



Identification of key modules and micro RNAs associated with colorectal cancer via a weighted gene co-expression network analysis and competing endogenous RNA network analysis

Lifang Li¹, Jingxiong Ruan², Yanfen Ma², Xin Xu², Hao Qin³, Xudong Tian², Jian Hu²

¹Emergency Department, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ²Clinical Laboratory, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ³Clinical Laboratory, Shaanxi Provincial Cancer Hospital, Xi'an, China

Contributions: (I) Conception and design: L Li, J Hu; (II) Administrative support: X Tian; (III) Provision of study materials or patients: J Ruan; (IV) Collection and assembly of data: Y Ma; (V) Data analysis and interpretation: X Xu, H Qin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jian Hu, PhD, MD. Clinical Laboratory, First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, China. Email: hobbyhujian@xjtu.edu.cn.

Background: Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide, and the incidence of CRC has increased rapidly in recent years. Due to the high invasiveness of colonoscopy and the low accuracy of alternative diagnostic methods, the diagnosis of CRC remains a serious problem. Thus, molecular biomarkers for CRC need to be identified.

Methods: In this study, RNA-sequencing data from The Cancer Genome Atlas (TCGA) database were used to identify the long non-coding RNAs (lncRNAs), messenger RNAs (mRNAs), and micro RNAs (miRNAs) that were differentially expressed between the CRC and normal tissues. Based on the gene expression and clinical features, the results of the weighted gene co-expression network analysis (WGCNA) and the binding relationships between miRNAs and lncRNAs and mRNAs were used to establish a CRC-related competing endogenous RNA (ceRNA) network.

Results: The core miRNAs (i.e., mir-874, mir-92a-1, and mir-940) in the network were identified. Among them, mir-874 was negatively correlated with the overall survival (OS) of patients. The protein-coding genes in the ceRNA network included *IZUMO4*, *WT1*, *NPEPL1*, *TEX22*, *PPFLA4*, and *SFXN3*, and the lncRNAs were *LINC00858* and *PRR7-AS1*. These genes were significantly highly expressed in CRC according to validations in other independent data sets.

Conclusions: In conclusion, this study established a network of the co-expressed ceRNAs associated with CRC and identified the genes and miRNAs related to the prognosis of CRC patients.

Keywords: Colorectal cancer (CRC); RNA-sequencing; weighted gene co-expression network analysis (WGCNA); competing endogenous RNA (ceRNA); prognosis

Submitted Mar 22, 2023. Accepted for publication Jun 09, 2023. Published online Jun 21, 2023.

doi: 10.21037/jgo-23-244

View this article at: <https://dx.doi.org/10.21037/jgo-23-244>

Introduction

Globally, colorectal cancer (CRC) is the 3rd most prevalent cancer in men and the 2nd most prevalent cancer in women (1). Historically, the incidence of CRC was far lower in China than Western nations. However, the prevalence of CRC has increased significantly in recent years and it is now

the most prevalent cancer of the digestive system (2). At the time of diagnosis, 80% of CRC patients are in an advanced stage, which has a serious effect on patient prognosis.

Currently, the prognosis and therapy options for individuals with CRC are assessed clinically based on their histological characteristics and pathological stage (3).

Patients are classified into one of the tumor, lymph node, and metastasis (TNM) stages, which range from stage I (the initial stage) to stage IV (the most advanced stage), and treatment is determined accordingly (4). However, even after stratifying and staging tumors appropriately, the best individual treatment and survival times for patients in the same stages vary greatly, which substantially impedes the accuracy of prognosis (4). Thus, a more precise molecular typing of CRC is needed, particularly for patients between stages II and III, who are the most challenging to precisely (5). The discovery of novel, more effective, and less intrusive biomarkers may increase the diagnostic and prognostic efficacy of the methods used for CRC patients.

Micro RNAs (miRNAs) are short non-coding RNAs, approximately 20–22 nucleotides in length, which exert a fine-tuning function by binding to targeted RNAs (6,7). A single miRNA affects hundreds of targeted genes primarily by causing translational repression and can also result in messenger RNA (mRNA) breakage and degradation (6,7).

