Cellular Physiology and Biochemistry

Cell Physiol Biochem 2018;49:1778-1791
DOI: 10.1159/000493623
Published online: 20 September, 2018
Accepted: 11 September, 2018

Original Paper

Analysis of lncRNA-Associated ceRNA Network Reveals Potential lncRNA Biomarkers in Human Colon Adenocarcinoma

Zhiyuan Zhang[a,b] Wenwei Qian[a,b] Sen Wang[a,b] Dongjian Ji[a,b] Qingyuan Wang[a,b] Jie Li[a,b] Wen Peng[a,b] Jiou Gu[a,b] Tao Hu[a,b] Bing Ji[a,b] Yue Zhang[a,b] Shijia Wang[a,b] Yueming Sun[b]

[a]The First School of Clinical Medicine, Nanjing Medical University, Nanjing, [b]Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Key Words
Colon adenocarcinoma • LncRNA • ceRNA network • Biomarkers • Bioinformatics analysis

Abstract
Background/Aims: Long non-coding RNAs (lncRNAs) acting as competing endogenous RNAs (ceRNAs) play significant roles in the development of tumors, but the functions of specific lncRNAs and lncRNA-related ceRNA networks have not been fully elucidated for colon adenocarcinoma (COAD). In this study, we aimed to clarify the lncRNA-microRNA (miRNA)-mRNA ceRNA network and potential lncRNA biomarkers in COAD. ***Methods:*** We extracted data from The Cancer Genome Atlas (TCGA) and identified COAD-specific mRNAs, miRNAs, and lncRNAs. The biological processes in Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were analyzed for COAD-specific mRNAs. We then constructed a ceRNA network of COAD-specific mRNAs, miRNAs and lncRNAs and analyzed the correlation between expression patterns and clinical features of the lncRNAs involved. After identifying potential mRNA targets of 4 lncRNAs related to overall survival (OS), we conducted stepwise analysis of these targets through GO and KEGG. Using tissue samples from our own patients, we also verified certain analytical results using quantitative real-time PCR (qRT-PCR). ***Results:*** Data from 521 samples (480 tumor tissue and 41 adjacent non-tumor tissue samples) were extracted from TCGA. A total of 258 specific lncRNAs, 206 specific miRNAs, and 1467 specific mRNAs were identified (absolute $\log_2$ [fold change] >2, false discovery rate <0.01). Analysis of KEGG revealed that specific mRNAs were enriched in cancer-related pathways. The ceRNA network was constructed with 64 lncRNAs, 18 miRNAs, and 42 mRNAs. Among these lncRNAs involved in the network, 3 lncRNAs (LINC00355, HULC, and IGF2-AS) were confirmed to be associated with certain clinical features and 4 lncRNAs (HOTAIR, LINC00355, KCNQ1OT1, and

Z. Zhang, W. Qian and S. Wang contributed equally to this work.

Yueming Sun
Dep.of General Surgery, The First Affiliated Hospital of Nanjing Med. University
Nanjing, Jiangsu (China)
Tel. +86 135051888397, E-Mail jssym@vip.sina.com

TSSC1-IT1) were found to be negatively linked to OS (log-rank p < 0.05). KEGG showed that the potential mRNA targets of these 4 lncRNAs may be concentrated in the MAPK pathway. Certain results were validated by qRT-PCR. ***Conclusion:*** This study providing novel insights into the lncRNA-miRNA-mRNA ceRNA network and reveals potential lncRNA biomarkers in COAD.

Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide, and colon adenocarcinoma (COAD) is a common type of CRC [1]. Considerable advancements have been made in the study of COAD in recent years, but the mechanisms remain unclear. Genetic events play significant roles in COAD. Thus, we are working to identify long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and mRNAs that are differentially expressed in COAD and to construct a competing endogenous RNA (ceRNA) network to reveal their potential interaction in COAD.

LncRNA, a type of RNA that was once was viewed as transcriptional "noise" without biological functions and protein-coding capacity, is defined as a type of RNA transcript of more than 200 nucleotides [1]. Recently, lncRNAs were reported to be closely involved in CRC. For example, homeobox transcript antisense intergenic RNA (HOTAIR), HULC, and linc00152 were reported to play vital roles in its development, and their high expression levels predicted a poor prognosis [2, 3]. LncRNAs are thought to act via various mechanisms. Salmena et al. [4] presented the ceRNA hypothesis whereby RNA transcripts can communicate with each other via miRNA response elements (MREs). This competition exerts a crucial role in tumorigenesis by affecting the expression levels of various RNAs through MREs.

Some studies have been conducted on COAD, but those with large sample sizes that detect expression patterns of COAD-specific lncRNAs have not yet been reported. Moreover, studies with small sample sizes cannot identify lncRNAs that are related to tumor-node-metastasis (TNM) stage, overall survival (OS), or other clinical features with statistical impact, and very few studies have sought to elucidate the ceRNA network in COAD.

To address these issues, we sought to elucidate the ceRNA network and lncRNAs in COAD in this study. We used bioinformatic tools and analyzed data from the The Cancer Genome Atlas (TCGA), a public platform containing RNA sequencing data of 480 COAD tumor tissue samples and 41 adjacent non-tumor tissue samples. Using tissue samples from our own patients, we also verified certain analytical results using quantitative real-time PCR (qRT-PCR). This approach was useful in revealing potential lncRNA biomarkers and constructing a ceRNA network in COAD.

Materials and Methods

Patients and samples

Information on 459 patients was extracted from TCGA. RNA expression patterns and clinical data such as pathologic stage and TNM information were obtained from TCGA. Exclusion criteria were set as follows: i) histologic diagnosis not COAD; ii) other malignancy aside from COAD; iii) incomplete data for analysis; and iv) preoperative chemoradiation received. Ultimately, data for 480 tumor tissue samples and 41 adjacent non-tumor tissue samples were analyzed. According to the pathologic stage, the number of tumor tissues in stages I, II, III, and IV were 81, 187, 133, and 66 respectively (staging was unknown for 13 tumor tissues). The study followed the TCGA guidelines.

For qRT-PCR analysis, we selected COAD tissue specimens and their paired adjacent non-tumor tissue specimens from 50 patients at The First Affiliated Hospital of Nanjing Medical University (Jiangsu, China). These patients (aged 45-80 years) were diagnosed with COAD based on histopathology and clinical history. Tissues were stored in RNAlater (Ambion, Austin, TX) at −80°C until RNA extraction.

RNA sequence datasets and analysis

The COAD RNA expression profile data (level 3) of the corresponding patients were downloaded from TCGA data portal (May 2017). TCGA provided the normalized count data of RNA sequencing including lncRNAs and mRNAs expression profiles via the RNASeqV2 system. The STAD level 3 microRNA sequencing (miRNAseq) data, downloaded from TCGA, were collected by Illumina HiSeq 2000 miRNAseq platforms (Illumina Inc., Hayward, CA). Next, we divided the tumor samples into 4 groups (tumor stages I, II, III, and IV) and used Empirical Analysis of Digital Gene Expression Data in R (edgeR) from R Studio (R version 3.4.1) to analyze differences in the expression levels between each of the 4 tumor stages and adjacent non-tumor tissues and between all tumor tissues and all adjacent non-tumor tissues (absolute $\log_2$ [fold change] [logFC] > 2.0, false discovery rate [FDR] < 0.01). Then, we chose intersections of differentially expressed COAD lncRNAs, mRNAs, and miRNAs from the 5 comparative groups for further analysis. This selection process is shown in Fig. 1. LncRNAs, miRNAs, and mRNAs were named COAD-specific lncRNAs, miRNAs, and mRNAs, respectively.

Functional enrichment analysis

We used DAVID 6.8 (Database for Annotation, Visualization, and Integrated Discovery, https://david.ncifcrf.gov/) for functional enrichment analysis. The biological processes in Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were searched for pathways at the significance level set ($p < 0.05$ and enrichment score >1.5).

