Construction of CeRNA Regulatory Network based on WGCNA Reveals Diagnosis Biomarkers for Colorectal Cancer

Jun Xiang  
The Second Affiliated Hospital of Harbin Medical University

HaoYu Jing  
The Second Affiliated Hospital of Harbin Medical University

Yun Xiao Liu  
The Second Affiliated Hospital of Harbin Medical University

HuFei Wang  
The Second Affiliated Hospital of Harbin Medical University

ZeWen Chang  
The Second Affiliated Hospital of Harbin Medical University

ShiHao Liu  
The Second Affiliated Hospital of Harbin Medical University

Lei Yu  
The Second Affiliated Hospital of Harbin Medical University

GuiYu Wang (✉ guiwyang@163.com)  
The Second Affiliated Hospital of Harbin Medical University

Research Article

Keywords: ceRNA, colorectal cancer, WGCNA, network, targets, TCGA

Posted Date: April 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1468089/v1

License: ☇ ① This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: Colorectal cancer is the third primary cause of death among cancers in the world. Although improvements in various treatments have greatly improved the survival time of colorectal cancer patients, since colorectal cancer is often at an advanced stage when diagnosed, the prognosis of patients still very poor. Since the ceRNA regulatory network was proposed in 2011, it has greatly promoted the study of the molecular mechanism of colorectal cancer occurrence and development.

Objective: Exploring the new molecular mechanism of the occurrence and development of colorectal cancer, and providing new targets for the diagnosis and treatment of colorectal cancer.

Method: We analyzed the colorectal cancer RNA-seq data in TCGA, such as, difference analysis, weighted gene co-expression network analysis (WGCNA) and construction of ceRNA regulatory network, et.al.

Results: Eventually, 19 hub genes with significant prognostic significance were identified in the ceRNA regulatory network, including 8 lncRNAs, 2 mRNAs and 9 miRNAs. These hub genes constitute the IncRNA-miRNA, miRNA-mRNA and IncRNA-miRNA-mRNA regulatory axis. Some of these ceRNA regulatory axes have been confirmed to exist in colorectal cancer or other cancers.

Conclusion: In this article, some new ceRNA regulatory axes have been discovered, which provides new molecular mechanisms for the occurrence and development of colorectal cancer, thereby providing new targets for clinical diagnosis and treatment.

Introduction

Colorectal cancer, one of the malignant tumors around the world, is the third dominating cause of cancer deaths worldwide and causes approximately 700,000 people demises per year(1, 2). According to a recent report by the International Agency for Research on Cancer, colorectal cancer accounts for 6.3% of all cancer deaths in China(3). Owing to the lack of specific and sensitive biomarkers, colorectal cancer patients are usually diagnosed as advanced cancer, and the five years overall survival rate is 40%-60% (4, 5). Although the progress of interventional therapy has greatly improved the overall survival (OS) of CRC in recent years, the prognosis of patients is still very poor due to the high recurrence and metastasis rate of advanced CRC(6). As a multistage disease, Colorectal cancer is the accumulation of multiple genetic or epigenetic changes and their complex mutual effect (7). In order to improve the early diagnosis, treatment and prognosis of colorectal cancer, it is urgent to study the molecular mechanism of colorectal cancer from initiation to metastasis. In recent years, there have been endless reports about lncRNA regulating disparate biological behaviours of colorectal tumor cells via ceRNA regulation network. For instance, lncRNA CCAT1 functions as ceRNA to antagonize miR-401 in down-regulating ITPKB in human colon cancer cell HCT116(8). In addition, it has been reported that lncRNA CLM regulates the expression of ZEB1 as a ceRNA of miR-215 to promote live metastasis of colorectal cancer(9). Besides the newly discovered IncRNA TUSC7 inhibits the proliferation of colorectal cancer cells through the molecular sponge miR-211 (10).
Non-coding RNA (ncRNA) refers to transcripts without protein coding function, and its number accounts for more than 98% of the entire genome transcript (11). The definition of long non-coding RNA (lncRNA) is RNA transcripts that are more than 200 nucleotides in length and cannot encode proteins. Although most lncRNAs have poly-A tails, they cannot be translated into proteins. Compared with protein-encoding mRNA, lncRNA shows greater tissue specificity, so it may become a biomarker for many diseases (12). MicroRNAs (miRNAs), non-coding RNAs with a length of 19-22 bases, were first discovered in eukaryotes, which can regulate endogenous genes. MiRNAs degrade messenger RNAs (mRNAs) or inhibit protein translation to regulate them at the posttranscriptional gene expression level (13, 14). lncRNA uses all sorts of mechanisms to regulate the expression, degradation and modification of proteins. The most critical regulation is the competitive endogenous RNA (ceRNAs) theory proposed by Salmena et al. (15). The ceRNA hypothesis describes the intricate post-transcriptional communication network of all transcripts, including lncRNA RNA species, which can be used as natural miRNA sponges to inhibit miRNA function by sharing miRNA response elements (MRE) (16). The importance of lncRNA-miRNA-mRNA regulatory network in different diseases was confirmed by subsequent studies (17).