Recent research has shown that the competing endogenous RNA (ceRNA) network plays a crucial role in cancer development and progression (8,9). CeRNA theory posits that in addition to their usual function of repressing RNAs, miRNAs participate in a competitive regulatory network of RNAs. Thus, any RNA that may be

targeted by miRNAs can influence the expression of genes that are targeted by the same miRNA (10,11). There is accumulating evidence that both long non-coding RNAs (lncRNAs) and miRNAs are involved in the biological processes (BPs) of cancer, which suggests that RNAs play a regulatory role. When miRNAs bind to lncRNAs, the expression of the competing mRNAs is not repressed, and the consequent changes in gene expression may contribute to the development of cancer. Currently, it has been shown that ceRNA networks play an important role in the onset and progression of several types of cancer (8,9,12). However, these studies have some limitations, such as small sample size, lack of validation in independent datasets, and incomplete consideration of clinical features. Moreover, the ceRNA networks identified by different studies are largely inconsistent, suggesting that the ceRNA interactions in CRC are complex and context-dependent. Therefore, it is necessary to perform a comprehensive and systematic analysis of the ceRNA network in CRC using large-scale and high-quality data.

In this study, we used RNA-sequencing data from The Cancer Genome Atlas (TCGA) database to identify the lncRNAs, mRNAs, and miRNAs that were differentially expressed between the CRC and normal tissues. Based on the gene expression and clinical features, we performed weighted gene co-expression network analysis (WGCNA) to identify the co-expressed modules of genes and miRNAs. Then, we integrated the results of WGCNA and the binding relationships between miRNAs and lncRNAs and mRNAs to establish a CRC-related ceRNA network. We further validated the expression and prognostic value of the core genes and miRNAs in the network using other independent datasets. Our study aimed to provide a comprehensive overview of the ceRNA interactions in CRC and to identify potential biomarkers for CRC diagnosis and prognosis. We present this article in accordance with the STREGA reporting checklist (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-23-244/rc>).

Highlight box

Key findings

- We identified the following core micro RNAs (miRNAs) in the competing endogenous (ceRNA) network related to colorectal cancer (CRC): mir-874, mir-92a-1, and mir-940.
- Mir-874 was negatively correlated with the overall survival of the patients. The protein-coding genes involved in the ceRNA network included *IZUMO4*, *WT1*, *NPEPL1*, *TEX22*, *PPF1A4*, and *SFXN3*, and the long non-coding RNAs were *LINC00858* and *PRR7-AS1*.

What is known and what is new?

- Currently, the best individual treatment and survival times for patients in the same stage vary greatly, which substantially impedes the accuracy of patient prognosis.
- This study constructed a ceRNA network for CRC to identify the miRNAs at the core of the network and their associations with patient prognosis.

What is the implication, and what should change now?

- Given the difficulty of obtaining histological samples from CRC patients, we intend to identify early novel markers to detect CRC early and predict the prognosis of CRC patients.

Methods

RNA-sequencing data from TCGA and the clinical characteristics of patients

We obtained RNA-sequencing data for CRC from TCGA database (<https://portal.gdc.cancer.gov/>), which contained 449 samples. TCGA data also included the corresponding clinical characteristics, including the overall survival (OS)

Table 1 Clinical information of the colorectal cancer patients in TCGA

| Characteristic | Numbers |
|----------------|---------|
| Age | |
| Median | 68 |
| Range (years) | 31–90 |
| Gender | |
| Male | 230 |
| Female | 209 |
| Stage | |
| Stage I | 89 |
| Stage IIC | 6 |
| Stage III | 2 |
| Stage IV | 19 |

TCGA, The Cancer Genome Atlas.

and other pathological characteristics, of the 439 CRC patients. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of differentially expressed mRNAs and lncRNAs

We converted the ENSEMBL IDs of the expression matrix to gene symbols according to GENCODE (<https://www.encodegenes.org/human/>). We identified the differentially expressed mRNAs, miRNAs, and lncRNAs between the CRC tissues and adjacent normal tissues using R package EdgeR (version 3.40.2). The mRNAs, miRNAs, and lncRNAs with an adjusted P value <0.05 and a $|\log_2$ fold change $|\geq 1$, were considered the significantly expressed RNAs associated with CRC.

Gene function and pathway enrichment analyses

We performed gene ontology (GO) (13) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (14) enrichment analyses of the significantly differentially expressed mRNAs using the R package clusterProfiler (version 4.6.0). We used the enrichment results of the GO analysis to describe the following 3 aspects of gene function: BP, molecular function (MF), and cellular component (CC). The KEGG pathway enrichment analysis results were used

to identify the possible biological pathways in which the genes were involved. The cut-off criterion for significant enrichment was set at $P < 0.05$.