Construction of the ceRNA network

The ceRNA network was constructed based on the theory that lncRNAs can affect miRNA and act as miRNA sponges to further regulate mRNA. We used COAD-specific lncRNAs, miRNAs, and mRNAs to construct the network. miRcode (http://www.mircode.org/) was used to predict the lncRNA-miRNA interactions based on COAD-specific miRNAs. We predicted mRNA targeted by miRNA using Targetscan (http://www.targetscan.org/), miRdb (http://www.mirdb.org/), and miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/). We ultimately retained intersections with the differentially expressed lncRNAs, miRNA, and mRNAs. The ceRNA network was constructed as shown in Fig. 2. Cytoscape v3.0 was used to construct the lncRNA-miRNA-mRNA ceRNA network. LncRNAs involved in the network were designated key lncRNAs.

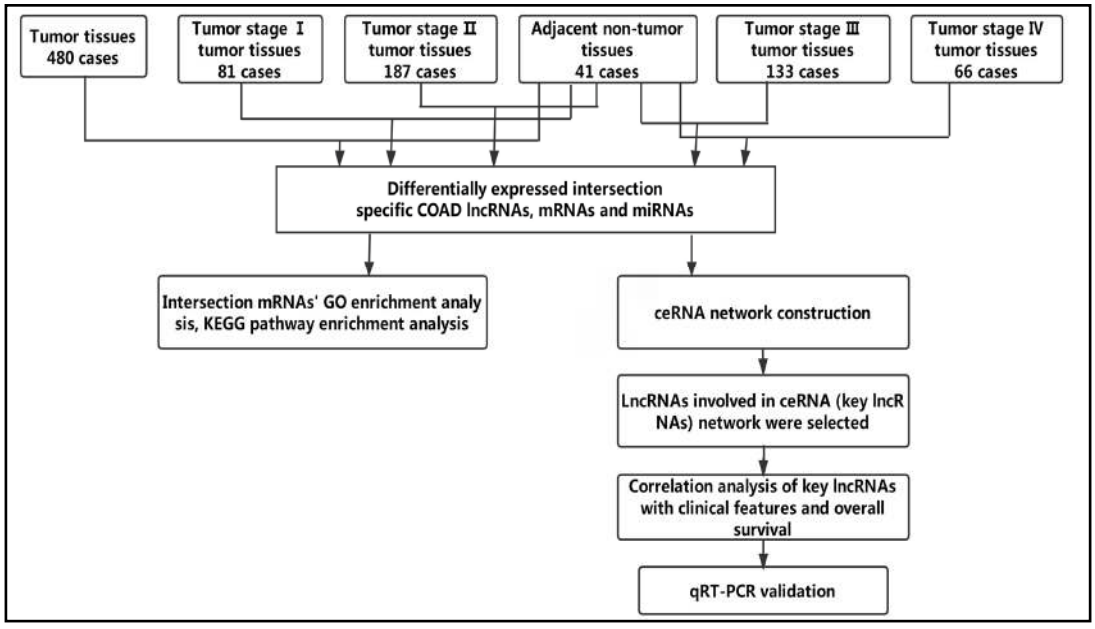

Fig. 1. Flow chart of bioinformatics analysis.

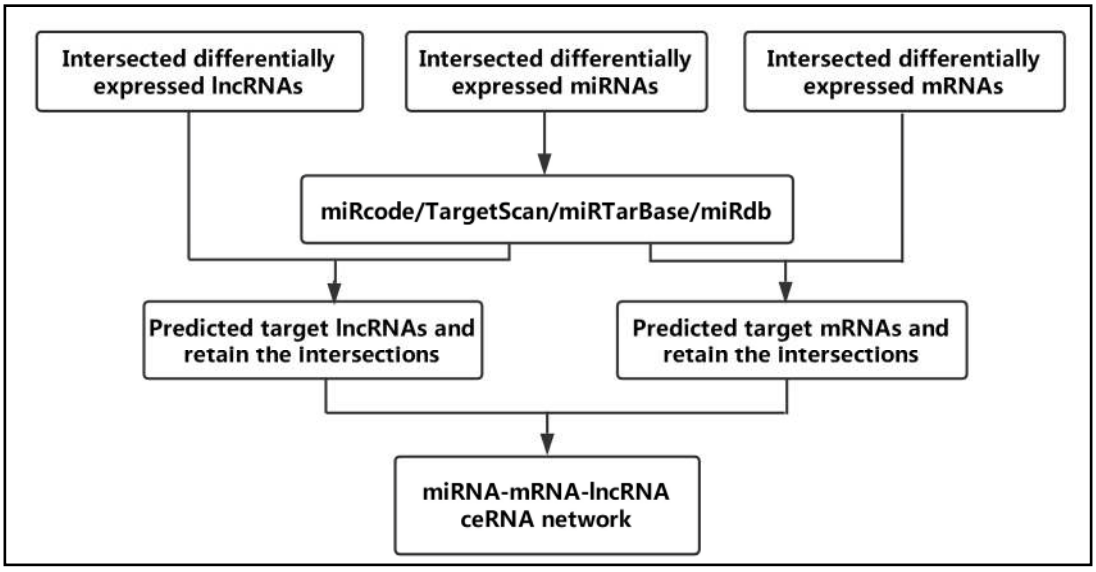

Fig. 2. Flow chart of construction of ceRNA network.

Clinical feature analysis of key lncRNAs

We chose key lncRNAs from the network to study their correlation with clinical features, including sex, tumor stage, TNM stage, distant metastasis, and lymphatic metastasis by edgeR (absolute logFC > 1.5, FDR < 0.01). Clinical data were extracted from TCGA. We also studied the association between key lncRNAs and COAD patients' OS (log-rank $p < 0.05$). Results of OS are shown by Kaplan-Meier survival curves.

mRNA targets of key lncRNAs related to survival

We used the Weighted Gene Co-expression Network Analysis (WGCNA) package in R Studio (R version 3.4.1) to build a co-expression network to analyze the potential mRNA targets of lncRNAs. The expression pattern of lncRNAs and mRNAs were obtained from data in TCGA.

RNA extraction and qRT-PCR validation

We used the TRIzol reagent (Invitrogen, Carlsbad, CA) to extract RNA from tissues and the PrimeScript RT reagent kit (Takara, Dalian, China) to synthesize complementary DNA. qRT-qPCR was carried out using the SYBR-Green PCR kit (Roche Diagnostics, Indianapolis, IN) on a StepOnePlus Real-Time PCR system (Applied Biosystems, Foster City, CA). PCR cycling conditions were as follows: 95°C for 30 s; 40 cycles of 95°C for 5 s and 60°C for 30 s; and dissociation at 95°C for 15 s, 60°C for 60 s, and 95°C for 15 s. Results were analyzed using the $2^{-\Delta\Delta Ct}$ method [5]. Fold change of qRT-PCR was presented as $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct = (CtRNAs - CtGApDH)_{tumor} - (CtRNAs - CtGApDH)_{adjacent\ non\text{-}tumor\ tissues}$, and $logFC = -\Delta\Delta Ct$. qRT-PCR reactions were all repeated three times.

Statistical analysis

Statistical analysis was performed by R Studio (R version 3.4.1) and GraphPad Prism 5.0 software (GraphPad Software Inc., La Jolla, CA). EdgeR from R Studio was used to identify different expression of RNAs. Paired t-test was used in qRT-PCR to compare differences. Statistical significance was set at $p < 0.05$.