Although people's understanding of lncRNA has made great progression in the past decades, only a small part of the annotated lncRNA in colorectal cancer has been well identified in terms of biological function. In this study, we analyzed the RNA sequencing data from colon cancer (COAD) and rectal cancer (READ) in TCGA, on behalf of identifying the lncRNA, miRNA and mRNA that were significantly changed during the development of colorectal cancer. On the side, the ceRNA regulatory network of these lncRNAs, miRNAs and mRNAs was constructed to make clear the molecular mechanisms involved in the initiation and progression of colorectal cancer, and provide new thinking for clinical diagnosis and treatments.

**Materials And Methods**

**Data Download and Processing**

The data in this study are from online public databases, downloading CRC chip data from GEO database. The inclusion criteria are: 1. There are at least 10 tumor and normal tissue samples; 2. The data includes tumor and normal samples; 3. The sample data has not undergone any processing. Finally obtained GEO data including GSE156355, GSE110224, GSE110223, GSE41657, GSE113512, GSE50117, GSE103512. In addition, the Counts data of colorectal cancer and clinical related information data were downloaded from the TCGA database. A total of 698 RNA-seq data were downloaded. It mainly includes 17580 mRNAs, 7365 lncRNAs and 802 miRNAs data.

The data obtained from GEO uses the R package "limma" for correction and normalization, and log2 conversion is performed on the data. The colorectal cancer data obtained in TCGA uses the ENSEMBL database for gene annotation. In the meantime, the average value of the gene, having repetition name, is taken, and the genes that are less than 30% expressed in the sample are finally removed.

**Screening of Differentially Expressed LncRNA and mRNA**
The R package "DESeq2" to analyze the differentially expression of the TCGA counts data. For the p-values, we use false discovery rate (FDR) to correct for the statistical significance of multiple tests. The final result uses $|\log_2 FC|>1$ and FDR<0.05 as cutoff criteria to screen differentially expressed genes (DEGs) and differentially expressed IncRNAs (DELs).

**Weighted Gene Co-expression Network Analysis**

The differentially expressed genes screened in the TCGA colorectal cancer data were selected to establish a weighted gene co-expression network. After selecting the appropriate samples and genes, using the "WGCNA" R package to calculate the “Pearson correlations coefficient” between all gene pairs in the selected samples to construct an adjacency matrix. Then, use the power of $\beta = 6$ (scale-free $R^2 = 0.90$) as the soft threshold to ensure that the network is scale-free. In order to further identify the functional modules in the co-expression network with these 5201 genes, the adjacency matrix is used to calculate the Topological Overlap Measure (TOM), which represents the overlap in the shared neighbors.

Modular eigengenes (MEs) are considered to be representative of gene expression profiles in the modules. Selecting modules related to tumor function for subsequent analysis.

**GO Enrichment Analysis and KEGG Pathway Analysis**

Performing GO enrichment and KEGG pathway analysis on the characteristic genes of different modules by the R package "cluster profiler". P-value<0.05 is considered to be a significant enrichment analysis result. The R package is used to analyze the KEGG pathway, and the threshold p-value<0.05. Using GO and KEGG pathways to analyze and predict potential functions.

**CeRNA Network Construction and Topology Analysis**

After predicting miRNAs in the miRcode and ENCORI databases, we obtained miRNAs that interact with different lncRNAs, and overlapped these miRNAs with miRNAs in TCGA colorectal cancer to obtain the final lncRNA-miRNA relationship file. We used miRTarBase and miRWalk databases to predict the target genes of these miRNAs, and overlapped the genes predicted by the two databases with the eigengenes genes of the module, and obtained miRNA-mRNA relationship pairs.

We merged the obtained lncRNA-miRNA and miRNA-mRNA two relationships pairs, and apply them to the Cytoscape software to visualize the topological network to obtain the topological network diagram of ceRNA. Performing topology analysis on the network, hub nodes was selected with a degree greater than 5 to construct a sub-network and perform subsequent analysis.

**Statistical Analysis of Hub Nodes**

Using the analysis tool R x64 4.0.4 and the online analysis tool "Xiantao Academic" (https://www.xiantao.love/) to perform statistical analysis on the sub-network node genes. It mainly
includes differentially expression analysis, ROC analysis and survival analysis. Kaplan-Meier method and log-rank test were used in survival analysis. P-value<0.05 was considered statistically significant.

Results

The workflow is shown in Figure 1.

**Determination of Significantly Differentially Expressed mRNA and Significantly Differentially Expressed LncRNA**

The mRNA expression profiles of 51 normal samples and 646 tumor samples were compared. A total of 5201 differentially expressed mRNAs were obtained through statistical testing, including 2653 up-regulated and 2548 down-regulated (Figures 2A and C).