Construction of the gene co-expression networks for the miRNAs, mRNAs, and lncRNAs

The gene co-expression networks were analyzed based on the expression of the significantly differently expressed miRNAs, mRNAs, and lncRNAs in the CRC samples using the R package weighted gene co-expression correlation analysis (WGCNA). Gene expression was normalized to counts per million (CPM) reads and log-transformed [$\text{Log}_{10}(\text{CPM} + 1)$]. The optimal soft threshold for the neighbor-joining calculation of the gene co-expression network was determined graphically. The transformed expression matrix was input into the functions of the WGCNA package (R, version 1.63), and the corresponding eigengenes were obtained. The cutreeDynamic function was used to prune the gene hierarchical clustering dendrogram and then produce the color-labeled co-expression modules; the related modules were then merged ($r > 0.75$).

Construction of the mRNA-miRNA-lncRNA ceRNA network

The lncRNAs and mRNAs in the key module related to CRC were selected, and the miRNAs that could target them were predicted using miRcode (<http://www.mircode.org/>). MiRDB (<http://www.mirdb.org/>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>), StarBase (<http://starbase.sysu.edu.cn/>), and TargetScan (<http://www.targetscan.org/>). In addition, Cytoscape software (National Resource for Network Biology, <https://cytoscape.org/>) was used to visualize the relationships in the ceRNA networks.

Results

Identification of the differentially expressed lncRNAs, miRNAs, and mRNAs in CRC

The RNA-sequencing expression data set from TCGA database provided the expression data for the lncRNAs, mRNAs, and miRNAs derived from the cancer and normal adjacent tissues of the 439 patients (Table 1). After normalizing the raw data using the R package DESeq2 (version 1.38.3), 4,318 significantly differentially expressed mRNAs, 126 significantly differentially expressed lncRNAs,