Results

COAD-specific lncRNAs in COAD patients

We identified 381 lncRNAs that were differentially expressed between tumor tissues and adjacent non-tumor tissues from TCGA (absolute logFC > 2, FDR < 0.01), of which 283 lncRNAs were upregulated and 98 lncRNAs were downregulated. Further analysis was carried out between tumor tissues in patients with stage I, II, III, and IV cancer and adjacent

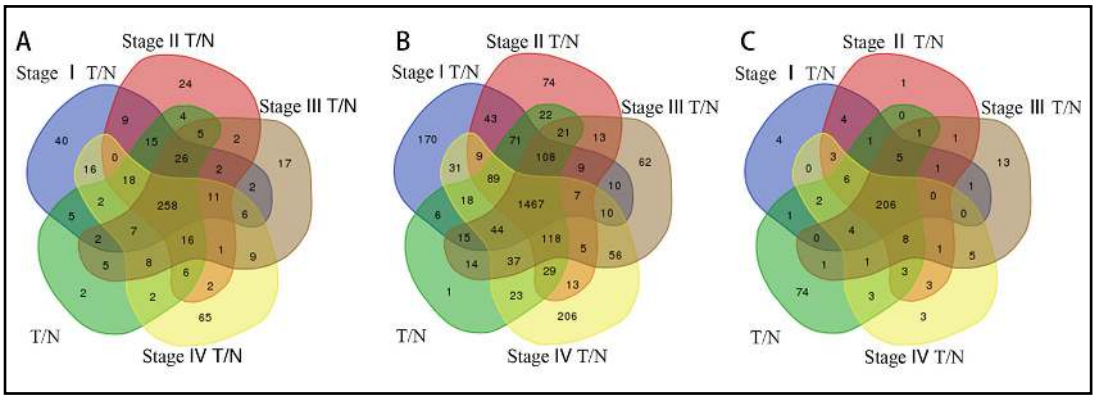

Fig. 3. Venn diagram analysis of differentially expressed lncRNAs(A), mRNAs(B) and miRNA(C) between integrated COAD tissues, 4 stages (stage Ⅰ, Ⅱ, Ⅲ and Ⅳ) and non-tumor tissues respectively. T represents COAD tissues; N represents adjacent non-tumor tissues.

Table 1. Key lncRNAs involved in the ceRNA network. FDR, the false discovery rate, using Benjamini and Hochberg (1995) method

lncRNAs	Log_2(fold-change)	-Log(FDR)	lncRNAs	Log_2(fold-change)	-Log(FDR)
HULC	7.50	10.99	MUC19	2.69	5.82
LINC00460	6.68	32.30	C15orf54	2.68	7.59
ERVMER61-1	6.28	8.84	BTBD9-AS1	2.62	3.25
NKX2-1-AS1	5.58	7.25	LMO7-AS1	2.58	17.56
CLDN10-AS1	5.17	15.38	PVT1	2.54	62.74
LINC00355	5.01	6.91	ST7-OT4	2.54	8.56
POU6F2-AS1	4.96	8.02	WASIR2	2.52	9.61
CRNDE	4.55	49.06	TSSC1-IT1	2.43	9.72
DSCAM-AS1	4.45	4.94	MALAT1	2.42	9.77
BOK-AS1	4.09	5.71	DLEU7-AS1	2.37	15.28
DLX6-AS1	3.95	9.60	STEAP2-AS1	2.32	9.02
LPP-AS1	3.89	4.49	C2orf48	2.28	18.31
MYO16-AS1	3.86	9.53	COL4A2-AS2	2.25	5.38
ABCA9-AS1	3.77	10.71	SHANK2-AS3	2.22	2.71
UCA1	3.76	15.01	EGOT	2.19	8.26
C17orf77	3.76	9.23	KCNQ1OT1	2.13	12.10
H19	3.63	11.86	FAM95B1	-2.22	21.34
IGF2-AS	3.39	7.14	LINC00484	-2.25	39.98
C8orf49	3.36	4.49	LINC00402	-2.31	18.53
MRPL23-AS1	3.35	7.49	LIFR-AS1	-2.41	39.50
ST7-AS2	3.33	4.89	SFTA1P	-2.61	28.95
LINC00524	3.31	7.87	LINC00461	-2.69	27.86
ATP11A-AS1	3.29	5.11	RBMS3-AS3	-2.75	34.98
HECW1-IT1	3.29	3.85	LINC00488	-2.81	22.27
HOTAIR	3.18	4.88	JAZF1-AS1	-2.83	41.55
GAS6-AS1	3.11	23.54	LINC00092	-2.85	86.33
E2F3-IT1	3.03	3.55	HCG23	-2.85	53.22
WT1-AS	2.99	5.58	ADAMTS9-AS2	-2.93	52.84
TBL1XR1-AS1	2.93	6.39	FRMD6-AS2	-3.29	29.26
SPATA13-AS1	2.92	3.77	C20orf166-AS1	-3.48	47.97
USP12-AS1	2.84	3.01	ADAMTS9-AS1	-3.67	57.37
VCAN-AS1	2.70	3.22	LINC00507	-4.28	72.52

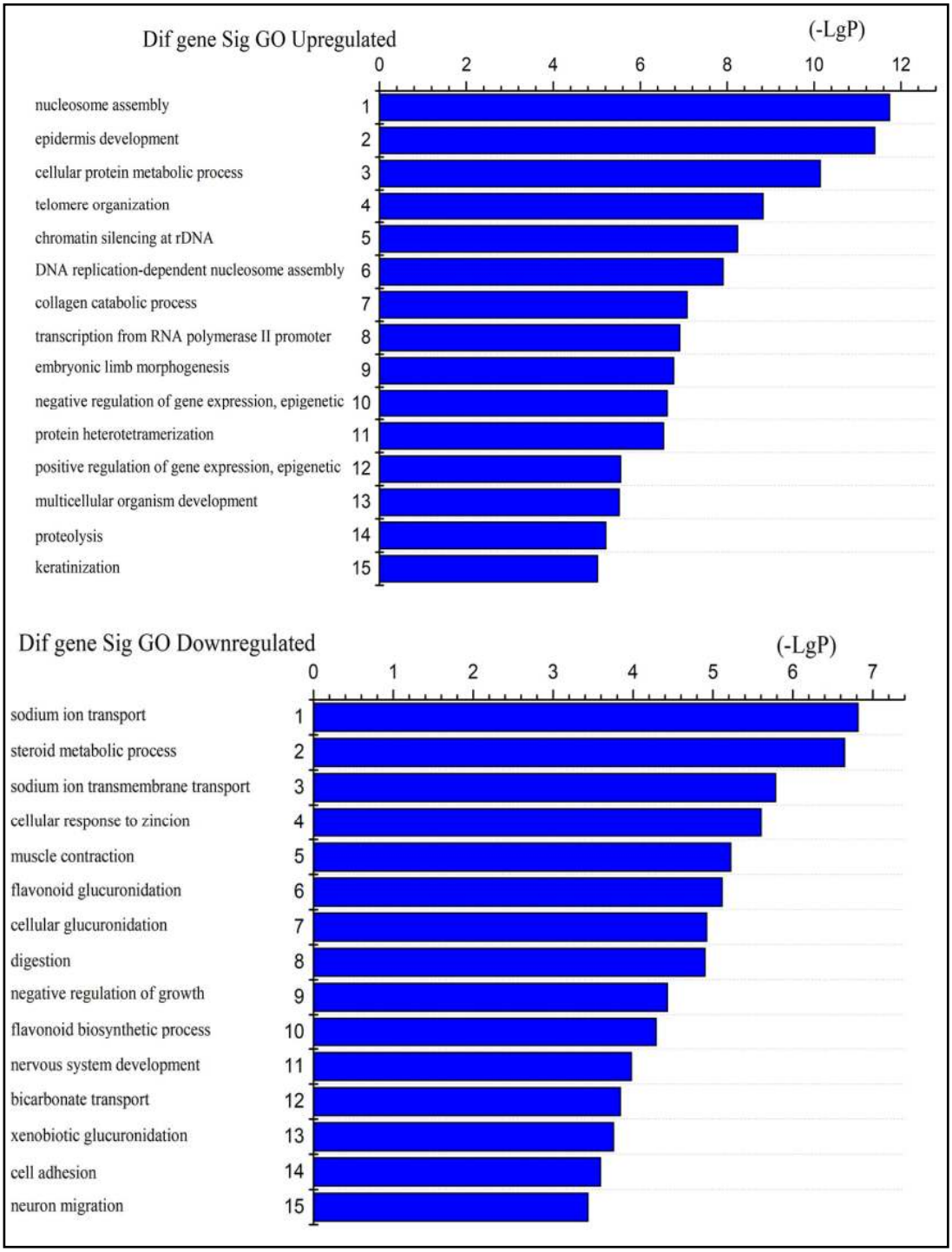

Fig. 4. Top 15 GO results of intersected upregulated and downregulated mRNAs respectively (-logP represents -log(P-value)).

