ONCOLOGY LETTERS 18: 3137-3149, 2019

Discovery of core genes in colorectal cancer by weighted gene co-expression network analysis

CUN LIAO[1*], XUE HUANG[2*], YIZHEN GONG[1] and QIUNING LIN[3]

[1]Department of Colorectal and Anal Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530000; [2]Department of Gastroenterology, Eighth Affiliated Hospital of Guangxi Medical University, Guigang, Guangxi Zhuang Autonomous Region 537100; [3]Department of Vascular Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530000, P.R. China

Received January 11, 2019; Accepted May 29, 2019

DOI: 10.3892/ol.2019.10605

Abstract. The aim of the present study was to investigate the interactions among messenger RNAs (mRNAs), microRNAs (miRNAs), and long noncoding RNAs (lncRNAs) in colorectal cancer (CRC), in order to examine its underlying mechanisms. The raw gene expression data was downloaded from the Gene Expression Omnibus (GEO) database. An online tool, GEO2R, which is based on the limma package, was used to identify differentially expressed genes. The co-expression between lncRNAs and mRNAs was identified utilizing the weighted gene co-expression analysis package of R to construct a coding non-coding (CNC) network. The function of the genes in the CNC network was determined by performing Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathways enrichment analysis. The interactions among miRNAs, mRNAs and lncRNAs were predicted using Lncbase and mirWalk to construct the competing endogenous RNA (ceRNA) network. The expression of the genes involved in the ceRNA network was further validated in The Cancer Genome Atlas dataset. A total of 3,183 dysregulated mRNAs, 78 dysregulated miRNAs and 2,248 dysregulated lncRNAs were screened in two GEO datasets. Combined with the results of the dysregulated genes, 169 genes were selected to construct the CNC network. 'p53 signaling pathway' and the 'cell cycle' were the most significant enriched pathways in the genes involved in the CNC network. Finally, a validated ceRNA network composed of 2 lncRNAs (MIR22HG and RP11-61I13.3), 5 miRNAs (hsa-miR-765, hsa-miR-198, hsa-miR-125a-3p, hsa-miR-149-3p and hsa-miR-650) and 5 mRNAs (ANK2, BTK, GBP2, PCSK5 and PDK4) was obtained. In conclusion, MIR22HG may regulate PCSK5, BTK and PDK4, and RP11-61I13.3 may regulate the ANK2, GBP2, PCSK5 through sponging miRNAs to act on the progression of CRC, and the potential function of these genes have been revealed. However, the diagnostic and prognostic value of these genes requires further validation.

Correspondence to: Dr Qiuning Lin, Department of Vascular Surgery, The First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning, Guangxi Zhuang Autonomous Region 530000, P.R. China
E-mail: linqiuning@126.com

*Contributed equally

Abbreviations: miRNAs, microRNAs; lncRNAs, long non-coding RNAs; CRC, colorectal cancer; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; DEMs, differentially expressed miRNAs; WGCNA, weighted gene co-repression analysis; CNC, coding non-coding; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ceRNA, competing endogenous RNA; TCGA, The Cancer Genome Atlas; FC, fold-change; HCC, hepatocellular carcinoma; COAD, colon adenocarcinoma; PCSK5, proprotein convertase subtilisin/kexin type 5

Key words: colorectal cancer, weighted gene co-repression analysis, competing endogenous RNA, bioinformatics analysis, The Cancer Genome Atlas

Introduction

Colorectal cancer (CRC) is one of the most frequently diagnosed cancers of the digestive system. It was estimated that there would be 97,220 new cases and 50,630 CRC-related deaths in the United States in 2018 (1). Much of the rising burden is attributable to population growth and aging, as well as sociodemographic changes. Recurrence is the major cause of CRC-related death (2).

Despite the diagnosis and treatment of CRC, the survival of the patient is closely associated to the stage of the tumor at diagnosis, and 40-50% of patient mortality is due to distant metastasis (3,4). Considering the enormous threat of CRC to human health, new diagnostic and therapeutic approaches are required for early cancer detection and effective treatment (5).

In order to find a more effective treatment for CRC, a more thorough understanding of its pathogenesis is important. Previously, there has been increasing evidence that microRNAs (miRNAs) are functional in cancer progression by downregulating their targets, including mRNA, long non-coding RNA (lncRNA), circular RNA and pseudogenes (6-8). Nevertheless,

Table I. Clinical information of the patients included in the GSE109454 dataset (18).

Sex	Age, years	TNM stage (22)	CEA, ng/ml
Male	55	II	12.43
Male	61	IVA	37.84
Female	56	IIIB	23.60
Male	59	IIIA	39.93
Female	53	II	15.51
Female	63	IIIA	18.89

CEA, carcinoembryonic antigen; TNM, Tumor-Node-Metastasis.

overexpression of these targets could abolish the downregulatory effect of miRNA in turn (9-11). Moreover, more than one miRNA target could compete with miRNAs as competing endogenous RNA (ceRNA) and overexpression of one ceRNA could indirectly upregulate other ceRNAs (12,13).

ceRNA crosstalk was first proposed by Poliseno *et al* (14). Phosphatase and tensin homolog pseudogene 1 (PTENP1) could upregulate PTEN by sponging their common miRNAs. It was subsequently confirmed that various genes could participate in the development of cancer through ceRNA mechanisms (15). In breast cancer, overexpression of NEAT1 induced by BRCA1-deficiency can silence hsa-miR-129-5p and upregulate WNT4, a target of hsa-miR-129-5p, which leads to activation of oncogenic WNT signaling (16). Liang *et al* (17) reported that the oncogenic functions of lncRNA H19 in CRC may be attributed to its ceRNA activity to sequester miR-138 and miR-200a and therefore, upregulate the expression levels of VIM, ZEB1 and ZEB2, the critical genes involved in epithelial mesenchymal transition. With the development of gene sequencing technology, it has been shown that the expression of various lncRNAs is either upregulated or downregulated in CRC tissues to regulate several signaling pathways, such as the Wnt, p13K/Akt and Ras pathways, through ceRNA crosstalk (18,19). Therefore, it has been proven that ceRNA crosstalk is a critical mechanism for the complex pathogenesis and multi-step development of CRC, which may be a potential starting point for the development of novel CRC treatment methods, which deserves further study. The aim of the present study, was to identify a ceRNA network between non-coding RNAs and coding genes in CRC and to reveal the potential mechanism of CRC progression, which may aid in establishing a novel clinical CRC treatment system.

Materials and methods

Gene datasets and clinical information. Raw gene expression data of the GSE109454 (20) and GSE41655 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41655) datasets were downloaded from the National Center of Biotechnology Information Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds/). In the GSE109454 dataset, tumor samples and adjacent non-tumor samples were obtained from 6 patients with CRC. Clinical information of these CRC patients from the original study is shown in Table I. In the GSE41655 dataset, there were 15 normal colorectal samples, 59 colorectal adenoma samples and 33 colorectal adenocarcinoma samples. Clinical information of these 107 cases from the original study is shown in Table II. Histological tumor type and grade were evaluated according to the World Health Organization cancer classification (21) and tumor stage according to the Union for International Cancer Control TNM classification (22).

Identification of differentially expressed genes (DEGs). GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) is an online tool provided by GEO. GEO2R is based on the R language limma package (v3.26.8) (23). GEO2R was used to screen DEGs between tumor and normal samples in the GSE41655 and GSE109454 datasets. At the same time, using false discovery rate correction, multiple t-tests were used to determine the statistical significance of the DEGs. $|\log_2$ fold change (FC)$|>1$ and a P-value <0.05 were set as the cut-off criteria. Subsequently, probes without a corresponding gene symbol were filtered.