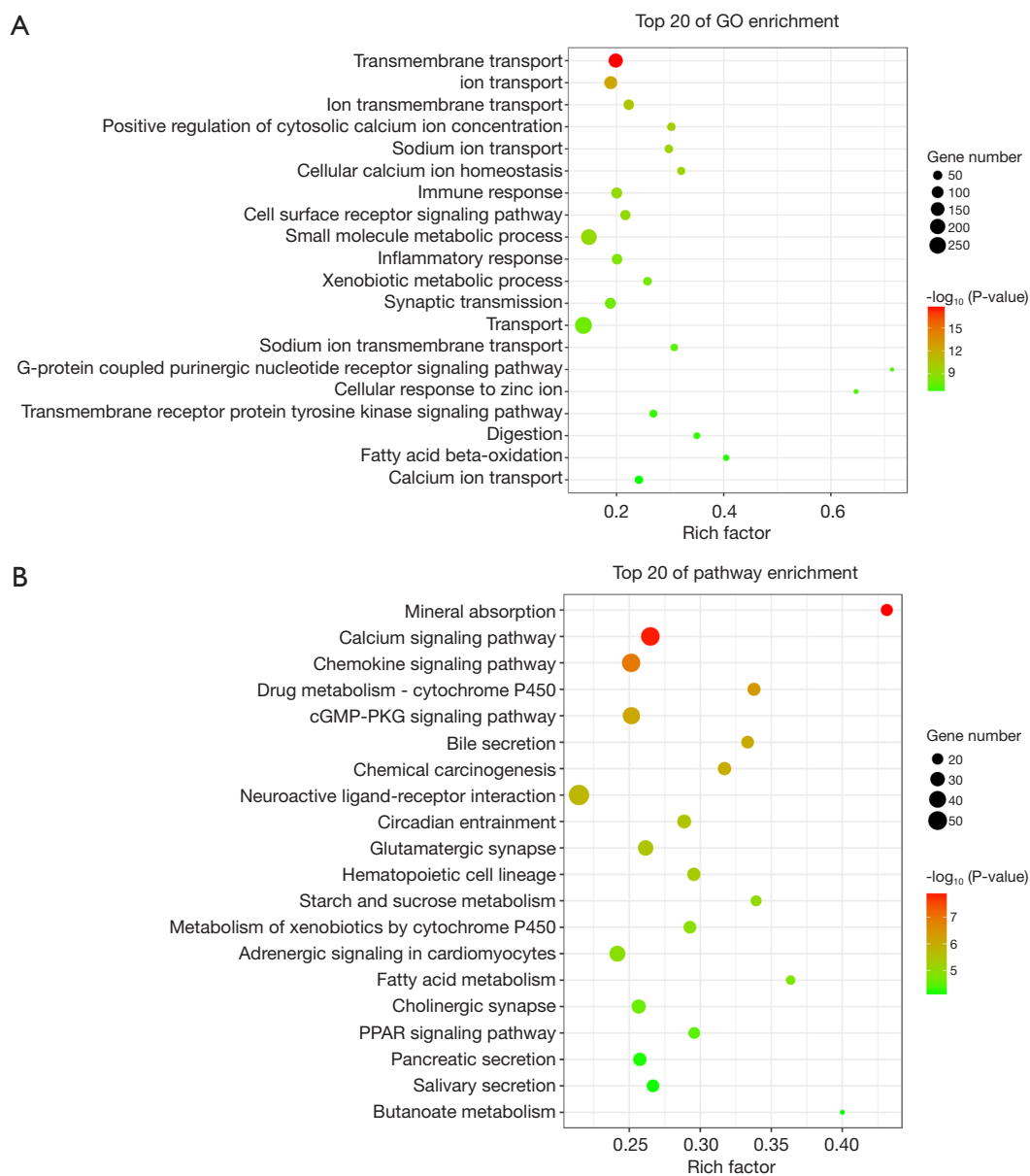


Figure 1 Gene function enrichment analysis of messenger RNAs with significant differences in expression between the colorectal cancer and the adjacent normal tissues. (A) Gene ontology functional enrichment analysis. (B) Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. GO, gene ontology.

and 428 significantly differently expressed miRNAs were identified between the CRC and adjacent normal tissues. In total, 2,386 and 1,932 significantly differently expressed mRNAs were upregulated and downregulated in the tumor tissues, respectively. In the tumor samples, 81 and 45 genes from the significantly differently expressed lncRNAs were upregulated and downregulated, respectively. In the tumor samples, 155 and 273 genes from the significantly

differently expressed miRNAs were elevated and decreased, respectively.

GO and KEGG pathway annotation and enrichment analyses were performed of the 4,318 significantly different mRNAs (Figure 1A,1B). The results of the GO analysis showed that these genes were significantly enriched in transmembrane transport, ion transport, the immune response, and the inflammatory response (Figure 1A). In

addition, the mineral absorption pathway, calcium signaling pathway, cyclic guanosine 3', 5'-monophosphate protein kinase G (cGMP-PKG) signaling pathway, and chemokine signaling pathway were also identified in the KEGG pathway enrichment analysis (Figure 1B).

Construction of the co-expressed gene modules in CRC by WGCNA

We conducted a WGCNA to construct co-expression networks and obtain the modules of the significantly differently expressed mRNAs, lncRNAs, and miRNAs in the 439 CRC patients. The Pearson correlation matrix of genes was converted to a strengthened adjacency matrix by a power β of 4 based on a scale-free topology with an R^2 of 0.97. All the selected genes were clustered using a topological overlap matrix-based dissimilarity measure, and the tree was divided into 11 modules (Figure 2A) based on a dynamic tree-cutting algorithm and labeled with different colors. The number of genes in each module is shown in Table 2. Among them, the turquoise module contained the most differently expressed mRNAs, lncRNAs and miRNAs in CRC, including 77 miRNAs, 45 lncRNAs, and 1,381 mRNAs. The genes in the same module had similar expression patterns and the first principal components [i.e., the module eigengene (MEs)] of each module were used to represent the gene expression profile of the whole module. Next, Pearson correlation coefficients were used to analyze the interactions of these co-expression modules (Figure 2B). Based on the co-expression relationships among the genes in the modules, a gene network showing the co-expression of mRNAs, lncRNAs, and miRNAs was built (available online: <https://cdn.amegroups.cn/static/public/jgo-23-244-1.xlsx>).