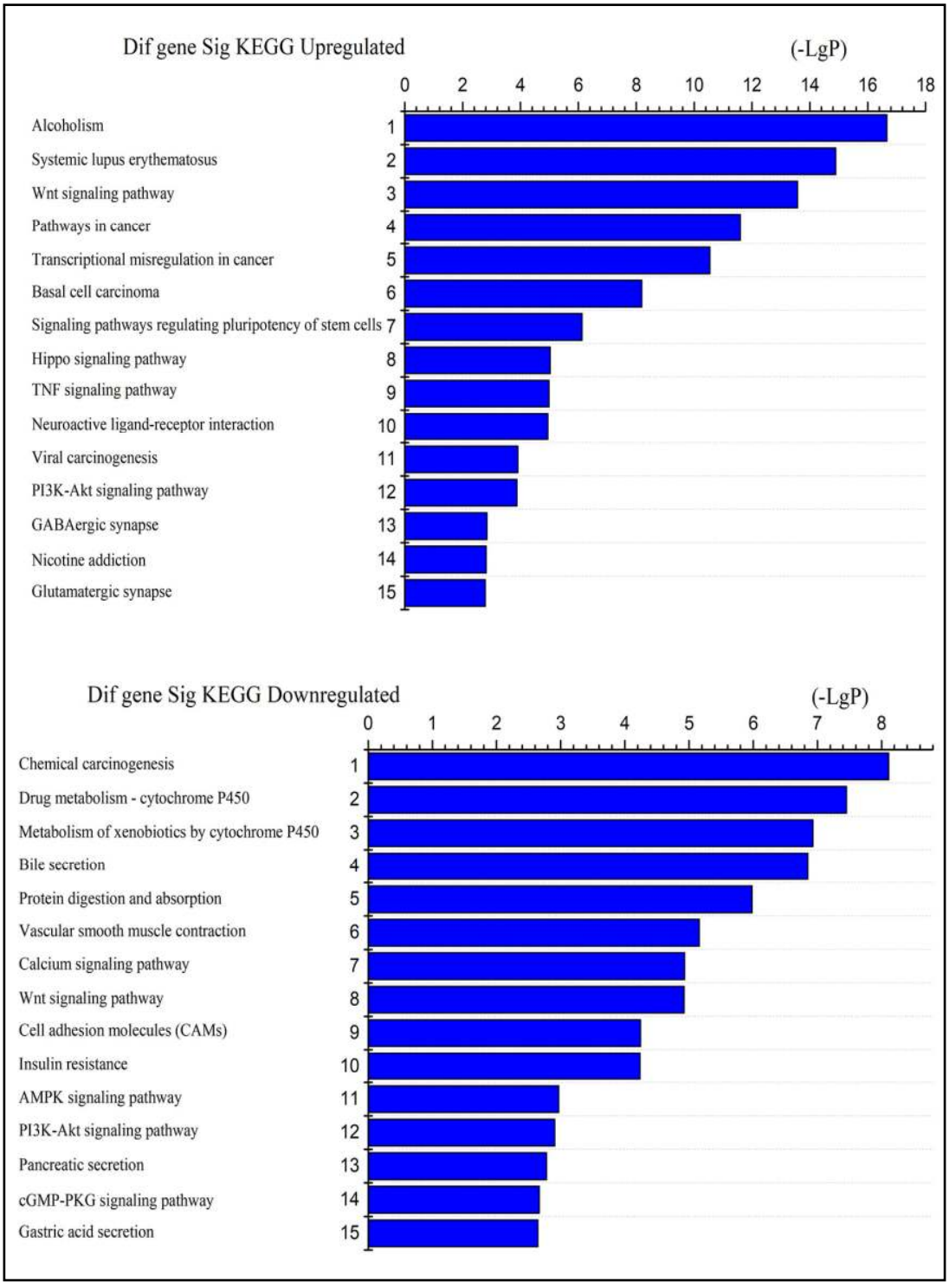

Fig. 5. Top 15 KEGG results of intersected upregulated and downregulated mRNAs respectively (-logP represents -log(P-value)).

non-tumor tissues (absolute logFC > 2, FDR < 0.01) and identified 419, 399, 378, and 427 differentially expressed lncRNAs, respectively. To enhance the reliability of the data, we selected 258 differentially expressed lncRNAs from intersections of the aforementioned 5 comparative groups for further analysis (Fig. 3A). Among these 258 specific lncRNAs, 64 lncRNAs were found to be involved in the lncRNA-miRNA-mRNA ceRNA network. Table 1 shows their logFC and FDR data based on results of comparison between the integrated tumor tissues and adjacent non-tumor tissues.

GO and pathway analysis of differentially expressed genes

To enhance the reliability and to understand the functions of the differentially expressed genes, we used intersected mRNAs for further analysis. A total of 2083 differentially expressed mRNAs were identified between the COAD patients' tumor tissue samples and the adjacent non-tumor tissue samples (absolute logFC > 2, FDR < 0.01). We identified 2107, 2098, 1996, and 2162 differentially expressed mRNAs between tumor tissues in patients with stage I, II, III, and IV cancer and adjacent non-tumor tissues, respectively (absolute logFC > 2, FDR < 0.01). We then chose the differentially expressed mRNAs in all 5 comparative groups, and 1467 differentially expressed mRNAs were finally selected for further study (Fig. 3B).

The 1467 differentially expressed mRNAs were divided into 2 groups (upregulated and downregulated) for analysis by DAVID 6.8 bioinformatics resources. We chose the top 15 GO biological process of upregulated and downregulated genes based on p-values, as shown in Fig. 4. The top 15 KEGG pathways of upregulated and downregulated genes are shown in Fig. 5. Among these pathways, the pI3K-Akt, AMPK, cell adhesion molecule (CAM), and Wnt signaling pathways are reported to be related to invasion and metastases of cancer [6-9]. Some other pathways are also known to be associated with cancers such as viral carcinogenesis, basal cell carcinoma, and transcriptional misregulation.

Prediction of miRNA targets and construction of the ceRNA network

We found 316 miRNAs that were differentially expressed between COAD tumor tissues and adjacent non-tumor tissues, and compared tumor tissues in patients with stage I, II, III, and IV cancer and adjacent non-tumor tissues. Next, we obtained 206 specific miRNAs by selecting the intersection of miRNA differentially expressed across the 5 comparative groups (absolute logFC > 2, FDR < 0.01) (Fig. 3C). We then studied whether these intersected miRNAs target the aforementioned 258 specific lncRNAs. Finally, 18 miRNAs were predicted to target 64 lncRNAs based on miRcode (Table 2).