The lncRNA expression profiles of 51 normal samples and 646 tumor samples were compared. A total of 2856 differentially expressed lncRNAs were obtained through statistical testing, including 1981 up-regulated and 875 down-regulated (Figure 2B and D).

**Weighted Gene Co-expression Network Analysis Results**

The aforementioned significantly differentially expressed mRNA was used to construct WGCNA, and a total of 5201 significantly different mRNA expression profiles were selected for WGCNA analysis (Figure 3A, B, D and E). Firstly, selecting the appropriate soft threshold value according to $R^2=0.9$, and finally choosing the soft threshold value $\beta=6$ to establish the relationship matrix, then convert the relationship matrix into an adjacency matrix, and introduce power exponent weighting to construct a scale-free network, and finally in the adjacency matrix Based on the establishment of TOM matrix, calculate the degree of TOM difference between genes (distTOM), and then establish gene feature modules, obtaining a total of 24 modules. GO enrichment and KEGG pathway analysis were performed on the genes of different module eigengenes. According to GO function annotation and KEGG pathway enrichment analysis results, tumor-related modules are selected for subsequent analysis. Finally, the yellow module is selected for the further analysis. The yellow module has a total of 519 genes. The GO function annotation and KEGG enrichment analysis results (Figure 3F-I), and the correlation between the yellow module and the member genes (Figure 3C).

**Construction of CeRNA Network and Topological Analysis**

The 519 mRNAs in the yellow module were predicted by miRNAs in the miRTarBase and miRWalk databases, and 21,725, 149,522 mRNA-miRNA relationship pairs were obtained, respectively. And 131,846 mRNA-miRNA relationship pairs were obtained through the intersection of the two shared miRNAs. At the same time, the differential lncRNA was predicted by the miRNAs in the ENCORI and miRcode databases, and the prediction results were 15788 and 5508 lncRNA-miRNA relationship pairs respectively. The two intersected by miRNA to obtain 1253 lncRNA-miRNA relationship pairs. The mRNA-miRNA and lncRNA-
miRNA are paired by miRNA, and then crossed with the colorectal cancer-related miRNA obtained from TCGA, and finally 31900 pairs of lncRNA-miRNA-mRNA relationship pairs are obtained, and a ceRNA network is constructed based on this (Figure 4A). Among them, 349 mRNAs were significantly up-regulated, 93 lncRNAs were significantly up-regulated, 34 lncRNAs were significantly down-regulated, and 28 miRNAs were significantly down-regulated. At the same time, perform topological analysis on the ceRNA network, and select genes with a degree of greater than 5 to construct a sub-network (Figure 4B), for the next step of analysis, which contains 61 lncRNAs, 48 mRNAs, and 28 miRNAs.

**Prognostic Analysis of Network Nodes**

The 61 lncRNAs, 48 mRNAs, and 28 miRNAs in the sub-network were grouped into Kaplan-Meier survival analysis with the smallest p value, and the Kaplan-Meier survival analysis of 14 lncRNAs, 20 mRNAs and 9 miRNAs was finally obtained. Significance, as shown in Table 1.

Among the 14 lncRNAs with significant survival analysis results, 8 lncRNA high expression groups including HOTAIR, ITPK1-AS1, MYO16-AS1, WASIR2, TSPEAR-AS1, SNHG7, TTC3-AS1, WT1-AS, etc. were found to survive the time is shorter, and its HR is greater than 1, indicating that these 8 lncRNAs are prognostic risk factors, as shown in Figure 5A-H. Then the ROC curve was drawn for these 8 lncRNAs, and the AUC values were calculated. As shown in Figure 6 A-H, it can be found that the AUC values of the 8 lncRNAs are all greater than 0.5, indicating that it has good predictive value as a prognostic risk factor.

Among the 20 mRNAs with significant survival analysis, the survival time of the high expression group of STC2 and TIGD1 was shorter, and the HR was greater than 1(Figure 7). It can be found that the survival analysis results of these 2 mRNAs indicate that STC2 and TIGD1 are both prognostic risks factor. Further ROC analysis shows that STC2 and TIGD1 can find that their AUC values are both greater than 0.9 (Figure 8), indicating that they have good predictive value as prognostic risk factors.

The survival analysis results of 9 significant miRNAs are shown in Figure 9. Figure 9B, D and G results show that miRNA miR-125b-5p, miR-193a-3p and miR-363-3p survival analysis results, the high expression group. The survival time is shorter, and the survival time of the low expression group of miRNA miR-17-5p, miR-129-5p, miR-206, miR-212-3p, miR-425-5p and miR-455-5p is significantly longer Short (Figure 9A, C, E, F, H, and I). At the same time, we also performed ROC analysis on 9 miRNAs and found that their AUC values were all greater than 0.5, especially miRNA miR-17-5p, miR-125b-5p, miR-129-5p, miR-193a-3p and miR-455-5p, the AUC values are all greater than 0.95 (Figure10 A-I).