Construction of the lncRNA-miRNA-mRNA ceRNA network in CRC

Based on the results of the differential expression analyses described above, the lncRNAs, mRNAs, and miRNAs in CRC were classified as upregulated or downregulated genes. The upregulated mRNAs and lncRNAs corresponded to the downregulated miRNAs, and the downregulated mRNAs and lncRNAs related to the upregulated miRNAs. MiRNA target prediction software (see method) was used to estimate the binding ability of the mRNAs, lncRNAs, and miRNAs. A lncRNA and mRNA that bound to the same

miRNA was considered a ceRNA. Thus, the final ceRNA network, which included 94 miRNAs, 1,859 mRNAs, and 117 lncRNAs, was built (available online: <https://cdn.amegroups.cn/static/public/jgo-23-244-2.xlsx>).

The intersection of the miRNAs in the turquoise module and the generated ceRNAs revealed that 20 miRNAs overlapped. Based on the degree of connectivity of the genes in the ceRNA network and the co-expression network of the WGCNA, hsa-mir-874, hsa-mir-92a-1, and hsa-mir-940 were identified as key miRNAs in both the ceRNA and co-expression networks. We screened the turquoise module based on these 3 miRNAs and their target RNAs (lncRNAs and mRNAs) to construct a network of CRC-associated ceRNAs (Figure 2C).

Correlations between the lncRNAs, miRNAs, and mRNAs in the ceRNA network and the prognosis of CRC patients

Based on the gene expression and OS data from TCGA, the genes in our ceRNA network were subjected to a univariate COX regression analysis to identify the genes linked with the OS of patients. According to the findings, 5 genes in the lncRNAs were significantly negatively related with OS (Figure 3A). All 26 mRNA genes significantly linked to OS were risk factors (i.e., the higher the expression, the worse the prognosis) (Figure 3B). Among the miRNAs, hsa-mir-874 was negatively associated with the OS of patients, while whole hsa-mir-92a-1 and hsa-mir-940 were not significantly correlated with patient prognosis (Figure 3C,3D).

Validation of the expression of the core miRNAs

To assess the reliability of these 3 core miRNAs (i.e., hsa-mir-874, hsa-mir-92a-1, and hsa-mir-940) in the ceRNA network, we examined the expression profiles of these genes in additional independent data sets. In the CRC-associated miRNA data set GSE115513, the expression levels of hsa-miR-874 and hsa-miR-940 were significantly lower in the CRC samples than the normal samples (Figure 4A,4B). The expression levels of hsa-miR-92a-1-5p were also reduced in CRC, but not significantly (Figure 4C). According to the GSCALite database (Guo Lab, College of Life Science and Technology, HUST, China), all the target genes of these 3 core miRNAs were significantly and uniquely elevated in CRC (Figure 4D).