Table 2. Specific miRNAs that may target specific lncRNAs

miRNAs	lncRNAs
hsa-mir-98	IGF2-AS, HECW1-IT1, WT1-AS, MUC19, LINC00488, FAM95B1, SPATA13-AS1, LINC00355, LMO7-AS1, ABCA9-AS1, JAZF1-AS1, LINC00484, ADAMTS9-AS2, KCNQ1OT1
hsa-mir-96	SHANK2-AS3, WT1-AS, MUC19, UCA1, ST7-OT4, LINC00488, FAM95B1, ERVMER61-1, TBL1XR1-AS1, ATP11A-AS1, GAS6-AS1, RBMS3-AS3, DLEU7-AS1, ADAMTS9-AS1, ADAMTS9-AS2, LIFR-AS1, LINC00461, MALAT1, NKX2-1-AS1, KCNQ1OT1, FRMD6-AS2
hsa-mir-454	H19, C20orf166-AS1, C15orf54, MUC19, E2F3-IT1, HOTAIR, ADAMTS9-AS1, ADAMTS9-AS2, NKX2-1-AS1, C8orf49, KCNQ1OT1
hsa-mir-424	C2orf48, C15orf54, HECW1-IT1, WT1-AS, MUC19, ST7-OT4, COL4A2-AS2, LINC00092, SFTA1P, STEAP2-AS1, LINC00355, ABCA9-AS1, DLX6-AS1, USP12-AS1, ATP11A-AS1, LINC00484, EGOT, DLEU7-AS1, LINC00461, PVT1, MALAT1, C8orf49, KCNQ1OT1
hsa-mir-338	IGF2-AS, H19, C2orf48, C15orf54, C17orf77, MUC19, ST7-OT4, FAM95B1, LPP-AS1, COL4A2-AS2, STEAP2-AS1, LINC00355, ERVMER61-1, WASIR2, DLX6-AS1, LINC00460, DSCAM-AS1, LINC00402, LINC00484, MYO16-AS1, DLEU7-AS1, ADAMTS9-AS2, LINC00461, CRNDE, MALAT1, C8orf49, KCNQ1OT1, FRMD6-AS2
hsa-mir-32	WT1-AS, MUC19, ST7-OT4, CLDN10-AS1, POU6F2-AS1, GAS6-AS1, JAZF1-AS1, LINC00484, ADAMTS9-AS2, LIFR-AS1, LINC00461, CRNDE, MALAT1, C8orf49, KCNQ1OT1
hsa-mir-223	C2orf48, C17orf77, WT1-AS, LINC00355, TBL1XR1-AS1, DLX6-AS1, GAS6-AS1, LINC00484, ADAMTS9-AS2, CRNDE, KCNQ1OT1
hsa-mir-217	WT1-AS, MUC19, BTBD9-AS1, HOTAIR, BOK-AS1, LINC00402, LINC00484, CRNDE, VCAN-AS1, PVT1, MALAT1, KCNQ1OT1
hsa-mir-21	LINC00488, HOTAIR, ERVMER61-1, JAZF1-AS1, EGOT, ADAMTS9-AS1, PVT1, MALAT1
hsa-mir-192	MUC19, LINC00488, POU6F2-AS1, ST7-AS2, HCG23, DLX6-AS1, ATP11A-AS1, DLEU7-AS1, LIFR-AS1, LINC00461, MALAT1, KCNQ1OT1
hsa-mir-182	C15orf54, WT1-AS, MUC19, UCA1, ST7-OT4, FAM95B1, SFTA1P, ERVMER61-1, TBL1XR1-AS1, GAS6-AS1, LINC00402, RBMS3-AS3, DLEU7-AS1, ADAMTS9-AS1, ADAMTS9-AS2, LIFR-AS1, MALAT1, NKX2-1-AS1, KCNQ1OT1, FRMD6-AS2
hsa-mir-17	IGF2-AS, H19, C2orf48, C20orf166-AS1, C17orf77, WT1-AS, MUC19, LMO7-AS1, HOTAIR, HCG23, DLX6-AS1, JAZF1-AS1, LINC00402, VCAN-AS1, PVT1, MALAT1, NKX2-1-AS1, C8orf49, KCNQ1OT1
hsa-mir-152	H19, MUC19, FAM95B1, COL4A2-AS2, STEAP2-AS1, HOTAIR, DLX6-AS1, ATP11A-AS1, LINC00484, ADAMTS9-AS2, PVT1, MALAT1, NKX2-1-AS1, C8orf49, KCNQ1OT1
hsa-mir-150	IGF2-AS, C2orf48, SHANK2-AS3, C20orf166-AS1, C15orf54, HECW1-IT1, C17orf77, MUC19, ST7-OT4, CLDN10-AS1, FAM95B1, COL4A2-AS2, TSSC1-IT1, LINC00092, MRPL23-AS1, BTBD9-AS1, LINC00355, LMO7-AS1, HOTAIR, WASIR2, DLX6-AS1, LINC00460, BOK-AS1, JAZF1-AS1, DSCAM-AS1, LINC00402, ADAMTS9-AS1, ADAMTS9-AS2, LIFR-AS1, LINC00461, PVT1, HULC, MALAT1, C8orf49, KCNQ1OT1
hsa-mir-144	MUC19, LINC00488, POU6F2-AS1, DLX6-AS1, ADAMTS9-AS1, ADAMTS9-AS2, LIFR-AS1, LINC00461, CRNDE, MALAT1
hsa-mir-143	C2orf48, C15orf54, HECW1-IT1, C17orf77, MUC19, UCA1, CLDN10-AS1, FAM95B1, LPP-AS1, E2F3-IT1, TSSC1-IT1, SFTA1P, MRPL23-AS1, STEAP2-AS1, HOTAIR, ATP11A-AS1, LINC00460, JAZF1-AS1, DSCAM-AS1, LINC00402, LINC00484, EGOT, ADAMTS9-AS2, LINC00461, CRNDE, PVT1, MALAT1, C8orf49, KCNQ1OT1, FRMD6-AS2, LINC00524
hsa-mir-141	H19, WT1-AS, ST7-OT4, FAM95B1, TSSC1-IT1, LINC00355, DLX6-AS1, DSCAM-AS1, LINC00402, LINC00484, EGOT, ADAMTS9-AS2, LINC00461, VCAN-AS1, MALAT1, KCNQ1OT1
hsa-mir-106a	C2orf48, C20orf166-AS1, C15orf54, C17orf77, WT1-AS, MUC19, LMO7-AS1, HCG23, DLX6-AS1, ATP11A-AS1, JAZF1-AS1, LINC00484, ADAMTS9-AS2, LIFR-AS1, LINC00461, VCAN-AS1, PVT1, MALAT1, C8orf49, LINC00507, KCNQ1OT1

In order to construct the ceRNA network, we used 18 miRNAs mentioned in Table 2 to predict mRNAs using Targetscan, miRdb, and miRTarBase. We then compared the predicted mRNAs and 1487 specific mRNAs and chose the mRNAs that exist in both groups. Finally, 42 mRNAs were found to interact with 18 miRNAs (Table 3). Among these mRNAs, some were validated to be cancer-related genes, such as TPM2, KLF4, and EPHA7 [10-12].

Based on information shown in Tables 2 and 3, we constructed an miRNA-lncRNA-mRNA ceRNA network using Cytoscape 3.0. Eighteen miRNAs, 42 mRNAs, and 64 lncRNAs were involved in the network (Supplementary Fig. S1 - For all supplemental material see www.karger.com/10.1159/000493623/). We designated the lncRNAs involved in the net as key lncRNAs.

Key lncRNA clinical feature analysis

To further study the lncRNAs, we analyzed the correlation between lncRNAs that were involved in the ceRNA network, and clinical features including sex, tumor stage, TNM stage, and lymphatic metastasis and distant metastasis status based on data from TCGA. Three lncRNAs were found to be associated with clinical features (absolute logFC > 1.5, FDR < 0.05). Results showed that linc00355 was associated with tumor stage, lymphatic metastasis, and distant metastasis, HULC was associated with tumor stage and lymphatic metastasis, and IGF2-AS was associated with distant metastasis (Table 4).