**Analysis of Expression Differences of Network Nodes**

At the same time, the survival analysis based on the sub-network nodes is significant, and we also analyzed whether there is a difference in the expression of the nodes in the tumor and normal tissues.

Through the analysis of the expression levels of the above 8 lncRNAs in TCGA, it is found that they are all highly expressed in the TCGA colorectal cancer data (Figure 11A-H). What is interesting is that mRNA STC2 and TIGD1 also show high expression (Figure 11I-J). In order to verify whether the above-mentioned
gene expression differences are universal, we used GEO data to verify their expression levels. The analysis results of some lncRNA and mRNA expression differences are shown in Figure 12. We can find that these lncRNA or mRNA also appear to be high expression in GEO data.

At the same time, we also analyzed the expression of 9 miRNAs and found that miR125b-5p, miR-129-5p and miR-363-3p are low in tumor tissues, while miR-17-5p and miR-193a-3p, miR-206, miR-212-3p, miR-425-5p and miR-455-5p are highly expressed in tumor tissues(Figure 13).

**Discussion**

In 2011, Salmena et al. proposed the establishment of a new RNA regulatory network hypothesis, that is, mRNA and lncRNA regulate each other through miRNA as a bridge. The basis for the two to regulate each other is the existence of miRNA reflection sites (MREs) (15). We already know that miRNA has a regulatory effect on mRNA, and lncRNA antagonizes this regulatory effect by competing with miRNA, which constitutes a regulatory network. This "competitive endogenous RNA" (ceRNA) communication forms a large-scale regulatory network spanning the transcriptome, which greatly expands the functional genetic information in the human genome and plays a role in pathological conditions including cancer (18). Previously, the Cancer Genome Atlas (TCGA) analyzed the relationship between gene expression data of colon cancer and the pathological stage (19), and identified potential prognostic miRNA biomarkers for predicting the overall survival of colon cancer (20). In this study, we combined the weighted gene co-expression network (WGCNA) with the ceRNA regulatory network for the first time. The purpose for finding new lncRNA-miRNA, miRNA-mRNA or lncRNA-miRNA-mRNA axis, which can regulate the development, metastasis, proliferation and invasion of colorectal cancer, is that can provide new ideas for studying the molecular mechanisms of the initiation and progression of colorectal cancer, and at the same time provide new targets for the treatment of colorectal cancer.

Studies have found that lncRNA is related to a variety of biological regulatory functions such as epigenome, transcription or posttranscriptional levels, and cancer pathogenicity (21-23). In this study, we analyzed and screened out 14 lncRNAs with significant survival analysis, including 3 down-regulated lncRNAs and 11 up-regulated lncRNAs, and further screened out 8 lncRNAs with significant high expression in colorectal cancer. Namely HOTAI, ITPK1-AS1, MYO16-AS1, WASIR2, TSPEAR-AS1, SNHG7, TTC3-AS1 and WT1-AS. Among these lncRNAs, a portion of lncRNA have been confirmed by previous molecular experiments, such as p21-mediated down-regulation of HOTAI to inhibit the proliferation, invasion and metastasis of colorectal cancer cells (24). Moreover, lncRNA SNHG7 sponging miR216b promotes liver metastasis of colorectal cancer by up-regulating GALNT1 (25), while down-regulation of SNHG7 can inhibit the phenotype of malignant bladder cancer (26). The decreased expression of another lncRNA WT1-AS can promote the proliferation and invasion of gastric cancer cells (27). The above results illustrate the reliability of our research to a certain extent. Hence, the lncRNAs, which were discovered, in this study may be used as biomarkers and have potential application value in the diagnosis, progression and treatment of colorectal cancer.
Previous studies have shown that miRNAs posttranscriptionally regulate gene expression in multifarious cancer-related signaling pathways and processes (28), and miRNAs can reduce mRNA stability or inhibit translation by binding to MRE. By analyzing TCGA colorectal cancer data, we have summarized 8 miRNAs. Some of the above miRNAs have been reported: for example, in colorectal cancer, miR-17-5p regulates EMT through targeted vitamins (29) and predicts the pathological staging and grading of colorectal cancer (30). At the same time, research reports have found that miR-206 regulates the resistance of intestinal cancer cells to 5-FU by targeting Bcl2 (31). In addition, miR-125b-5p and miR-17-5p can predict liver metastasis and chemotherapy response in advanced colorectal cancer (32). At present, the understanding of other miRNAs discovered in this study is limited, and molecular studies are needed to confirm. These colorectal cancer-specific miRNAs may become specific potential biomarkers in the diagnosis and progression of colorectal cancer some time.