Weighted gene co-repression analysis (WGCNA). To screen significant co-expression modules, WGCNA was performed as previously described (24,25). Probe sets were first filtered based on the variance of expression value across all samples. Probe sets with repeating gene symbols were also removed based on the expression variance. The R package WGCNA (v1.61) (26) was used to construct the co-expression networks. Independent signed networks were constructed from the CRC samples and normal samples. The adjacency matrix was constructed using a soft threshold power of 12 to make the soft threshold >0.8. Network interconnectedness was measured by calculating the topology overlap using the TOMdist function with a signed TOM-Type. The average hierarchical clustering was performed to group the genes based on the topological overlap dissimilarity measure (1-TOM) of their connection strengths. The network modules were identified using a dynamic tree cut algorithm, with a minimum cluster size of 30 and a merge threshold function of 0.25. Genes that were not assigned to particular modules were designated as grey. The module that had the strongest association with CRC was selected for further analysis and the co-expression in this module was filtered as follows: i) The weight score of co-expression >0.3; ii) only the coding and non-coding co-expression relationships were retained; iii) only the genes that were differentially expressed were retained; and iv) only the genes in one pair that had the same expression tendency were retained.

Gene function analysis. To determine the function of the selected genes in CRC, Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for the genes in the co-expression network were implemented using DAVID (v6.8; https://david.ncifcrf.gov/). In the present study, pathway and process enrichment analysis was performed using the following ontological sources: GO Biological Processes (BPs) and KEGG Pathway. All genes in the *H. sapiens* genome were used as background in the enrichment analysis. The P-value was calculated based

Table II. Clinical information of the patients included in the GSE41655 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41655).

Sex	Age, years	Pathological grade (21)	pTNM stage (22)
Male	35	Normal epithelium	
Male	60	Normal epithelium	
Male	53	Normal epithelium	
Male	68	Normal epithelium	
Female	61	Normal epithelium	
Female	55	Normal epithelium	
Male	29	Normal epithelium	
Male	37	Normal epithelium	
Female	38	Normal epithelium	
Female	22	Normal epithelium	
Female	62	Normal epithelium	
Male	69	Normal epithelium	
Female	46	Normal epithelium	
Male	51	Normal epithelium	
Female	39	Normal epithelium	
Male	49	Low-grade dysplasia	
Female	51	Low-grade dysplasia	
Male	64	High-grade dysplasia	
Male	69	Low-grade dysplasia	
Male	54	Low-grade dysplasia	
Male	60	Low-grade dysplasia	
Male	61	Low-grade dysplasia	
Female	74	Low-grade dysplasia	
Male	64	Low-grade dysplasia	
Male	61	High-grade dysplasia	
Female	70	High-grade dysplasia	
Male	83	High-grade dysplasia	
Female	53	High-grade dysplasia	
Male	49	High-grade dysplasia	
Female	29	High-grade dysplasia	
Male	49	High-grade dysplasia	
Female	70	High-grade dysplasia	
Male	80	High-grade dysplasia	
Male	65	High-grade dysplasia	
Male	56	High-grade dysplasia	
Female	76	High-grade dysplasia	
Male	56	High-grade dysplasia	
Male	74	High-grade dysplasia	
Male	61	High-grade dysplasia	
Male	35	Low-grade dysplasia	
Male	59	Low-grade dysplasia	
Female	53	High-grade dysplasia	
Female	68	Low-grade dysplasia	
Male	61	Low-grade dysplasia	
Female	43	Low-grade dysplasia	
Male	37	Low-grade dysplasia	
Male	55	High-grade dysplasia	
Male	82	High-grade dysplasia	
Female	62	Low-grade dysplasia	

Table II. Continued.

Sex	Age, years	Pathological grade (21)	pTNM stage (22)
Male	72	Low-grade dysplasia	
Male	62	Low-grade dysplasia	
Female	34	Low-grade dysplasia	
Female	61	Low-grade dysplasia	
Female	58	Low-grade dysplasia	
Female	53	Low-grade dysplasia	
Male	76	Low-grade dysplasia	
Male	29	Low-grade dysplasia	
Male	37	Low-grade dysplasia	
Male	38	Low-grade dysplasia	
Female	22	Low-grade dysplasia	
Male	69	Low-grade dysplasia	
Male	69	Low-grade dysplasia	
Male	65	Low-grade dysplasia	
Male	49	High-grade dysplasia	
Male	69	Low-grade dysplasia	
Male	63	Low-grade dysplasia	
Male	57	Low-grade dysplasia	
Female	56	Low-grade dysplasia	
Female	69	Low-grade dysplasia	
Female	72	Low-grade dysplasia	
Female	65	Low-grade dysplasia	
Male	75	Low-grade dysplasia	
Male	57	Low-grade dysplasia	
Female	35	Low-grade dysplasia	
Male	69	Adenocarcinoma	T2N0M0
Male	63	Adenocarcinoma	T4N1M1
Male	64	Adenocarcinoma	T3N1M0
Female	75	Adenocarcinoma	T1N0M0
Male	49	Adenocarcinoma	T3N1M0
Female	70	Adenocarcinoma	T3N1M0
Male	80	Adenocarcinoma	T1N0M0
Male	65	Adenocarcinoma	T2N0M0
Male	56	Adenocarcinoma	T3N0M0
Female	76	Adenocarcinoma	T2N1M0
Male	56	Adenocarcinoma	T1N0M0
Female	68	Adenocarcinoma	T2N1M0
Male	74	Adenocarcinoma	T4N1M0
Male	61	Adenocarcinoma	T4N1M0
Male	70	Adenocarcinoma	T3N1M0
Female	80	Adenocarcinoma	T1N0M0
Male	71	Adenocarcinoma	T2N0M0
Male	35	Adenocarcinoma	T3N0M0
Male	59	Adenocarcinoma	T3N0M0
Male	60	Adenocarcinoma	T3N0M0
Female	68	Adenocarcinoma	T2N0M0
Male	61	Adenocarcinoma	T3N1M0
Male	51	Adenocarcinoma	T3N1M0
Male	55	Adenocarcinoma	T4N2M1
Male	66	Adenocarcinoma	T3N1M0
Male	72	Adenocarcinoma	T4N2M0

Table II. Continued.

Sex	Age, years	Pathological grade (21)	pTNM stage (22)
Female	34	Adenocarcinoma	T4N2M0
Female	63	Adenocarcinoma	T1N0M0
Male	63	Adenocarcinoma	T1N0M0
Male	69	Adenocarcinoma	T4N1M0
Female	69	Adenocarcinoma	T2N0M0
Male	61	Adenocarcinoma	T4N1M0
Male	45	Adenocarcinoma	T2N0M0

pTNM, pathological Tumor-Node-Metastasis.

on accumulative hypergeometric distribution, and the q-value was calculated using the Benjamini-Hochberg program to consider multiple tests (27). When hierarchical clustering was performed on enriched terms, the κ score was used as a measure of similarity, and sub-trees with similarity >0.3 were then treated as clusters.

Prediction of the lncRNA-miRNA interactions and miRNA-mRNA interactions. The interactions between differentially expressed miRNAs (DEMs) and DEGs were predicted using miRWalk v3.0 (http://mirwalk.umm.uni-heidelberg.de/), which integrated the predictions of miRDB (28) and TargetScan (29). A score ≥0.95 was considered as the critical criterion for miRWalk predictive analysis. Considering the inhibition effect of miRNA on mRNA expression, the interactions of miRNA and mRNA with the same expression tendency were deleted. The interactions between miRNA and lncRNA were predicted using DIANA-LncBase v2.0 (30), and the score ≥0.4 was considered to be the cut-off criterion for predictive analysis in the LncBase prediction module. After the predicted targets were intersected with DEGs in the GSE109454 dataset and DEMs in the GSE41655 dataset, the mRNAs, miRNAs and lncRNAs were selected to construct the lncRNA-mRNA-miRNA regulatory network. The cytoscape software (v3.40; https://cytoscape.org/) was used to visualize the network.