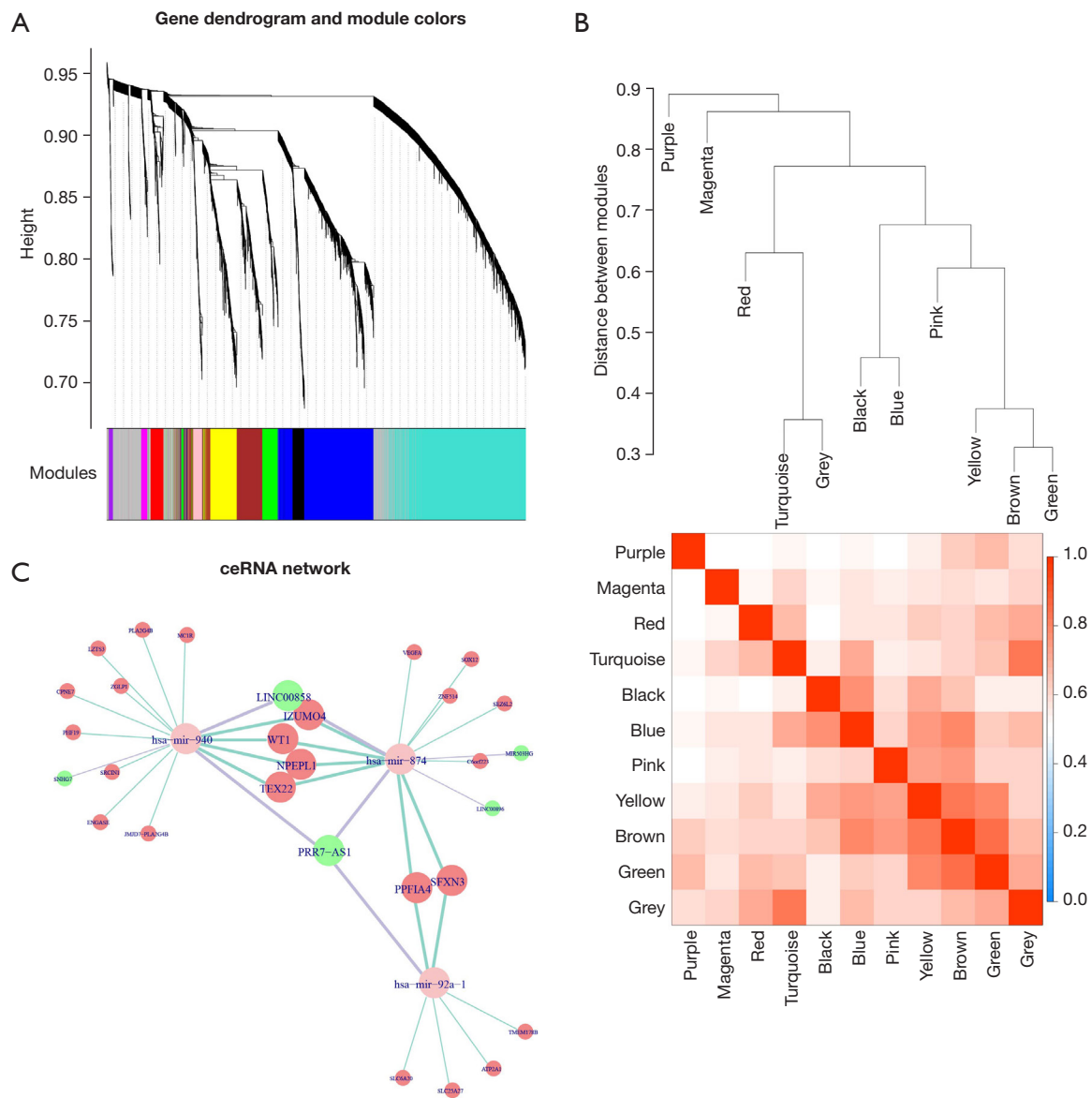


Figure 2 WGCNA and construction of the ceRNA network for the colorectal cancer-related genes. (A) Hierarchical clustering based on the WGCNA; different gene modules were obtained based on the set criteria for the cut-off clustering results; the different modules are indicated by the branches of the clustering tree and the different colors. (B) Correlation between the co-expression modules; the upper panel is the hierarchical tree based on the dissimilarity between the 2 modules; the bottom panel is a heat map of the correlation between the 2 modules. (C) ceRNA network showing the interaction between miRNA and mRNA or mRNA and lncRNAs, including the miRNAs, lncRNAs, and mRNAs in the turquoise module. The size of the nodes in the network indicates the degree of connectivity; the color indicates the fold change of gene expression in the tumor tissue compared to the adjacent normal tissue; red indicates upregulation in the tumor; green indicates downregulation in the tumor. ceRNA, competing endogenous RNA; WGCNA, weighted gene co-expression correlation analysis; miRNA, micro RNA; mRNA, messenger RNA; lncRNA, long non-coding RNA.

Table 2 The number of genes in the gene modules

| Module | No. of genes |
|-----------|--------------|
| Black | 149 |
| Blue | 985 |
| Brown | 478 |
| Green | 237 |
| Grey | 809 |
| Magenta | 71 |
| Pink | 111 |
| Purple | 47 |
| Red | 150 |
| Turquoise | 1,503 |
| Yellow | 332 |

Correlation between the expression of the core miRNAs and immune cells

To evaluate the immune cell infiltration in CRC tumors, we applied the single-sample gene set enrichment analysis (ssGSEA) algorithm to estimate the relative abundance of 24 immune cell types based on gene expression profiles. Next, we examined the correlation between two CRC-related microRNAs (miRNAs), hsa-miR-874 and hsa-miR-940, and the immune cell ratio within the tumor. Our results revealed that hsa-miR-874 was significantly and positively associated with B cells ($r=0.32$, $P<0.001$), while hsa-miR-940 was significantly and positively associated with dendritic cells (DCs) ($r=0.28$, $P<0.001$) (Figure 5). This result suggests the correlation of these core miRNAs in CRC with immune cells in the tumor microenvironment.