The ceRNA network we constructed mainly via lncRNA-miRNA and miRNA-mRNA relationship pairs, as shown in Table 2 and Table 3. In these analysis results, some regulatory networks have been reported, such as the HOTAIR sponge miR-17-5p, which plays a tumor-promoting role in cervical cancer (33). At the same time, the HOTAIR/miR-206 axis occupies an important position in the occurrence and development of various tumors. For example, the HOTAIR sponge miR-206 regulates STC2 and affects the biological functions of head and neck squamous cell cancer cells (34). HOTAIR uses miR- The 206/TBX3 axis maintains the stemness of ovarian cancer stem cells (35) and HOTAIR regulates the proliferation of breast cancer cells through the miR-206-mediated Bcl-w signaling pathway (36). In addition, it has been reported that propofol inhibits cervical cancer progression by regulating the HOTAIR/miR-129-5p/RPL14 axis (37) and HOTAIR promotes breast cancer progression by regulating the miR-129-5p/FZD7 axis (38). Based on previous reports and our analysis results, it is found that HOTAIR plays an important role in the occurrence and development of various cancers. We can assume that the HOTAIR-miR-17-5p, HOTAIR-miR-206 or miR-129-5p axis also plays an important role in the occurrence and development of colorectal cancer, especially the HOTAIR/miR-206/STC2 axis in the colorectal It plays the same role in cancer, just like the HOTAIR/miR-206/STC2 axis in head and neck squamous cell carcinoma.

In this study, there are currently two main limitations. The first is that clinical information such as TNM, tumor staging, and histology have not been analyzed. This may be a potential limitation that needs further research to clarify; and the other is that we just based on the colorectal cancer data in TCGA, the theoretical analysis is carried out, and the specific molecular mechanism needs to be confirmed by experiments. In the future, molecular biology methods such as qPCR, luciferase reporter system, and co-immunoprecipitation will help verify our findings and clarify the molecular mechanism of the ceRNA network.

**Conclusion**

Our analysis results will provide a new molecular mechanism for the occurrence and development of colorectal cancer, thereby providing promising clues for clinical diagnosis and treatment.
Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. Research involving human participants and human data, performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable

Availability of data and materials

The data of this study are from GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA database (https://portal.gdc.cancer.gov/).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Heilongjiang Province Provincial undergraduate colleges and universities Central support for the reform and development of local colleges and universities (grant number 2020GSP05), National Natural Science Foundation of China Youth Science Fund Project (grant number 82002506) and Health and Family Planning Commission of Heilongjiang province (grant number 2018032).

Authors' contributions

XJ and JY carried out data analysis; XJ and LX drafted the manuscript; XJ and GL participated in study design and data collection. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Brody H. Colorectal cancer. Nature. 2015;521(7551):S1.

2. Wang X, Peng L, Gong X, Zhang X, Sun R, Du J. miR-423-5p Inhibits Osteosarcoma Proliferation and Invasion Through Directly Targeting STMN1. Cell Physiol Biochem. 2018;50(6):2249-59.
3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.

4. Bibbins-Domingo K. Colorectal Cancer Screening Recommendations-Reply. JAMA. 2016;316(16):1717.

5. Force USPST, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, Jr., et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2016;315(23):2564-75.

6. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394(10207):1467-80.

7. Donadon M, Ribero D, Morris-Stiff G, Abdalla EK, Vauthey JN. New paradigm in the management of liver-only metastases from colorectal cancer. Gastrointest Cancer Res. 2007;1(1):20-7.

8. Li B, Shi C, Zhao J, Li B. Long noncoding RNA CCAT1 functions as a ceRNA to antagonize the effect of miR-410 on the down-regulation of ITPKB in human HCT-116 and HCT-8 cells. Oncotarget. 2017;8(54):92855-63.

9. Chen DL, Lu YX, Zhang JX, Wei XL, Wang F, Zeng ZL, et al. Long non-coding RNA UICLM promotes colorectal cancer liver metastasis by acting as a ceRNA for microRNA-215 to regulate ZEB2 expression. Theranostics. 2017;7(19):4836-49.

10. 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(2):635-44.

11. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. Nat Rev Cancer. 2018;18(1):5-18.

12. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res. 2017;77(15):3965-81.

13. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.

14. Ou C, Sun Z, Li X, Li X, Ren W, Qin Z, et al. MiR-590-5p, a density-sensitive microRNA, inhibits tumorigenesis by targeting YAP1 in colorectal cancer. Cancer Lett. 2017;399:53-63.

15. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353-8.

16. Giroud M, Scheideler M. Long Non-Coding RNAs in Metabolic Organs and Energy Homeostasis. Int J Mol Sci. 2017;18(12).
17. Hu X, Sood AK, Dang CV, Zhang L. The role of long noncoding RNAs in cancer: the dark matter matters. Curr Opin Genet Dev. 2018;48:8-15.