Validation in the cancer genome atlas (TCGA) datasets. TCGA is a public platform for researchers to download and assess free datasets (https://cancergenome.nih.gov/) (31). RNA-sequencing (RNA-seq) data and clinical data from 31 cancer types are included in TCGA. In the present study, to improve the reliability of the analysis, the expression of hub genes was validated in TCGA datasets using an easy-use online tool GEPIA (v1.0; http://gepia.cancer-pku.cn/). The tumor type was limited to colon adenocarcinoma (COAD). Finally, RNA-seq data and clinical data from 275 primary tumor samples of COAD and 349 corresponding normal samples in TCGA datasets were used. Box and whisker plots were used to display the expression of the genes involved in the ceRNA network. $|\log_2 FC|>1$ and P-value <0.05 were set as the cut-off criteria to determine the differential expression between primary tumor samples and normal samples.

Results

Identification of DEGs. DEGs in the GSE109454 dataset and DEMs in the GSE41655 dataset were screened with a threshold of P<0.05 and $|\log_2 FC|>1$. As shown in Fig. 1, a total of 2,599 downregulated genes (1,430 mRNAs and 1,169 lncRNAs) and 2,832 upregulated genes (1,753 mRNAs and 1,079 lncRNAs) were screened in the GSE109454 dataset. In the GSE41655 dataset, a total of 116 DEMs, including 71 downregulated miRNAs and 45 upregulated miRNAs, between colorectal adenoma samples and normal colorectal samples, were screened (Fig. 2A and C), and a total of 109 DEMs, including 55 downregulated miRNAs and 54 upregulated miRNAs, between colorectal adenocarcinoma samples and normal colorectal samples, were screened (Fig. 2B and D). After the DEMs in the two groups were intersected, 78 DEMs were selected for further analysis.

Construction of co-expression network. Using the R package for WGCNA, 39 modules were generated (Fig. 3A) and all of the uncorrelated genes were assigned to the grey module. The number of the genes in every module is shown in Table III. The trait of the samples in the present study was divided into tumor and non-tumor. Out of 39 modules, the blue module was most positively associated with CRC (r=0.98; $P=1\times10^{-8}$; Fig. 3B). A total of 2,556,690 co-expression relationships in the blue module were therefore further filtered, and 169 genes (86 mRNAs and 83 lncRNAs) and 245 relationships were selected to construct the CNC network (Fig. 4).

GO and KEGG pathway enrichment analysis of genes in the CNC network. GO and KEGG enrichment analysis was performed for the genes in the blue module and the result is shown in Table IV. 'p53 signaling pathway' and 'cell cycle' were the most significantly enriched pathways in the KEGG pathway enrichment analysis. 'Cell division', 'chromosome segregation' and 'mitotic nuclear division' were the top 3 enriched BP terms.

Construction and validation of ceRNA network. To identify how lncRNAs regulate mRNAs through miRNAs in the CNC network, online prediction tools were used to determine miRNA-lncRNA and miRNA-mRNA interactions. The predicted miRNAs were then intersected with DEMs identified in the GSE41655 dataset. Subsequently a ceRNA network was constructed, which was composed of 24 mRNAs, 16 miRNAs and 10 lncRNAs (Fig. 5). To improve the reliability of the analysis, the expression of genes involved in the ceRNA network was validated in TCGA datasets. The results showed that 3 lncRNAs (MIR22HG, CHL1-AS2 and RP11-61I13.3) were downregulated in TCGA datasets (Fig. 6). As shown in Fig. 7, ANK2, BTK, FGL2, GBP2, PCSK5 and PDK4 were downregulated, and ARMC10, CENPF, CENPF, DKC1, GINS4, PAICS, PARPBP, RAD54B, RAE1 and TOMM34 were upregulated in TCGA datasets. Subsequently, these validated mRNAs and lncRNAs were selected to construct a ceRNA network. Finally, a validated ceRNA network composed of 2 lncRNAs (MIR22HG and RP11-61I13.3), 5 miRNAs (hsa-miR-765, hsa-miR-198, hsa-miR-125a-3p, hsa-miR-149-3p and hsa-miR-650) and 5 mRNAs (ANK2,

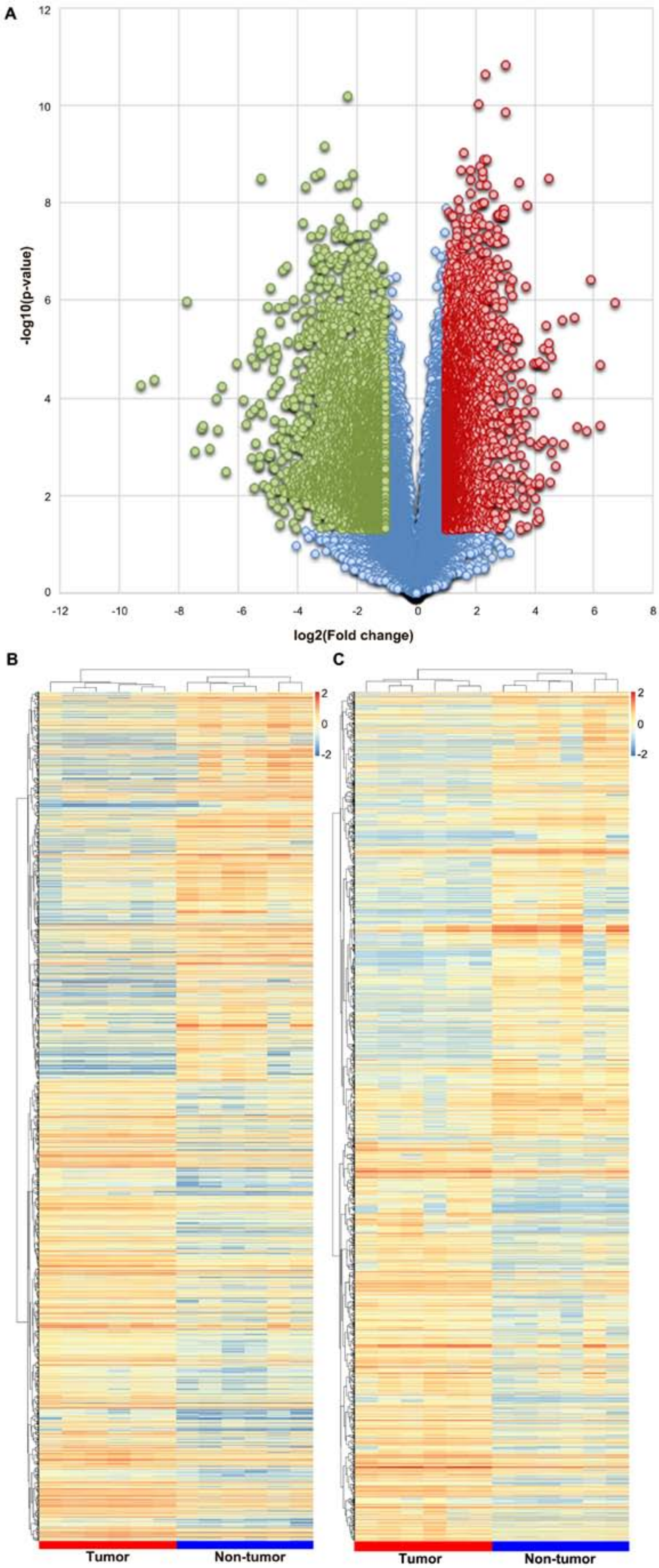

Figure 1. Identification of differentially expressed genes between tumor and normal samples in the GSE109454 dataset. (A) The volcano plot. The green dots indicate the significantly downregulated genes, the red dots indicate the significantly upregulated genes, while the blue dots indicate the genes with no significant difference. The horizontal axis represents the $\log_2$ (fold change) and the vertical axis represents the $-\log_{10}$ (P-value). (B and C) Heatmaps of the (B) differentially expressed mRNAs and (C) differentially expressed lncRNAs. Each row presents a dysregulated RNA transcript and each column represents a sample. Orange represents upregulation and blue represents downregulation.

