mine the expression and prognosis of the three crucial mRNAs and six verified lncRNAs. The results of data mining were mostly consistent with those of GeneChip and qRT-PCR.

MYC, SLC7A5, and TGFBI all ranked in the top 20 according to fold change and had a clear correlation with the validated lncRNAs. Acting as a proto-oncogene, MYC was significantly associated with CRC. A study suggested that MYC was a target gene of Wnt signaling and promoted the tumorigenesis of CRC after activation.[22] Another vital mRNA in our study was SLC7A5. It was reported that the overexpression of SLC7A5 induced by MYC in colon cancer tissues and cells promoted colon cancer cell proliferation.[23]*TGFBI* gene was originally identified by Skonier *et al*[24] as a major TGF-β-responsive gene in the lung adenocarcinoma cell line A549. A study suggested that the expression level of *TGFBI* was high in CRC, serving as an independent poor prognostic factor for CRC patients.[25] Our research suggested similar results as the studies above.

Based on the results above, we constructed a lncRNA-mRNA coexpression network and a lncRNA-miRNA-mRNA ceRNA network to further explore the correlations between them and try to reveal the underlying mechanisms. The lncRNA-miRNA-mRNA ceRNA network was constructed based on the ceRNA hypothesis that lncRNAs regulate the activity of mRNAs by sequestering and binding miRNAs, thereby acting as miRNA sponges.[26] AC017002.2 was in a relatively highlighted position in our

ceRNA network. We found that AC017002.2-miR-5000-3p-MYC constituted the ceRNA network. For miR-5000-3p, a current study implied that up-regulated miR-5000-3p in side population cells might be a potential miRNA biomarker of multiple drug resistance in colon cancer.[27] Further experiments on AC017002.2-miR-5000-3p-MYC are necessary to verify their combination and reveal more phenotypes.

In summary, we identified the lncRNAs AC017002.2, LINC00152, CASC19, NONHSAT058834, NONHSAT007692, and ENST00000415991.1 as well as the crucial mRNAs MYC, TGFBI, and SLC7A5 in RC through integrated gene chip analysis. Some of these lncRNAs have not been reported in RC yet. These results might provide some potential biomarkers for the diagnosis and treatment of RC.

Acknowledgements

The authors are sincerely grateful to the Department of Central Laboratory, Beijing Shijitan Hospital, Capital Medical University for the help of experiment skills.

Funding

This work was supported by a grant from the Beijing Municipal Health Commission's Key Clinical Specialty (Oncology) (Project No. 2018ZLLCZDZK1).