18. Lin C, Yang L. Long Noncoding RNA in Cancer: Wiring Signaling Circuitry. Trends Cell Biol. 2018;28(4):287-301.

19. Zhou XG, Huang XL, Liang SY, Tang SM, Wu SK, Huang TT, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. Onco Targets Ther. 2018;11:2815-30.

20. Wei HT, Guo EN, Liao XW, Chen LS, Wang JL, Ni M, et al. Genomescale analysis to identify potential prognostic microRNA biomarkers for predicting overall survival in patients with colon adenocarcinoma. Oncol Rep. 2018;40(4):1947-58.

21. Renganathan A, Felley-Bosco E. Long Noncoding RNAs in Cancer and Therapeutic Potential. Adv Exp Med Biol. 2017;1008:199-222.

22. Slaby O, Laga R, Sedlacek O. Therapeutic targeting of non-coding RNAs in cancer. Biochem J. 2017;474(24):4219-51.

23. Sun W, Yang Y, Xu C, Guo J. Regulatory mechanisms of long noncoding RNAs on gene expression in cancers. Cancer Genet. 2017;216-217:105-10.

24. Lin K, Jiang H, Zhang LL, Jiang Y, Yang YX, Qiu GD, et al. Down-Regulated LncRNA-HOTAIR Suppressed Colorectal Cancer Cell Proliferation, Invasion, and Migration by Mediating p21. Dig Dis Sci. 2018;63(9):2320-31.

25. Shan Y, Ma J, Pan Y, Hu J, Liu B, Jia L. LncRNA SNHG7 sponges miR-216b to promote proliferation and liver metastasis of colorectal cancer through upregulating GALNT1. Cell Death Dis. 2018;9(7):722.

26. Xu C, Zhou J, Wang Y, Wang A, Su L, Liu S, et al. Inhibition of malignant human bladder cancer phenotypes through the down-regulation of the long non-coding RNA SNHG7. J Cancer. 2019;10(2):539-46.

27. Du T, Zhang B, Zhang S, Jiang X, Zheng P, Li J, et al. Decreased expression of long non-coding RNA WT1-AS promotes cell proliferation and invasion in gastric cancer. Biochim Biophys Acta. 2016;1862(1):12-9.

28. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov. 2017;16(3):203-22.

29. Kim TW, Lee YS, Yun NH, Shin CH, Hong HK, Kim HH, et al. MicroRNA-17-5p regulates EMT by targeting vimentin in colorectal cancer. Br J Cancer. 2020;123(7):1123-30.
30. Fu F, Jiang W, Zhou L, Chen Z. Circulating Exosomal miR-17-5p and miR-92a-3p Predict Pathologic Stage and Grade of Colorectal Cancer. Transl Oncol. 2018;11(2):221-32.

31. Meng X, Fu R. miR-206 regulates 5-FU resistance by targeting Bcl-2 in colon cancer cells. Onco Targets Ther. 2018;11:1757-65.

32. Sur D, Balacescu L, Cainap SS, Visan S, Pop L, Burz C, et al. Predictive Efficacy of MiR-125b-5p, MiR-17-5p, and MiR-185-5p in Liver Metastasis and Chemotherapy Response Among Advanced Stage Colorectal Cancer Patients. Front Oncol. 2021;11:651380.

33. Ji F, Wuerkenbieke D, He Y, Ding Y, Du R. Long Noncoding RNA HOTAIR: An Oncogene in Human Cervical Cancer Interacting With MicroRNA-17-5p. Oncol Res. 2018;26(3):353-61.

34. Li T, Qin Y, Zhen Z, Shen H, Cong T, Schiferle E, et al. Long non-coding RNA HOTAIR/microRNA-206 sponge regulates STC2 and further influences cell biological functions in head and neck squamous cell carcinoma. Cell Prolif. 2019;52(5):e12651.

35. Zhang Y, Guo J, Cai E, Cai J, Wen Y, Lu S, et al. HOTAIR maintains the stemness of ovarian cancer stem cells via the miR-206/TBX3 axis. Exp Cell Res. 2020;395(2):112218.

36. Ding W, Ren J, Ren H, Wang D. Long Noncoding RNA HOTAIR Modulates MiR-206-mediated Bcl-w Signaling to Facilitate Cell Proliferation in Breast Cancer. Sci Rep. 2017;7(1):17261.

37. Sun N, Zhang W, Liu J, Yang X, Chu Q. Propofol Inhibits the Progression of Cervical Cancer by Regulating HOTAIR/miR-129-5p/RPL14 Axis. Onco Targets Ther. 2021;14:551-64.

38. Wu D, Zhu J, Fu Y, Li C, Wu B. LncRNA HOTAIR promotes breast cancer progression through regulating the miR-129-5p/FZD7 axis. Cancer Biomark. 2021;30(2):203-12.