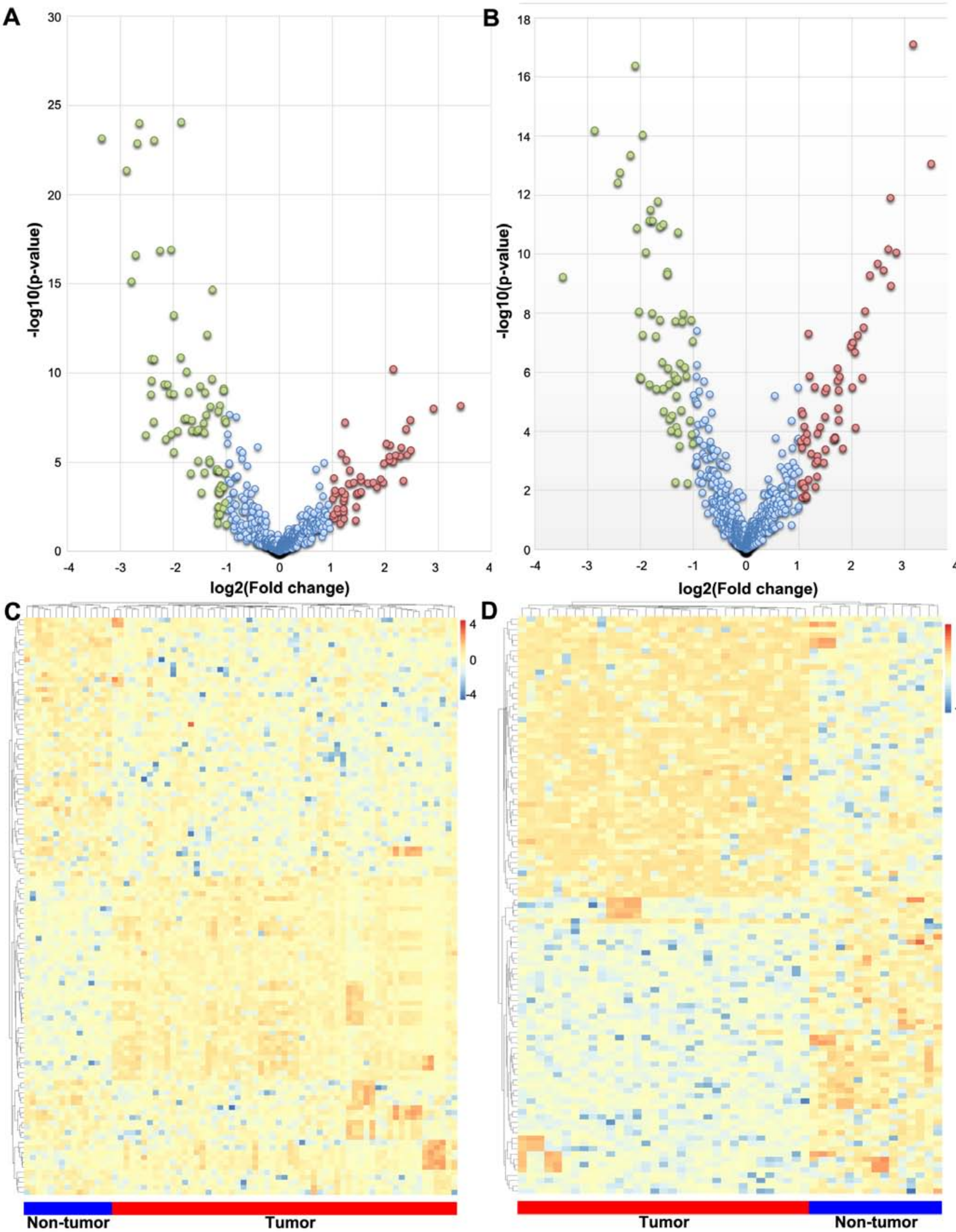

Figure 2. Identification of DEMs between tumor and normal samples in the GSE41655 dataset. (A) The volcano plots of the DEMs between colorectal adenoma samples and normal colorectal samples. (B) The volcano plots of the DEMs between colorectal adenocarcinoma samples and normal colorectal samples. The green dots indicate the significantly downregulated genes, the red dots indicate the significantly upregulated genes and the blue dots indicate the genes with no significant difference. The horizontal axis represents the $\log_2$ (fold change) and the vertical axis represents the $-\log_{10}$ (P-value). (C) Heatmaps of DEMs between colorectal adenoma samples and normal colorectal samples. (D) Heatmaps of DEMs between colorectal adenocarcinoma samples and normal colorectal samples. Each row presents a dysregulated RNA transcript and each column represents a sample. Orange represents upregulation and blue represents downregulation. DEMs, differentially expressed mRNAs.