We also analyzed the OS of the 64 key lncRNAs to study the prognostic characteristics based on the data from TCGA. As demonstrated by the Kaplan-Meier survival curves shown in Fig. 6A, HOTAIR, LINC00355, KCNQ1OT1, and TSSC1-IT1 were negatively associated with OS (log-rank $p < 0.05$).

Prediction and analysis of mRNA targeting 4 lncRNAs related to survival

The top 20 potential targeted mRNAs of each lncRNA based on the correlation rate between lncRNAs and mRNAs are shown in Table 5. We analyzed the biological processes in GO and KEGG based on all mRNAs targeted by the 4 lncRNAs (Table 6). Results from GO showed that the 4 lncRNAs may be involved in RNA transcription and translation. The MAPK signaling pathway, which plays a vital role in cancer, was shown to be involved in KEGG [13] and, in our study, it indicated these 4 lncRNAs may act through the MAPK signaling pathway.

qRT-PCR validation

To verify the bioinformatics analysis results, we investigated the expression patterns of 4 lncRNAs related to OS by qRT-PCR of tissue samples from 50 COAD patients. The logFC results of these 4 lncRNAs obtained by TCGA analysis and qRT-PCR are shown in Fig. 6B. We confirmed the correlation between 3 lncRNAs related to clinical features and their expression patterns, and the 50 patients were divided into 2 groups according to the expression levels of these lncRNAs. Patients with expression of lncRNAs that was either higher or lower than the median were allocated to high- or low-expression groups, respectively. The results are shown in Table 7.

Table 3. Specific miRNAs that may target specific mRNAs

miRNAs	mRNAs
hsa-mir-106a	FAM129A, CADM2, CFL2, FOXQ1, FJX1
hsa-mir-141	PHLPP2, MACC1, EPHA7, ELAVL4, KIAA1549
hsa-mir-143	COL1A1
hsa-mir-144	GRIK3
hsa-mir-150	HILPDA, EREG, SLC7A11
hsa-mir-152	NPTX1, BMP3, KLF4
hsa-mir-17	FJX1, FOXQ1, CFL2, FAM129A, CADM2, SLC16A9
hsa-mir-182	NPTX1, CHL1, ULBP2
hsa-mir-21	OSR1, TGFBI
hsa-mir-217	DACH1
hsa-mir-223	EPB41L3
hsa-mir-32	UGP2, PHLPP2, PAX9, PBLD
hsa-mir-338	NOVA1
hsa-mir-424	TPM2, AXIN2, PHLPP2, CBX2, PSAT1, TMEM100
hsa-mir-454	CFL2
hsa-mir-96	TRIB3
hsa-mir-98	IGF2BP1, CPA4, PRSS22, IGF2BP3, TRIM71, HAND1
hsa-mir-192	GRHL1

Table 4. Correlation between COAD key lncRNAs and their clinical features. LnCRNAs correlated with clinical features with absolue logFC>1.5, FDR <0.05

Comparisons	Upregulated	Downregulated
Tumor stage (Stage III, IVvs. Stage I, II)	LINC00355, HULC	
Lymphatic metastasis (Yes vs. No)	LINC00355, HULC	
Distant metastasis (Yes vs. No)	LINC00355, IGF2-AS	

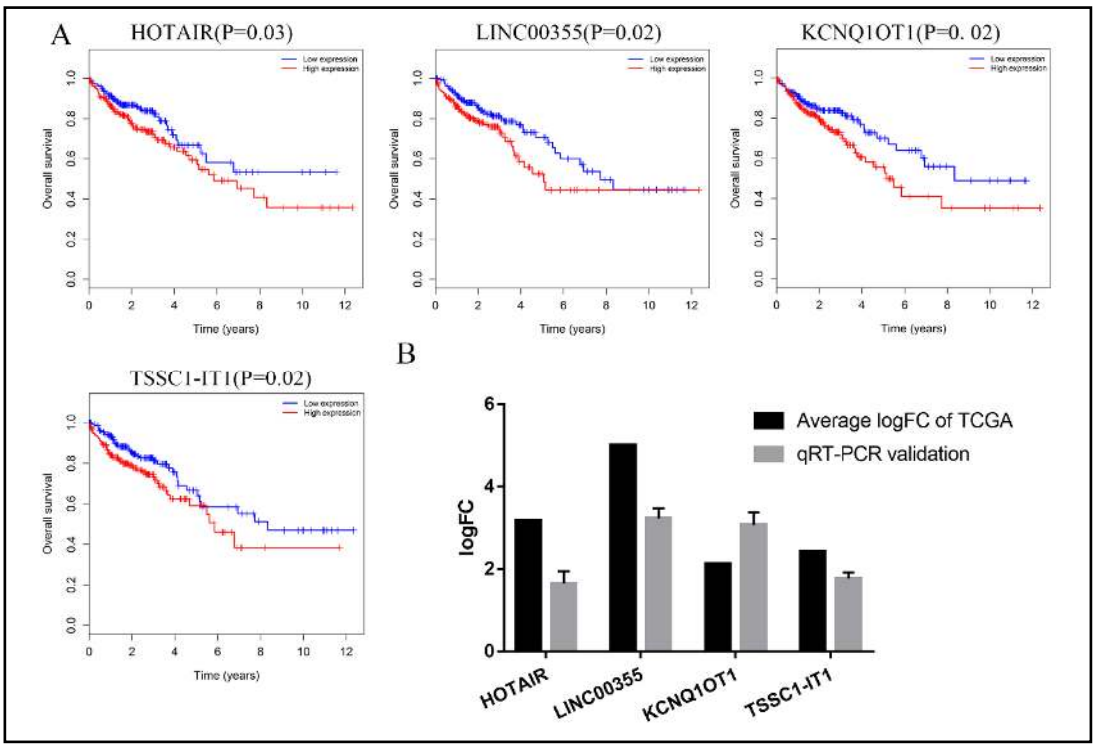

Fig. 6. Kaplan-Meier survival curves for 4 lncRNAs associated with overall survival (A). Horizontal axis, overall survival time, years; vertical axis, survival function. Correlation between bioinformatics results and qRT-PCR results (B). Comparison of logFC (-ΔΔCt) of lncRNAs between TCGA and qRT-PCR results. LogFC represents $\log_2$(fold change).

Table 5. Top 20 mRNA targets of each key lncRNAs related to OS

lncRNAs	mRNA targets
HOTAIR	REV3L, INO80D, PHIP, AGO3, TAOK1, ATM, GPATCH2L, RC3H1, SHPRH, RICTOR, MAP3K2, ATAD2B, RIC1, PIKFYVE, WDPCP, USP34, PHC3, MIR133A1HG, BRWD1, ZBTB37
LINC00355	ATM, REV3L, GPATCH2L, RICTOR, SHPRH, PHIP, AGO3, WDPCP, INO80D, RIC1, RC3H1, TAOK1, USP34, ATAD2B, MFSD4B, FAM217A, PIKFYVE, ZBTB37, NBEAL1, GDAP2
TSSC1-IT1	SHPRH, INO80D, LRRTM2, TNRC6B, GPATCH2L, PHC3, GPR52, RC3H1, AGO3, ZBTB20, REV3L, NBEAL1, RNF169, TAS2R19, MFSD4B, IFNK, CLDN20, RIC1, FBXL13, MATR3
KCNQ1OT1	SHPRH, INO80D, AGO3, LRRTM2, RC3H1, MIR133A1HG, GPATCH2L, REV3L, MATR3, ZBTB20, PHC3, GPR52, GDAP2, POU5F2, MFSD4B, NBEAL1, TAS2R19, ATM, C1orf195, AIRN

Discussion

CRC is the third most common cancer in the world, and it has the fourth highest cancer mortality rate [14]. COAD is a frequently observed type of CRC. Despite great progress that has been made in the treatment of CRC through advancements in medical science and technology, surgical techniques, and chemotherapy, the mortality rate of COAD remains high [15], which may be due to insufficient understanding of the underlying mechanisms and a lack of efficient biomarkers. Previous studies have shown that dysregulated genes may have important functions in cancer [16, 17] and have great potential as biomarkers. To better understand and identify efficient new biomarkers of COAD, we are studying the mechanisms of COAD. Recent studies have shown that lncRNAs play vital roles in the development of cancer [2, 3, 18-21], but there have been very few studies that focused on the profiles of lncRNAs in CRC.