Tables

Table 1. Kaplan-Meier Survival Analysis Significantly Related RNA
| RNA        | Log-rank P | HR   | 95%CI       |
|------------|------------|------|-------------|
| IncRNA     | ARHGEF26-AS1 | 0.016 | 1.52       | 1.08-2.16 |
|            | HOTAIR     | 0.006 | 1.61       | 1.13-2.30 |
|            | ITPK1-AS1  | 0.008 | 1.65       | 1.16-2.34 |
|            | KCNQ1OT1   | 0.025 | 1.69       | 1.13-2.52 |
|            | LINC00184  | 0.035 | 0.69       | 0.49-0.98 |
|            | LINC00484  | 0.05  | 1.41       | 1.00-2.00 |
|            | MYO16-AS1  | 0.045 | 1.43       | 0.99-2.05 |
|            | SNHG7      | 0.003 | 1.72       | 1.22-2.43 |
|            | TSPEAR-AS1 | 0.025 | 1.51       | 1.02-2.24 |
|            | TTC3-AS1   | 0.024 | 1.50       | 1.06-2.11 |
|            | UCA1       | 0.011 | 0.63       | 0.42-0.94 |
|            | WASIR2     | 0.05  | 1.48       | 1.03-2.12 |
|            | WDFY3-AS2  | 0.043 | 1.57       | 1.06-2.33 |
|            | WT1-AS     | 0.018 | 1.52       | 1.05-2.21 |
| mRNA       | CCDC113    | 0.019 | 0.65       | 0.43-0.97 |
|            | CD44       | 0.013 | 0.56       | 0.38-0.83 |
|            | CDC6       | 0.006 | 0.62       | 0.44-0.88 |
|            | CDC25A     | 0.018 | 0.66       | 0.46-0.95 |
|            | CLSPN      | 0.024 | 0.65       | 0.43-0.99 |
|            | DSN1       | 0.004 | 0.59       | 0.39-0.90 |
|            | GINS4      | 0.006 | 0.62       | 0.43-0.89 |
|            | GPSM2      | <0.001| 0.48       | 0.32-0.71 |
|            | HM13       | 0.028 | 0.65       | 0.46-0.93 |
|            | NANP       | 0.007 | 0.59       | 0.41-0.89 |
|            | NEBL       | 0.027 | 0.66       | 0.45-0.99 |
|            | PAICS      | 0.001 | 0.53       | 0.38-0.75 |
|            | PLAGL2     | 0.019 | 0.66       | 0.45-0.96 |
|            | RAD18      | 0.046 | 0.70       | 0.47-1.02 |
| lncRNA  | p-value | q-value | 95% CI  |
|---------|---------|---------|---------|
| RAN     | 0.047   | 0.70    | 0.48-1.02 |
| RBM28   | 0.046   | 0.70    | 0.49-1.01 |
| RUBCNL  | 0.002   | 0.58    | 0.41-0.83 |
| SLC7A6  | 0.046   | 0.69    | 0.46-1.03 |
| STC2    | 0.044   | 1.43    | 1.01-2.02 |
| TIGD1   | 0.007   | 1.60    | 1.12-2.28 |
| miRNA   |         |         |         |
| miR-17-5p| 0.022  | 0.65    | 0.43-0.95 |
| miR-125b-5p | 0.013 | 1.56    | 1.07-2.27 |
| miR-129-5p | 0.032  | 0.66    | 0.46-0.95 |
| miR-193a-3p | 0.004  | 1.65    | 1.15-2.35 |
| miR-206  | 0.018   | 0.61    | 0.42-0.88 |
| miR-212-3p | 0.014  | 0.65    | 0.46-0.93 |
| miR-363-3p | 0.008  | 1.71    | 1.19-2.46 |
| miR-425-5p | 0.024  | 0.67    | 0.47-0.96 |
| miR-455-5p | 0.041  | 0.69    | 0.47-1.01 |

Table 2. IncRNA-miRNA relationship pairs.
| IncRNA     | miRNA          |
|-----------|----------------|
| HOTAIR    | hsa-miR-17-5p  |
|           | hsa-miR-129-5p |
|           | hsa-miR-206    |
|           | hsa-miR-193a-3p|
| ITPK1-AS1 | hsa-miR-212-3p |
|           | hsa-miR-129-5p |
|           | hsa-miR-17-5p  |
|           | hsa-miR-455-5p |
| MYO16-AS1 | hsa-miR-125b-5p|
|           | hsa-miR-129-5p |
|           | hsa-miR-425-5p |
| SNHG7     | hsa-miR-425-5p |
|           | hsa-miR-193a-3p|
| TSPEAR-AS1| hsa-miR-193a-3p|
|           | hsa-miR-212-3p |
| TTC3-AS1  | hsa-miR-193a-3p|
|           | hsa-miR-363-3p |
| WASIR2    | hsa-miR-129-5p |
|           | hsa-miR-455-5p |
|           | hsa-miR-193a-3p|
| WT1-AS    | hsa-miR-17-5p  |
|           | hsa-miR-23b-3p |
|           | hsa-miR-206    |
|           | hsa-miR-129-5p |
|           | hsa-miR-125b-5p|
|           | hsa-miR-193a-3p|