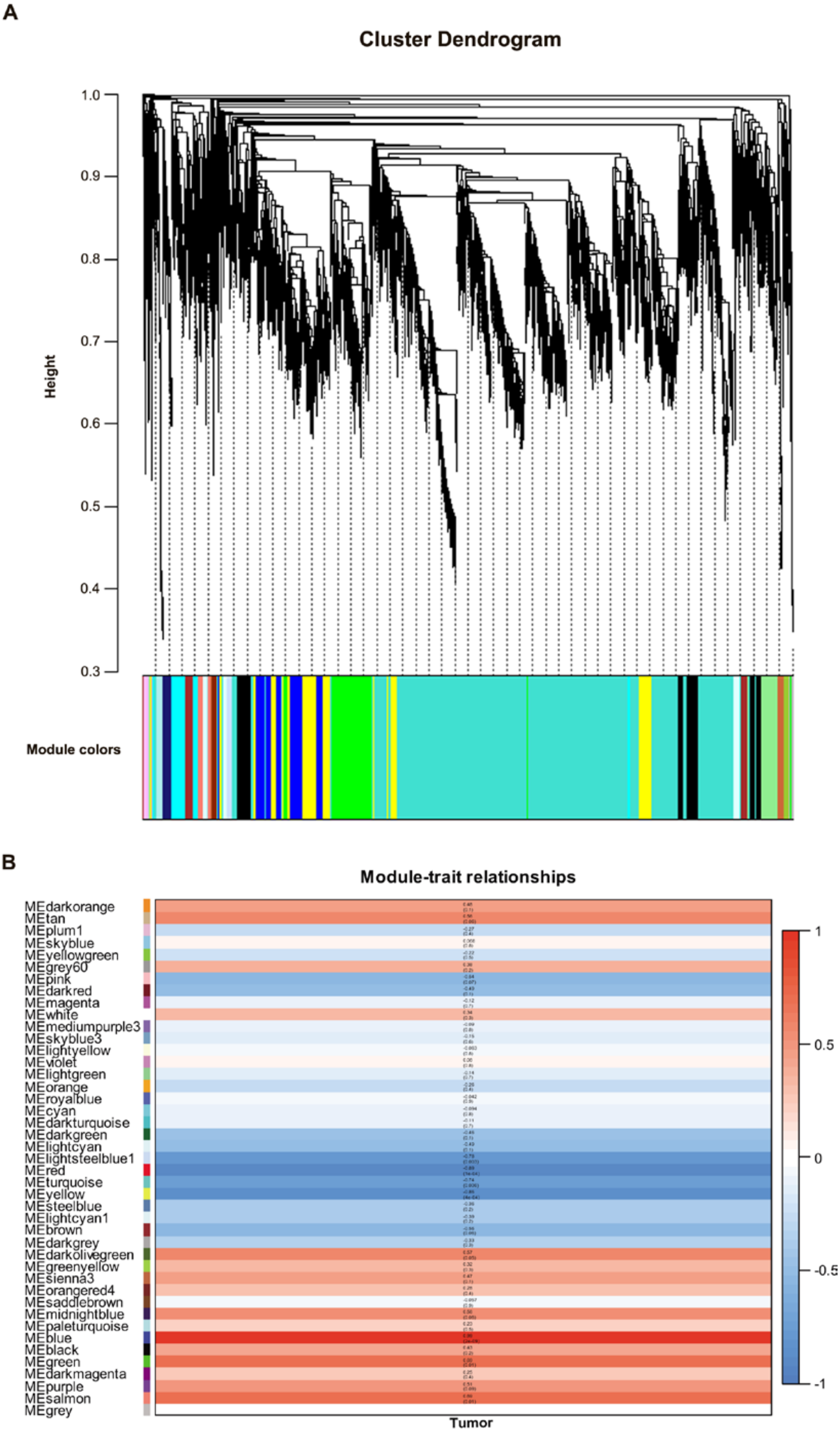

Figure 3. (A) Hierarchical clustering dendrogram of all the probe sets used in the weighted gene co-expression network analysis based on the 1-TOM dissimilarity measure. Each line of the dendrogram represents an individual probe. The multi-color bar represents the 39 modules identified using the dynamic tree cut algorithm. (B) Correlations between module eigengenes and tumor status. The numbers within the heatmap indicate correlations and P-values for the module-trait associations, respectively. Red wells indicate positive correlation, blue wells indicate negative correlation and white wells indicate no correlation.

Table III. Number of genes in the 39 modules.

Module colors	Frequency
Black	426
Blue	2,249
Brown	1,417
Cyan	274
Dark green	133
Dark grey	103
Dark magenta	56
Dark olive green	59
Dark orange	84
Dark red	145
Dark turquoise	103
Green	507
Green yellow	293
Grey	191
Grey 60	223
Light cyan	246
Light green	177
Light yellow	169
Magenta	349
Midnight blue	264
Orange	97
Pale turquoise	60
Pink	391
Plum 1	39
Purple	313
Red	439
Royal blue	154
Saddle brown	63
Salmon	282
Sienna 3	52
Sky blue	72
Sky blue 3	47
Steel blue	63
Tan	291
Turquoise	3,809
Violet	59
White	80
Yellow	909
Yellow green	48

BTK, GBP2, PCSK5 and PDK4) was obtained (Fig. 8). PCSK5 was at the center of the validated ceRNA network, and interacted with 4 miRNAs.

Discussion

In developing countries, the incidence of CRC is rapidly increasing. CRC is the fifth most common malignant tumor after lung, liver, esophagus and breast cancer in China (32). Statistically, there were 376,300 new cases and 191,000 CRC-related deaths in China in 2015 (32). Therefore, CRC has become a public health issue and it is essential to understand the etiological factors and mechanisms of CRC progression to improve prevention and survival rates.

In the present study, 3,183 dysregulated mRNAs, 78 dysregulated miRNAs and 2,248 dysregulated lncRNAs were screened in two GEO datasets. WGCNA analysis was conducted to identify the co-expression between coding genes and non-coding genes. Combined with the dysregulated genes, 169 genes were selected to construct the CNC network. GO and KEGG pathway enrichment analysis was performed to identify the biological function of the genes in the CNC network. The genes in the CNC network were significantly enriched in 'cell cycle' and 'p53 signaling pathway'. DEMs identified from the GS41655 dataset were predicted to be involved in miRNA-lncRNA and miRNA-mRNA interactions were used to construct the ceRNA network. The expression of 3 lncRNAs (MIR22HG, CHL1-AS2 and RP11-61I13.3) and 15 mRNAs (ANK2, ARMC10, BTK, CENPF, DKC1, FGL2, GBP2, GINS4, PAICS, PARPBP, PCSK5, PDK4, RAD54B, RAE1 and TOMM34) in the ceRNA network were validated and found to be dysregulated in TCGA datasets.

PCSK5 is at the center of the validated ceRNA network, and interacts with 4 miRNAs. PCSK5 belongs to the subtilisin-like proprotein convertase family. The members of this family are proprotein convertases that process a potential precursor protein into its biologically active product (33). Reports of the involvement of PCSK5 in cancer are rare. In triple-negative breast cancer (TNBC), a lack of PCSK5 could lead to the bioactivity of growth differentiation factor (GDF11) as a tumor-suppressor. PCSK5 reconstitution mobilizes the latent TNBC reservoir of GDF11 *in vitro* and inhibits TNBC metastasis to the lung of syngeneic hosts (34). However, the function of PCSK5 in CRC is largely unknown. The present study indicated that PCSK5 could be regulated by MIR22HG through miR-198 and miR-149-3p, and regulated by RP11-61I13.3 through miR-765 and miR-125-3p.

MIR22HG has been identified as a tumor suppressor in several studies. In hepatocellular carcinoma (HCC), MIR22HG expression is significantly downregulated, and can regulate the expression of high mobility group box 1 and its downstream pathways by deriving miR-22-3p to suppress HCC cell proliferation, invasion and metastasis (35). In endometrial cancer (EC), MIR22HG can act as sponge of miR-141-3p, which could inhibit EC cells proliferation and induce EC cells apoptosis by targeting death-associated protein kinase 1 (36). In addition, the present study suggested that MIR22HG could regulate the expression of BTK and PDK4 through miR-650. Therefore, MIR22HG might be an important regulator in CRC. BTK is a new regulator of p53 and BTK induction, which leads to p53 phosphorylation. Inhibiting BTK can reduce p53-dependent senescence and apoptosis (37). In CRC, miR-650 is a direct regulator of N-myc downstream-regulated gene 2, which is a potential tumor suppressor gene (38). Therefore, MIR22HG might be an important regulator in CRC.

RP11-61I13.3 is a novel lncRNA and little is known about its function. RP11-61I13.3 can also regulate GBP2 and ANK2

Table IV. Gene function enrichment analysis of the genes included in the coding non-coding network.