Table 6. GO and KEGG about mRNAs targets of key lncRNA related to OS

Items	-LogP
GO	
Sensory perception of taste	5.29
Detection of chemical stimulus involved in sensory perception of bitter taste	4.72
Posttranscriptional gene silencing by RNA	1.56
MiRNA mediated inhibition of translation	1.33
KEGG	
Taste transduction	5.15
MAPK signaling pathway	1.45

Table 7. Expression of lncRNAs related to clinical features according to patients' clinicopathological characteristics

Characteristics	Number	LINC00355 expression		P-value	HULC expression		P-value
		High group	Low group		High group	Low group	
Gender							
Female	27	15	12	0.39	14	13	0.78
Male	23	10	13		11	12	
Lymphatic metastasis							
No	27	8	19	0.001	9	18	0.01
Yes	23	17	6		16	7	
Tumor stage							
Stage Ⅰ, Ⅱ	24	8	16	0.02	8	16	0.02
Stage Ⅲ Ⅳ	26	17	9		17	9	
TNM staging system							
T1+T2	29	17	12	0.15	14	15	0.77
T3+T4	21	8	13		11	10	
Distant metastasis							
No	32	12	20	0.02	18	14	0.24
Yes	18	13	5		7	11	

Characteristics	Number	IGF2-AS expression		P-value
		High group	Low group	
Gender				
Female	27	13	14	0.78
Male	23	12	11	
Lymphatic metastasis				
No	27	15	12	0.39
Yes	23	10	13	
Tumor stage				
Stage Ⅰ, Ⅱ	24	11	13	0.57
Stage Ⅲ Ⅳ	26	14	12	
TNM staging system				
T1+T2	29	12	17	0.15
T3+T4	21	13	8	
Distant metastasis				
No	32	12	20	0.02
Yes	18	13	5	

LncRNAs have sophisticated functions through diverse pathways, and the ceRNA hypothesis makes the relationship between lncRNAs, miRNAs, and mRNAs more complicated. The ceRNA hypothesis suggests a novel regulatory mechanism that can be mediated by lncRNAs [4]. Previous studies have revealed several potential ceRNAs in CRC, but the ceRNA network of COAD has not been described clearly.

In this study, we identified differentially expressed lnRNAs, mRNAs, and miRNAs between integrated tumor tissues and adjacent non-tumor tissues according to data from TCGA. We then compared the expression profile between adjacent non-tumor tissues and COAD tumor tissues in patients with stage I, II, III, and IV cancer. Intersections were selected from the 5 comparative groups for further analysis. Through GO and KEGG, we further analyzed the functions and pathways involving the differentially expressed mRNAs. Then, by utilizing bioinformatics tools, we constructed a ceRNA network with COAD-specific mRNAs, miRNAs, and lncRNAs. We analyzed correlations between lncRNAs that were involved in the ceRNA network with several clinical features and ultimately identified 4 lncRNAs from the network to be correlated with OS. Our findings were validated by qRT-PCR conducted on COAD tissues taken from 50 patients.

Using GO and KEGG, we analyzed the COAD-specific mRNAs. Based on the hypothesis of ceRNA, lncRNAs can be mediated by mRNAs; thus, the specific lncRNAs may also function or concentrate on the potential pathways in a manner similar to mRNAs. The results for the GO biological process showed that specific genes may focus on several domains such as cellular functions, metabolism, and immune functions. Some pathways that appeared in the KEGG analysis have been previously reported to be associated with cancer. PI3K/AKT signaling is involved in the processes of downregulating apoptosis and stimulating cell growth and proliferation. Usually, activation of PI3K/AKT is regulated by both extracellular and intracellular growth signals [22]. Li et al. demonstrated that the AMPK pathway is involved in the processes of tumor invasion and migration; specifically, that liver kinase B1 phosphorylates and activates AMPK and further reduces the cancer cell proliferation and metabolism [23]. It has been reported that CAMs play important roles in the process of metastasis [8], and that miR-612 can suppress the stemness of liver cancer via the Wnt pathway [9].

We used specific miRNAs, lncRNAs, and mRNAs to construct a ceRNA network using bioinformatics tools. This network contains key miRNAs, mRNAs, and lncRNAs, and we demonstrated their interactions. Moreover, this network can help explain the COAD at the genetic level. Several interactions have already been confirmed previously. For example, H19 can interact with miR-141 and regulate cell proliferation and migration in gastric cancer [24], and UCA1 regulates miR-143 and further modulates breast cancer cell growth and apoptosis [25]. Several lncRNAs from the network have been verified to mediate CRC. Yang et

al. discovered that HULC promotes CRC progression through epigenetic repression of NKD2 expression [26]. CRNDE promotes CRC cell and chemoresistance via miR-181a-5p-mediated regulation of Wnt/β-catenin signaling [27]. H19, UCA1, PVT1, and MALAT1 are also involved in the progression of CRC [28-31]. These studies also confirm that our analytical results are credible. LncRNAs from the ceRNA network were selected in order to analyze their relationship with OS. Four lncRNAs were shown to have significant effects on OS: HOTAIR, LINC00355, KCNQ1OT1, and TSSC1-IT1 were negatively associated with OS. HOTAIR is a well-known lncRNA that has been reported to show involvement in various types of cancer [3], although LINC00355, KCNQ1OT1, and TSSC1-IT1 still need to be investigated in CRC and other cancers. We further predicted the mRNA targets of 4 lncRNAs in order to analyze the potential of GO and KEGG. These 4 lncRNAs were shown to potentially function in RNA transcription and translation. The MAPK signaling pathway, which has been reported to play crucial roles in CRC, is shown in KEGG [13]. The 4 identified lncRNAs may function through the MAPK signaling pathway and may play key roles in cancer.

We chose the lncRNAs from the ceRNA network to analyze the correlations with several clinical features such as sex, TNM stage, pathologic stage, lymphatic metastasis, and distant metastasis. Among these key lncRNAs, HULC and LINC00355 appeared often. HULC was validated to perform important roles in CRC [26], and further studies are required to understand the functions of LINC00355.

Finally, in order to validate our bioinformatic analysis, we selected 4 lncRNAs related to OS to detect their expressions in 50 paired tumor tissues and adjacent non-tumor tissues from COAD patients by qRT-PCR. We could confirm the correlations between 3 lncRNAs related to clinical features and their expression patterns through qRT-PCR. These results were highly consistent and further support the credibility of our analysis.

Conclusion

In conclusion, we constructed a ceRNA network with COAD-specific lncRNAs, miRNAs, and mRNAs, and we identified key COAD lncRNAs by bioinformatics analysis and studied their correlations with clinical features based on data from TCGA. To our knowledge, lncRNA profiling from such large-scale samples are rare. Our findings provide a method for identifying potential lncRNAs biomarkers. In addition, we revealed the potential ceRNA network in COAD to help understand the mechanism of COAD at the genetic level.

Acknowledgements

The work is funded by Jiangsu Key Medical Discipline (General Surgery) (ZDXKA2016005).

Disclosure Statement

The authors declare no conflicts of interest.

References

1 Jathar S, Kumar V, Srivastava J, Tripathi V: Technological Developments in lncRNA Biology. Adv Exp Med Biol 2017;1008:283-323.