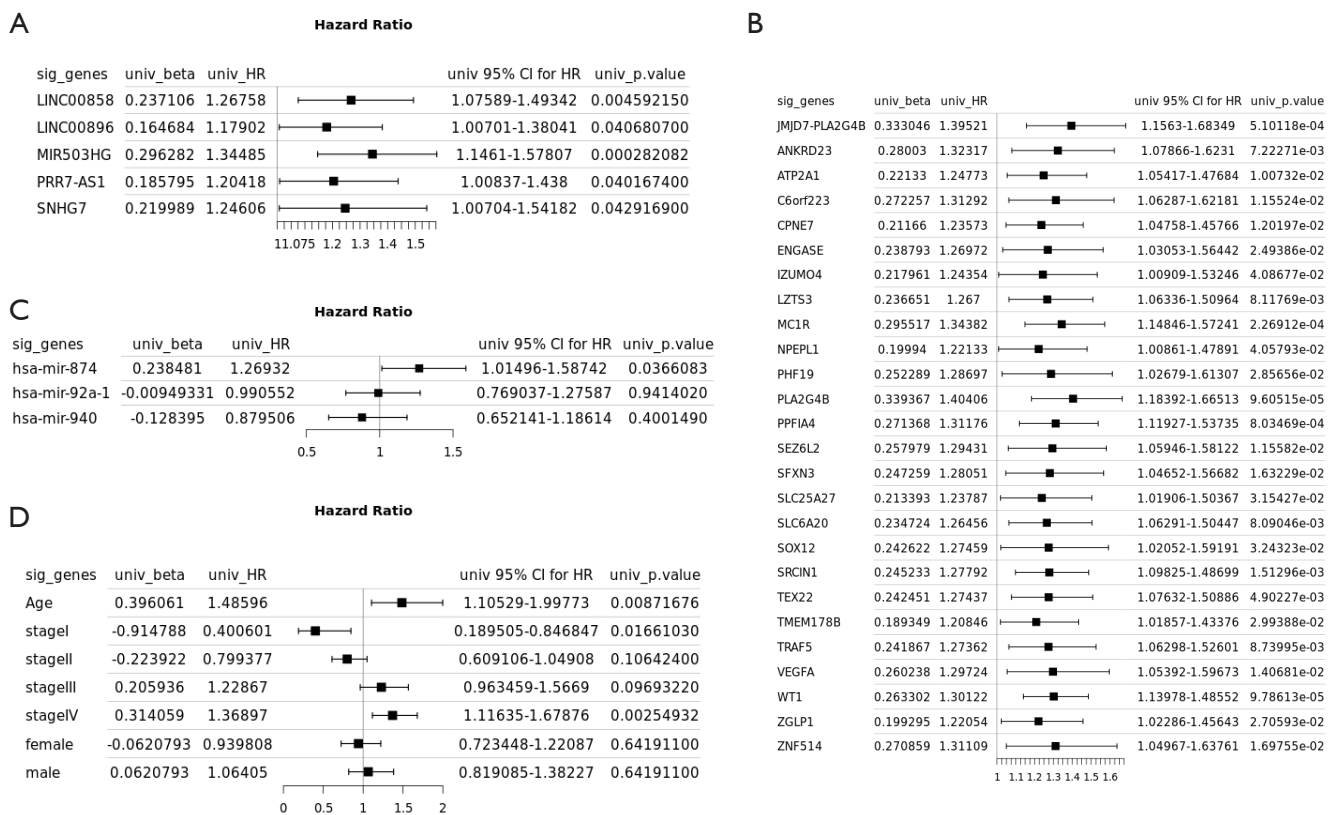


Figure 3 Univariate Cox regression analysis of the correlation between the overall survival of patients and the lncRNAs (A), mRNAs (B), and miRNAs (C) in the constructed ceRNA network and the clinical characteristics of patients (D). CI, confidence interval; HR, hazard ratio; lncRNA, long non-coding RNA; mRNA, messenger RNA; miRNA, micro RNA; ceRNA, competing endogenous RNA.