Category	Term	Count	P-value	Genes
KEGG	Cell cycle	5	1.66×10^{-3}	CDK1, MAD2L1, CCNB2, E2F5, ATR
	p53 signaling pathway	3	3.18×10^{-2}	CDK1, CCNB2, ATR
BP	Cell division	13	7.04×10^{-8}	CDK1, MAD2L1, CCNB2, KIF11, OIP5, NEK2, KIF20B, NUF2, SDCCAG3, CENPF, AURKA, UBE2C
	Mitotic nuclear division	10	2.17×10^{-6}	CDK1, KIF11, CCNB2, OIP5, NEK2, KIF20B, NUF2, CENPF, CENPW, AURKA
	Chromosome segregation	6	1.58×10^{-5}	KIF11, OIP5, NEK2, NUF2, CENPF, CENPW
	DNA strand elongation involved in DNA replication	4	4.21×10^{-5}	GINS1, RFC3, GINS4, POLD2
	Positive regulation of telomere maintenance via telomerase	4	4.34×10^{-4}	DKC1, NEK2, ATR, CCT6A
	DNA replication	6	7.71×10^{-4}	CDK1, RFC3, GINS4, POLD2, ATR, DSCC1
	DNA repair	7	7.81×10^{-4}	CDK1, RAD51AP1, FANCD2, PARPBP, RUVBL2, ACTL6A, ATR
	CENP-A containing nucleosome assembly	4	1.04×10^{-3}	CENPL, OIP5, CENPQ, CENPW
	DNA duplex unwinding	4	1.11×10^{-3}	GINS1, GINS4, RUVBL2, RAD54B
	Sister chromatid cohesion	5	1.35×10^{-3}	CENPL, MAD2L1, CENPQ, NUF2, CENPF
	Positive regulation of telomerase RNA localization to Cajal body	3	2.15×10^{-3}	DKC1, RUVBL2, CCT6A
	Spindle organization	3	2.45×10^{-3}	KIF11, AURKA, AUNIP
	Cell cycle	6	3.38×10^{-3}	DTYMK, SDCCAG3, AURKA, ATR, CDKN3, SUV39H2
	G2/M transition of mitotic cell cycle	5	3.79×10^{-3}	CDK1, PLK4, CCNB2, NEK2, AURKA
	Regulation of mitotic nuclear division	3	5.50×10^{-3}	MKI67, NEK2, KIF20B
	Anaphase-promoting complex-dependent catabolic process	4	5.92×10^{-3}	CDK1, MAD2L1, AURKA, UBE2C
	Cytoplasmic translation	3	5.96×10^{-3}	RPL6, RPL8, FTSJ1
	Reciprocal meiotic recombination	3	8.52×10^{-3}	CCNB1IP1, RAD54B, TRIP13
	Histone H4 acetylation	3	9.08×10^{-3}	NCOA1, RUVBL2, ACTL6A
	Positive regulation of type I hypersensitivity	2	01.3871	FCER1A, BTK
	Mitotic nuclear envelope disassembly	3	1.78×10^{-2}	CDK1, CCNB2, RAE1
	Regulation of mitotic centrosome separation	2	1.85×10^{-2}	KIF11, NEK2
	Interstrand cross-link repair	3	2.18×10^{-2}	RAD51AP1, FANCD2, ATR
	Mitotic centrosome separation	2	2.30×10^{-2}	KIF11, AURKA
	Regulation of growth	3	2.52×10^{-2}	ARMC10, RUVBL2, ACTL6A
	Cell proliferation	6	2.77×10^{-2}	CDK1, MKI67, DKC1, DLGAP5, DTYMK, CENPF
	Positive regulation of DNA-directed DNA polymerase activity	2	3.21×10^{-2}	RFC3, DSCC1
	Box C/D snoRNP assembly	2	3.21×10^{-2}	NUFIP1, RUVBL2
	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process	4	3.44×10^{-2}	CDK1, MAD2L1, AURKA, UBE2C
	Positive regulation of establishment of protein localization to telomere	2	4.10×10^{-2}	DKC1, CCT6A
	Negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	3	4.31×10^{-2}	CDK1, MAD2L1, UBE2C
	Positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	3	4.87×10^{-2}	CDK1, MAD2L1, UBE2C

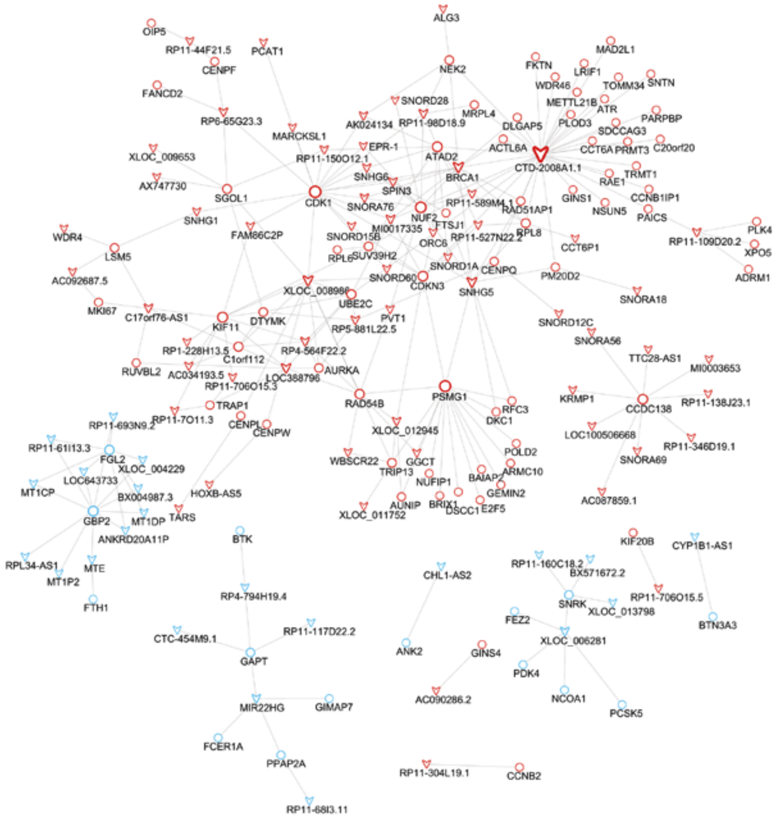

Figure 4. Coding non-coding network. Red represents upregulation and blue represents downregulation. The V-shape represents lncRNA and circles represent mRNA.

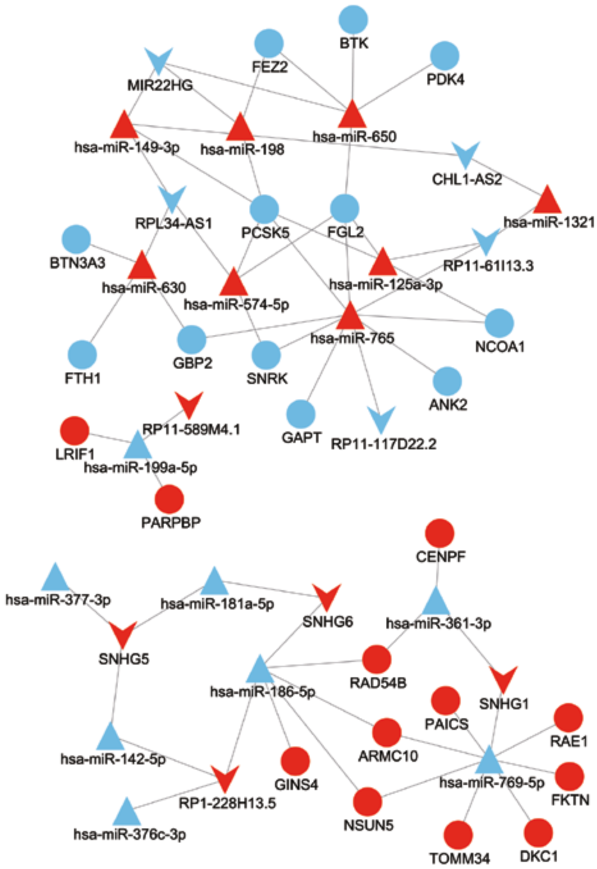

Figure 5. Competing endogenous RNA network. Red represents upregulation and blue represents downregulation. The V-shape represents lncRNA, circles represent mRNA and triangles represent microRNA.

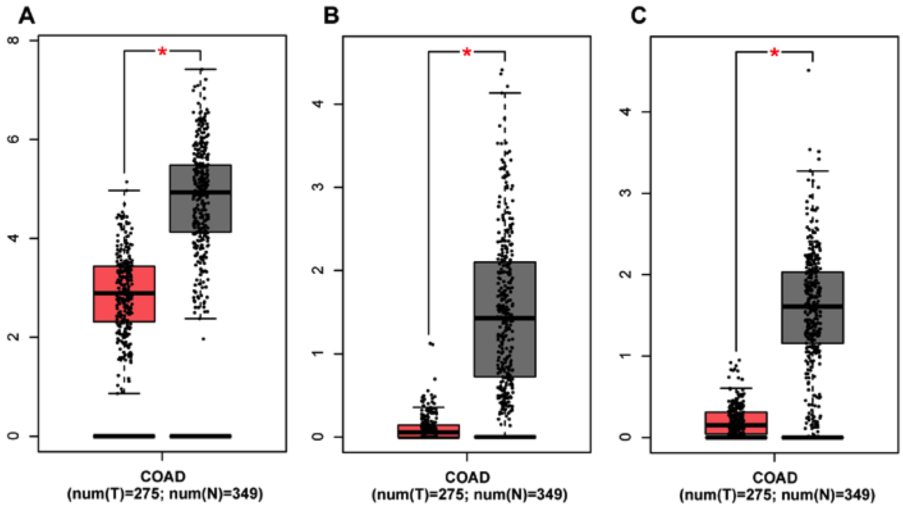

Figure 6. Gene expression of the selected long non-coding RNAs validated in The Cancer Genome Atlas dataset. The dysregulation of (A) MIR22HG, (B) CHL1-AS2 and (C) RP11-61I13.3. The red box corresponds to tumor samples and the grey box corresponds to non-tumor samples. The vertical axis represents the relative expression of genes. *P<0.05. COAD, colon adenocarcinoma; T, tumor; N, normal.