2 Li J, Wang X, Tang J, Jiang R, Zhang W, Ji J, Sun B: HULC and Linc00152 Act as Novel Biomarkers in Predicting Diagnosis of Hepatocellular Carcinoma. Cell Physiol Biochem 2015;37:687-696.

3 Xu CZ, Jiang C, Wu Q, Liu L, Yan X, Shi R: A Feed-Forward Regulatory Loop between HuR and the Long Noncoding RNA HOTAIR Promotes Head and Neck Squamous Cell Carcinoma Progression and Metastasis. Cell Physiol Biochem 2016;40:1039-1051.
4 Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 2011;146:353-358.
5 Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-408.
6 Park JH, Kim JJ, Bae YS: Involvement of PI3K-AKT-mTOR pathway in protein kinase CKII inhibition-mediated senescence in human colon cancer cells. Biochem Biophys Res Commun 2013;433:420-425.
7 Schuster S, Penke M, Gorski T, Gebhardt R, Weiss TS, Kiess W, Garten A: FK866-induced NAMPT inhibition activates AMPK and downregulates mTOR signaling in hepatocarcinoma cells. Biochem Biophys Res Commun 2015;458:334-340.
8 Saadatmand S, de Kruijf EM, Sajet A, Dekker-Ensink NG, van Nes JG, Putter H, Smit VT, van de Velde CJ, Liefers GJ, Kuppen PJ: Expression of cell adhesion molecules and prognosis in breast cancer. Br J Surg 2013;100:252-260.
9 Tang J, Tao ZH, Wen D, Wan JL, Liu DL, Zhang S, Cui JF, Sun HC, Wang L, Zhou J, Fan J, Wu WZ: MiR-612 suppresses the stemness of liver cancer via Wnt/beta-catenin signaling. Biochem Biophys Res Commun 2014;447:210-215.
10 Dube S, Thomas A, Abbott L, Benz P, Mitschow C, Dube DK, Poiesz BJ: Expression of tropomyosin 2 gene isoforms in human breast cancer cell lines. Oncol Rep 2016;35:3143-3150.
11 Zhu LF, Chen QR, Chen SZ, Wang LY, Luo XF, Ren JH, Yuan XH, Wu XQ, Zeng YL, Xiao M, Chen YQ, Chen YY, Lin MH, Wu ZJ, Chen ZZ, Hu JD, Yang T: The Construction and Identification of Induced Pluripotent Stem Cells Derived from Acute Myelogenous Leukemia Cells. Cell Physiol Biochem 2017;41:1661-1674.
12 Xiang C, Lv Y, Wei Y, Wei J, Miao S, Mao X, Gu X, Song K, Jia S: Effect of EphA7 Silencing on Proliferation, Invasion and Apoptosis in Human Laryngeal Cancer Cell Lines Hep-2 and AMC-HN-8. Cell Physiol Biochem 2015;36:435-445.
13 Ahronian LG, Sennott EM, Van Allen EM, Wagle N, Kwak EL, Faris JE, Godfrey JT, Nishimura K, Lynch KD, Mermel CH, Lockerman EL, Kalsy A, Gurski JM, Jr., Bahl S, Anderka K, Green LM, Lennon NJ, Huynh TG, Mino-Kenudson M, Getz G, et al.: Clinical Acquired Resistance to RAF Inhibitor Combinations in BRAF-Mutant Colorectal Cancer through MAPK Pathway Alterations. Cancer Discov 2015;5:358-367.
14 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A: Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
15 Issa IA, Noureddine M: Colorectal cancer screening: An updated review of the available options. World J Gastroenterol 2017;23:5086-5096.
16 Eldai H, Periyasamy S, Al Qarni S, Al Rodayyan M, Muhammed Mustafa S, Deeb A, Al Sheikh E, Afzal M, Johani M, Yousef Z, Aziz MA: Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner. PLoS One 2013;8:e76251.
17 Li X, Peng S, Chen J, Lu B, Zhang H, Lai M: SVM-T-RFE: a novel gene selection algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles. Biochem Biophys Res Commun 2012;419:148-153.
18 Zhang W, Yuan W, Song J, Wang S, Gu X: LncRna CPS1-IT1 Suppresses Cell Proliferation, Invasion and Metastasis in Colorectal Cancer. Cell Physiol Biochem 2017;44:567-580.
19 Hu CE, Du PZ, Zhang HD, Huang GJ: Long Noncoding RNA CRNDE Promotes Proliferation of Gastric Cancer Cells by Targeting miR-145. Cell Physiol Biochem 2017;42:13-21.
20 Yu B, Ye X, Du Q, Zhu B, Zhai Q, Li XX: The Long Non-Coding RNA CRNDE Promotes Colorectal Carcinoma Progression by Competitively Binding miR-217 with TCF7L2 and Enhancing the Wnt/beta-Catenin Signaling Pathway. Cell Physiol Biochem 2017;41:2489-2502.
21 Xu J, Zhang R, Zhao J: The Novel Long Noncoding RNA TUSC7 Inhibits Proliferation by Sponging MiR-211 in Colorectal Cancer. Cell Physiol Biochem 2017;41:635-644.
22 Danielsen SA, Eide PW, Nesbakken A, Guren T, Leithe E, Lothe RA: Portrait of the PI3K/AKT pathway in colorectal cancer. Biochim Biophys Acta 2015;1855:104-121.
23 Li N, Huang D, Lu N, Luo L: Role of the LKB1/AMPK pathway in tumor invasion and metastasis of cancer cells (Review). Oncol Rep 2015;34:2821-2826.

24 Zhou X, Ye F, Yin C, Zhuang Y, Yue G, Zhang G: The Interaction Between MiR-141 and lncRNA-H19 in Regulating Cell Proliferation and Migration in Gastric Cancer. Cell Physiol Biochem 2015;36:1440-1452.
25 Tuo YL, Li XM, Luo J: Long noncoding RNA UCA1 modulates breast cancer cell growth and apoptosis through decreasing tumor suppressive miR-143. Eur Rev Med Pharmacol Sci 2015;19:3403-3411.
26 Yang XJ, Huang CQ, Peng CW, Hou JX, Liu JY: Long noncoding RNA HULC promotes colorectal carcinoma progression through epigenetically repressing NKD2 expression. Gene 2016;592:172-178.
27 Han P, Li JW, Zhang BM, Lv JC, Li YM, Gu XY, Yu ZW, Jia YH, Bai XF, Li L, Liu YL, Cui BB: The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/beta-catenin signaling. Mol Cancer 2017;16:9.
28 Han Y, Yang YN, Yuan HH, Zhang TT, Sui H, Wei XL, Liu L, Huang P, Zhang WJ, Bai YX: UCA1, a long non-coding RNA up-regulated in colorectal cancer influences cell proliferation, apoptosis and cell cycle distribution. Pathology 2014;46:396-401.
29 Takahashi Y, Sawada G, Kurashige J, Uchi R, Matsumura T, Ueo H, Takano Y, Eguchi H, Sudo T, Sugimachi K, Yamamoto H, Doki Y, Mori M, Mimori K: Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. Br J Cancer 2014;110:164-171.
30 Ohtsuka M, Ling H, Ivan C, Pichler M, Matsushita D, Goblirsch M, Stiegelbauer V, Shigeyasu K, Zhang X, Chen M, Vidhu F, Bartholomeusz GA, Toiyama Y, Kusunoki M, Doki Y, Mori M, Song S, Gunther JR, Krishnan S, Slaby O, et al.: H19 Noncoding RNA, an Independent Prognostic Factor, Regulates Essential Rb-E2F and CDK8-beta-Catenin Signaling in Colorectal Cancer. EBioMedicine 2016;13:113-124.
31 Xu C, Yang M, Tian J, Wang X, Li Z: MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. Int J Oncol 2011;39:169-175.