Conflicts of interest

None.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424. doi: 10.3322/caac.21492.
2. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? Cancer Commun (Lond) 2019;39:22. doi: 10.1186/s40880-019-0368-6.
3. Zhang X, Hong R, Chen W, Xu M, Wang L. The role of long noncoding RNA in major human disease. Bioorg Chem 2019;92:103214. doi: 10.1016/j.bioorg.2019.103214.
4. Chen G, Gu Y, Han P, Li Z, Zhao JL, Gao MZ. Long noncoding RNA SBF2-AS1 promotes colorectal cancer proliferation and invasion by inhibiting miR-619-5p activity and facilitating HDAC3 expression. J Cell Physiol 2019;234:18688–18696. doi: 10.1002/jcp.28509.
5. Ozawa T, Matsuyama T, Toiyama Y, Takahashi N, Ishikawa T, Uetake H, *et al*. CCAT1 and CCAT2 long noncoding RNAs, located within the 8q.24.21 'gene desert', serve as important prognostic biomarkers in colorectal cancer. Ann Oncol 2017;28:1882–1888. doi: 10.1093/annonc/mdx248.
6. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, *et al*. The gene ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258–D261. doi: 10.1093/nar/gkh036.
7. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, *et al*. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36:D480–D484. doi: 10.1093/nar/gkm882.
8. Ma L, Li A, Zou D, Xu X, Xia L, Yu J, *et al*. LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. Nucleic Acids Res 2015;43:D187–D192. doi: 10.1093/nar/gku1167.
9. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 2019;47:W556–W560. doi: 10.1093/nar/gkz430.
10. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, *et al*. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1–6. doi: 10.1016/s1476-5586(04)80047-2.
11. Goswami CP, Nakshatri H. PROGgeneV2: enhancements on the existing database. BMC Cancer 2014;14:970. doi: 10.1186/1471-2407-14-970.
12. Greco S, Gaetano C, Martelli F. Long noncoding competing endogenous RNA networks in age-associated cardiovascular diseases. Int J Mol Sci 2019;20:E3079. doi: 10.3390/ijms20123079.
13. Tao F, Xu Y, Yang D, Tian B, Jia Y, Hou J, *et al*. LncRNA NKILA correlates with the malignant status and serves as a tumor-suppressive role in rectal cancer. J Cell Biochem 2018;119:9809–9816. doi: 10.1002/jcb.27300.
14. Tsai KW, Lo YH, Liu H, Yeh CY, Chen YZ, Hsu CW, *et al*. Linc00659, a long noncoding RNA, acts as novel oncogene in regulating cancer cell growth in colorectal cancer. Mol Cancer 2018;17:72. doi: 10.1186/s12943-018-0821-1.
15. Zhang Q, Ding Z, Wan L, Tong W, Mao J, Li L, *et al*. Comprehensive analysis of the long noncoding RNA expression profile and construction of the lncRNA-mRNA co-expression network in colorectal cancer. Cancer Biol Ther 2020;21:157–169. doi: 10.1080/15384047.2019.1673098.
16. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–674. doi: 10.1016/j.cell.2011.02.013.
17. Al-Samadi A, Moossavi S, Salem A, Sotoudeh M, Tuovinen SM, Konttinen YT, *et al*. Distinctive expression pattern of interleukin-17 cytokine family members in colorectal cancer. Tumour Biol 2016;37:1609–1615. doi: 10.1007/s13277-015-3941-x.
18. Ye LC, Chen T, Zhu DX, Lv SX, Qiu JJ, Xu J, *et al*. Downregulated long non-coding RNA CLMAT3 promotes the proliferation of colorectal cancer cells by targeting regulators of the cell cycle pathway. Oncotarget 2016;7:58931–58938. doi: 10.18632/oncotarget.10431.
19. Yu J, Han Z, Sun Z, Wang Y, Zheng M, Song C. LncRNA SLCO4A1-AS1 facilitates growth and metastasis of colorectal cancer through beta-catenin-dependent Wnt pathway. J Exp Clin Cancer Res 2018;37:222. doi: 10.1186/s13046-018-0896-y.
20. Bian Z, Zhang J, Li M, Feng Y, Yao S, Song M, *et al*. Long non-coding RNA LINC00152 promotes cell proliferation, metastasis, and confers 5-FU resistance in colorectal cancer by inhibiting miR-139-5p. Oncogenesis 2017;6:395. doi: 10.1038/s41389-017-0008-4.
21. Wang XD, Lu J, Lin YS, Gao C, Qi F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. World J Gastroenterol 2019;25:1697–1714. doi: 10.3748/wjg.v25.i14.1697.
22. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, *et al*. Identification of c-MYC as a target of the APC pathway. Science 1998;281:1509–1512. doi: 10.1126/science.281.5382.1509.
23. Venkateswaran N, Lafita-Navarro MC, Hao YH, Kilgore JA, Perez-Castro L, Braverman J, *et al*. MYC promotes tryptophan uptake and metabolism by the kynurenine pathway in colon cancer. Genes Dev 2019;33:1236–1251. doi: 10.1101/gad.327056.119.
24. Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD, Purchio AF. cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. DNA Cell Biol 1992;11:511–522. doi: 10.1089/dna.1992.11.511.
25. Zhu J, Chen X, Liao Z, He C, Hu X. TGFBI protein high expression predicts poor prognosis in colorectal cancer patients. Int J Clin Exp Pathol 2015;8:702–710.
26. Lyu L, Xiang W, Zhu JY, Huang T, Yuan JD, Zhang CH. Integrative analysis of the lncRNA-associated ceRNA network reveals lncRNAs as potential prognostic biomarkers in human muscle-invasive bladder cancer. Cancer Manag Res 2019;11:6061–6077. doi: 10.2147/CMAR.S207336.
27. Xia ZS, Wang L, Yu T, Zhong W, Lian GD, Wu D, *et al*. MiR-5000-3p, miR-5009-3P and miR-552: potential microRNA biomarkers of side population cells in colon cancer. Oncol Rep 2014;32:589–596. doi: 10.3892/or.2014.3232.

How to cite this article: Wang DZ, Chen GY, Li YF, Zhang NW. Comprehensive analysis of long non-coding RNA and mRNA expression profile in rectal cancer. Chin Med J 2020;133:1312–1321. doi: 10.1097/CM9.0000000000000753