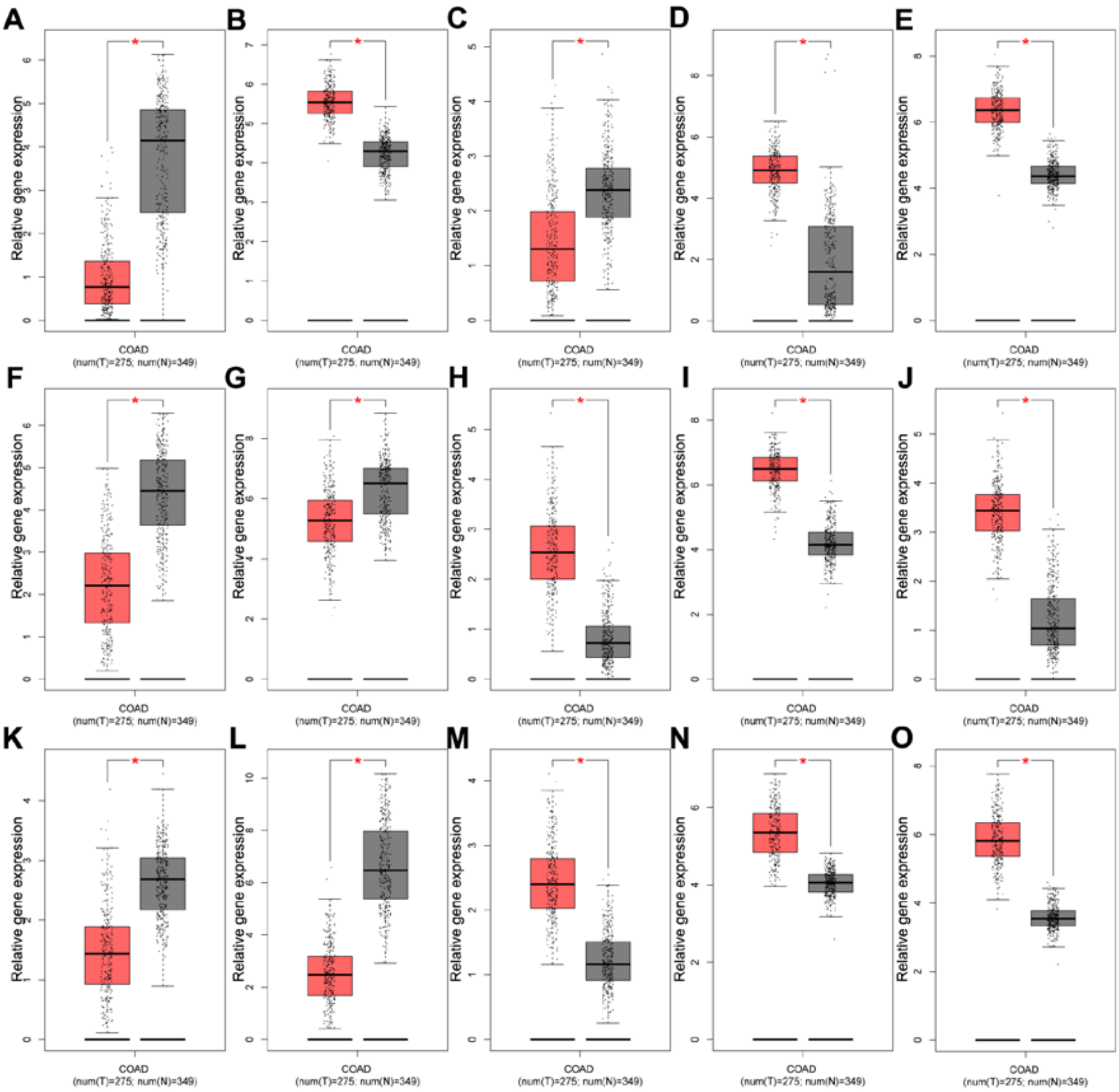

Figure 7. Gene expression of the selected mRNAs validated in The Cancer Genome Atlas dataset. The dysregulation of (A) ANK2, (B) ARMC10, (C) BTK, (D) CENPF, (E) DKC1, (F) FGL2, (G) GBP2, (H) GINS4, (I) PAICS, (J) PARPBP, (K) PCSK5, (L) PDK4, (M) RAD54B, (N) RAE1 and (O) TOMM34. The red box corresponds to tumor samples and the grey box corresponds to non-tumor samples. The vertical axis represents the relative expression of genes. *P<0.05. COAD, colon adenocarcinoma; T, tumor; N, normal.

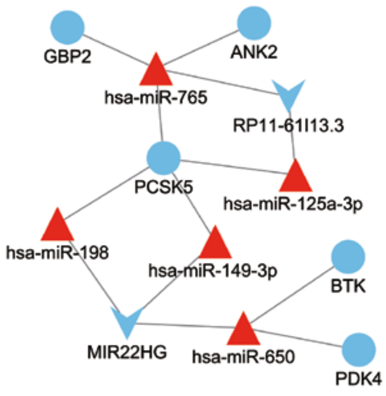

Figure 8. Competing endogenous RNA network constructed from validated genes. Red represents upregulation and blue represents downregulation. V-shape represents lncRNA, circle represents mRNA, and triangle represents miRNA.

through miR-765, and miR-765 has been demonstrated to be an onco-miRNA in HCC (39) and oral squamous cancer (40). Meanwhile, GBP2 has been associated with a better prognosis in several cancer types, including breast cancer and ovarian cancer (41,42).

In conclusion, two lncRNAs (MIR22HG and RP11-61I13.3) have been identified that can regulate several mRNAs through sponging miRNAs in CRC, and the potential functions of these genes have been revealed. The ceRNA network involved in PCSK5 has been shown to play an important role in CRC. However, the diagnostic and prognostic value of these genes still requires further validation.

Acknowledgements

Not applicable.

Funding

The current study was supported by the Program for Improvement Scientific Research Ability of Young and Middle-Aged Teachers of Higher Education of Guangxi (grant no. 2017KY0093) and the 2018 Innovation Project of Guangxi Graduate Education (grant no. YCBZ2018036).

Availability of data and materials

The datasets generated and/or analyzed during the present study are available in the NCBI GEO (https://www.ncbi.nlm.nih.gov/geo) (GSE109454 and GSE41655) and GEPIA (https://gepia.cancer-pku.cn/) depositories.