*Table 3. mRNA-miRNA relationship pairs.*
| mRNA  | miRNA             |
|-------|-------------------|
| STC2  | hsa-miR-129-5p    |
|       | hsa-miR-455-5p    |
|       | hsa-miR-125b-5p   |
|       | hsa-miR-125a-5p   |
| TIGD1 | hsa-miR-129-5p    |
|       | hsa-miR-455-5p    |
|       | hsa-miR-490-3p    |
|       | hsa-miR-212-3p    |

**Figures**

**Figure 1**

Analysis flow chart.

**Figure 2**

**Differential expression analysis of mRNA and lncRNA in TCGA colorectal cancer.**

Figure A shows the volcano map of mRNA, Figure B shows the volcano map of lncRNA, Figure C shows the clustering heat map of differentially expressed mRNA, Figure D shows the clustering heat map of differentially expressed lncRNA.

**Figure 3**

**WGCNA and enrichment analysis results.**

Figure A shows the scale-free fitting index of the network topology obtained by the soft-threshold power analysis method. Figure B shows the connectivity of characteristic genes. Red indicates positive correlation and blue indicates negative correlation. Figure C is the analysis of gene significance and module membership in the key module yellow. D-level cluster analysis is used to detect co-expression clusters with corresponding color assignments. Each color represents a module in the gene co-expression
network constructed by WGCNA. Picture E is to visualize some random genes from the network, using a heat map to describe the TOM between genes in the analysis. On a linear scale, the depth of red is positively correlated with the correlation strength between the pair of modules. The F-I diagrams are respectively the GO function annotation and KEGG pathway enrichment analysis diagram in the yellow module.

**Figure 4**

**Construction of ceRNA regulation network using two relationship pairs.**

A is a ceRNA network diagram, and B is a sub-network composed of hub nodes with a degree greater than 5 in the ceRNA network. Diamond represents IncRNA, Triangle represents mRNA, Ellipse represents miRNA, red represents up-regulated genes, green represents down-regulated genes, and blue represents genes that are not differentially expressed.

**Figure 5**

**Survival analysis of IncRNA.**

A-H is the survival analysis chart of IncRNA HOTAIR, ITPK1-AS1, MYO16-AS1, WASIR2, TSPEAR-AS1, SNHG7, TTC3-AS1, WT1-AS, respectively.

**Figure 6**

**ROC analysis of IncRNA.**

A-H are the ROC curves of IncRNA HOTAIR, ITPK1-AS1, MYO16-AS1, WASIR2, TSPEAR-AS1, SNHG7, TTC3-AS1, WT1-AS, respectively.

**Figure 7**

**Survival analysis of mRNA.**

A-B are the survival analysis diagrams of mRNA STC2 and TIGD1 respectively.
Figure 8

ROC analysis of mRNA.

A-B are the ROC curves of mRNA STC2 and TIGD1 respectively.

Figure 9

Survival analysis of miRNA.

A-I are miRNA miR-17-5p, miR-125b-5p, miR-129-5p, miR-193a-3p, miR-206, miR-212-3p, miR-363-3p, miR-425, respectively -5p and miR-455-5p survival analysis results.

Figure 10

ROC analysis of miRNA.

A-I are miRNA miR-17-5p, miR-125b-5p, miR-129-5p, miR-193a-3p, miR-206, miR-212-3p, miR-363-3p, miR-425-5p and miR-455-5p ROC analysis results.

Figure 11

Differential expression of lncRNA and mRNA in TCGA colorectal cancer data.

A-J are the expression differences of lncRNA HOTAIR, ITPK1-AS1, MYO16-AS1, WASIR2, TSPEAR-AS1, SNHG7, TTC3-AS1, WT1-AS and mRNA STC2 and TIGD1.

Figure 12

Differential expression of lncRNA and mRNA in GEO database.

Figure A shows the differential expression analysis of HOTAIR in GSE41657, Figure B and Figure C shows the differential expression analysis of SNHG7 in GSE103512 and GSE113513, respectively, and DH shows the expression analysis of STC2 in GSE156355, GSE41657, GSE110223, GSE110224, and GSE113513, respectively. I-K is the difference analysis of TIGD1 expression in GSE156355, GSE103512 and GSE113513, respectively.
Figure 13

Differential expression of miRNA in TCGA colorectal cancer data.

A-I are miRNA miR-17-5p, miR-125b-5p, miR-129-5p, miR-193a-3p, miR-206, miR-212-3p, miR-363-3p, miR-425, respectively -5p and miR-455-5p in the TCGA colorectal cancer data expression difference results.