Authors' contributions

CL designed the study, analyzed the data and wrote the manuscript. XH analyzed the data and the literature, wrote the manuscript and contributed to the finalization of the manuscript. YG analyzed the data. QL designed the study, analyzed the data and wrote the manuscript. All authors reviewed and approved the manuscript, and agree to be accountable for all aspects of the research.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD and Jemal A: Cancer statistics, 2018. CA Cancer J Clin 68: 7-30, 2018.
2. Sato H, Kotake K, Sugihara K, Takahashi H, Maeda K and Uyama I; Study Group for Peritoneal Metastasis from Colorectal Cancer By the Japanese Society for Cancer of the Colon and Rectum: Clinicopathological factors associated with recurrence and prognosis after R0 resection for stage IV colorectal cancer with peritoneal metastasis. Dig Surg 33: 382-391, 2016.
3. Corte H, Manceau G, Blons H and Laurent-Puig P: MicroRNA and colorectal cancer. Dig Liver Dis 44: 195-200, 2012.
4. Gonzalez-Pons M and Cruz-Correa M: Colorectal cancer biomarkers: Where are we now? Biomed Res Int 2015: 149014, 2015.
5. Xie X, Zheng X, Han Z, Chen Y, Zheng Z, Zheng B, He X, Wang Y, Kaplan DL, Li Y, *et al* A biodegradable stent with surface functionalization of combined-therapy drugs for colorectal cancer. Adv Healthc Mater 7: e1801213, 2018.
6. Valastyan S: Roles of microRNAs and other non-coding RNAs in breast cancer metastasis. J Mammary Gland Biol Neoplasia 17: 23-32, 2012.
7. Zhang N, Wang X, Huo Q, Sun M, Cai C, Liu Z, Hu G and Yang Q: MicroRNA-30a suppresses breast tumor growth and metastasis by targeting metadherin. Oncogene 33: 3119-3128, 2014.
8. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y and Goodall GJ: The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 10: 593-601, 2008.
9. Seitz H: Redefining microRNA targets. Curr Biol 19: 870-873, 2009.
10. Xia T, Chen S, Jiang Z, Shao Y, Jiang X, Li P, Xiao B and Guo J: Long noncoding RNA FER1L4 suppresses cancer cell growth by acting as a competing endogenous RNA and regulating PTEN expression. Sci Rep 5: 13445, 2015.
11. Yue B, Sun B, Liu C, Zhao S, Zhang D, Yu F and Yan D: Long non-coding RNA Fer-1-like protein 4 suppresses oncogenesis and exhibits prognostic value by associating with miR-106a-5p in colon cancer. Cancer Sci 106: 1323-1332, 2015.
12. Salmena L, Poliseno L, Tay Y, Kats L and Pandolfi PP: A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? Cell 146: 353-358, 2011.
13. Karreth FA, Reschke M, Ruocco A, Ng C, Chapuy B, Léopold V, Sjoberg M, Keane TM, Verma A, Ala U, *et al*: The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell 161: 319-332, 2015.
14. Poliseno L, Salmena L, Jiangwen Z, Carver B, Haveman WJ and Pandolfi PP: A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465: 1033-1038, 2010.
15. Qu J, Li M, Zhong W and Hu C: Competing endogenous RNA in cancer: A new pattern of gene expression regulation. Int J Clin Exp Med 8: 17110-17116, 2015.
16. Lo PK, Zhang Y, Wolfson B, Gernapudi R, Yao Y, Duru N and Zhou Q: Dysregulation of the BRCA1/long non-coding RNA NEAT1 signaling axis contributes to breast tumorigenesis. Oncotarget **7**: 65067-65089, 2016.
17. Liang WC, Fu WM, Wong CW, Wang Y, Wang WM, Hu GX, Zhang L, Xiao LJ, Wan DC, Zhang JF and Waye MM: The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. Oncotarget 6: 22513-22525, 2015.

18. Li LJ, Zhao W, Tao SS, Leng RX, Fan YG, Pan HF and Ye DQ: Competitive endogenous RNA network: Potential implication for systemic lupus erythematosus. Expert Opin Ther Targets 21: 639-648, 2017.
19. Xu J, Zhang R and Zhao J: The novel long noncoding RNA TUSC7 inhibits proliferation by sponging MiR-211 in colorectal cancer. Cell Physiol Biochem 41: 635-644, 2017.
20. He F, Song Z, Chen H, Chen Z, Yang P, Li W, Yang Z, Zhang T, Wang F, Wei J, *et al*: Long noncoding RNA PVT1-214 promotes proliferation and invasion of colorectal cancer by stabilizing Lin28 and interacting with miR-128. Oncogene 38: 164-179, 2019.
21. Bosma FT: WHO classification of tumours of the digestive system, 2010.
22. Rice TW: 7th Edition AJCC/UICC Staging, Esophagus and Esophagogastric Junction, 2017.
23. Diboun I, Wernisch L, Orengo CA and Koltzenburg M: Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics 7: 252, 2006.
24. Zhang B and Horvath S: A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4: Article17, 2005.
25. Gao B, Shao Q, Choudhry H, Marcus V, Dong K, Ragoussis J and Gao ZH: Weighted gene co-expression network analysis of colorectal cancer liver metastasis genome sequencing data and screening of anti-metastasis drugs. Int J Oncol 49: 1108-1118, 2016.
26. Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559, 2008.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29, 2000.
28. Wong N and Wang X: miRDB: An online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 43 (Database Issue): D146-D152, 2015.
29. Lewis BP, Burge CB and Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15-20, 2005.
30. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, Zagganas K, Tsanakas P, Floros E, Dalamagas T and Hatzigeorgiou AG: DIANA-LncBase v2: Indexing microRNA targets on non-coding transcripts. Nucleic Acids Res 44: D231-D238, 2016.
31. Tomczak K, Czerwińska P and Wiznerowicz M: The cancer genome atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol (Pozn) 19: A68-A77, 2015.
32. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ and He J: Cancer statistics in China, 2015. CA Cancer J Clin 66: 115-132, 2016.
33. Cao H, Mok A, Miskie B and Hegele RA: Single-nucleotide polymorphisms of the proprotein convertase subtilisin/kexin type 5 (PCSK5) gene. J Hum Genet 46: 730-732, 2001.
34. Bajikar SS, Wang CC, Borten MA, Pereira EJ, Atkins KA and Janes KA: Tumor-auppressor inactivation of GDF11 occurs by precursor sequestration in triple-negative breast cancer. Dev Cell 43: 418-435.e13, 2017.
35. Zhang DY, Zou XJ, Cao CH, Zhang T, Lei L, Qi XL, Liu L and Wu DH: Identification and functional characterization of long non-coding RNA MIR22HG as a tumor suppressor for hepatocellular carcinoma. Theranostics 8: 3751-3765, 2018.
36. Cui Z, An X, Li J, Liu Q and Liu W: LncRNA MIR22HG negatively regulates miR-141-3p to enhance DAPK1 expression and inhibits endometrial carcinoma cells proliferation. Biomed Pharmacother 104: 223-228, 2018.
37. Althubit M, Rada M, Samuel J, Escorsa JM, Najeeb H, Lee KG, Lam KP, Jones GD, Barlev NA and Macip S: BTK modulates p53 activity to enhance apoptotic and senescent responses. Cancer Res 76: 5405-5414, 2016.
38. Feng L, Xie Y, Zhang H and Wu Y: Down-regulation of NDRG2 gene expression in human colorectal cancer involves promoter methylation and microRNA-650. Biochem Biophys Res Commun 406: 534-538, 2011.
39. Xie BH, He X, Hua RX, Zhang B, Tan GS, Xiong SQ, Liu LS, Chen W, Yang JY, Wang XN and Li HP: Mir-765 promotes cell proliferation by downregulating INPP4B expression in human hepatocellular carcinoma. Cancer Biomark 16: 405-413, 2016.
40. Zheng Z, Luan X, Zha J, Li Z, Wu L, Yan Y, Wang H, Hou D, Huang L, Huang F, *et al*: TNF-α inhibits the migration of oral squamous cancer cells mediated by miR-765-EMP3-p66Shc axis. Cell Signal 34: 102-109, 2017.
41. Godoy P, Cadenas C, Hellwig B, Marchan R, Stewart J, Reif R, Lohr M, Gehrmann M, Rahnenführer J, Schmidt M and Hengstler JG: Interferon-inducible guanylate binding protein (GBP2) is associated with better prognosis in breast cancer and indicates an efficient T cell response. Breast Cancer 21: 491-499, 2014.
42. Wang X, Wang SS, Zhou L, Yu L and Zhang LM: A network-pathway based module identification for predicting the prognosis of ovarian cancer patients. J Ovarian Res 9: 73, 2016.