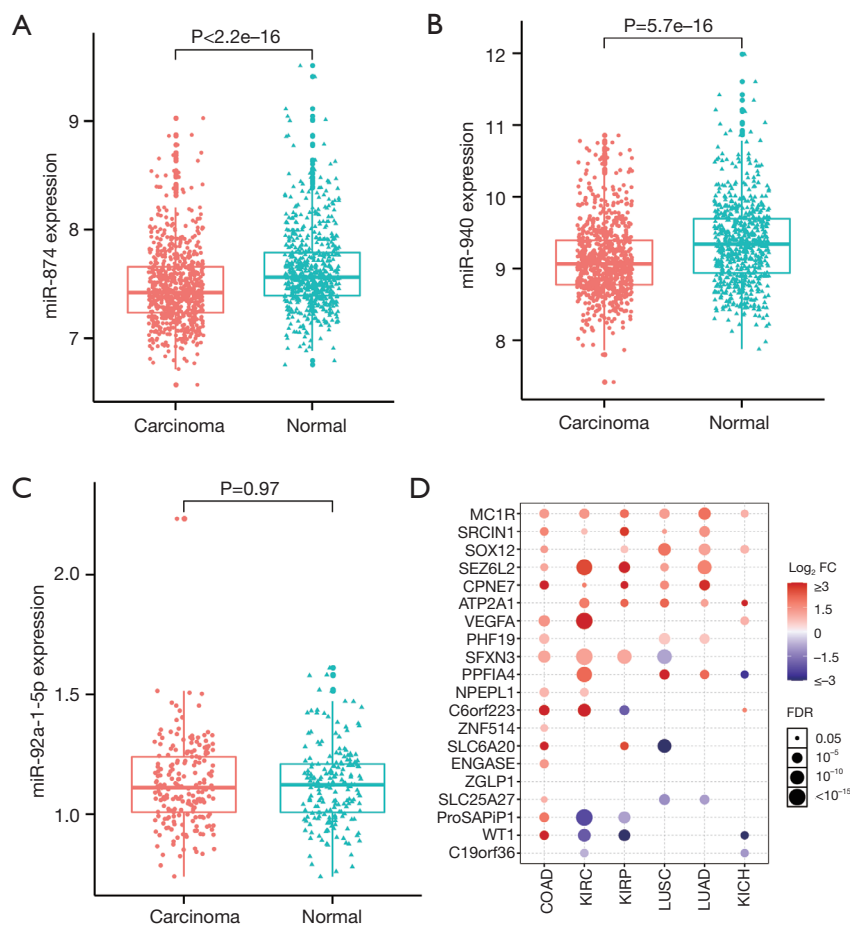


Figure 4 Validation of the expression of the 3 core miRNAs in the ceRNA network in the independent data sets. The expression levels of hsa-miR-874 (A), hsa-miR-940 (B), and hsa-miR-92a-1-5p (C) in data set GSE115513 were compared between the tumor and normal tissues. (D) The fold change (tumor *vs.* adjacent normal) of the expression of the mRNAs in the ceRNA network in COAD, KIRC, KIRP, LUSC, LUAD, and KICH according the GSCALite database (Guo Lab, College of Life Science and Technology, HUST, China). FC, fold change; FDR, false discovery rate; miRNA, micro RNA; ceRNA, competing endogenous RNA; mRNA, messenger RNA; COAD, colon adenocarcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma; KICH, kidney chromophobe.

Discussion

With >1.84 million new cases reported in 2018, CRC is the 3rd most commonly diagnosed cancer worldwide, following lung cancer and breast cancer (15,16). CRC is the 2nd deadliest malignancy in the world, and accounts for 9.2% of all cancer-related fatalities (16). The number of CRC survivors is on the rise due to improved early detection and treatment options; however, the number of patients with recurrence and metastasis after the primary tumor diagnosis is also on the rise. To enhance patient prognosis, it is crucial to identify non-invasive biomarkers with high sensitivity and

specificity in CRC for the early detection and monitoring of therapeutic responses. The histologic grade of malignancy is often used to evaluate tumor differentiation and patient prognosis, as it indicates tumor progression. However, given the difficulty of obtaining histological samples from CRC patients, we sought to investigate novel markers to detect CRC early and predict the prognosis of CRC patients.

MiRNAs are essential regulators of various biological and physiological processes, and their dysregulation is closely associated with the pathological context (6,7,17). MiRNAs are stable, easily measurable, and detectable in body fluids, and thus can be evaluated non-invasively (17-19). Given

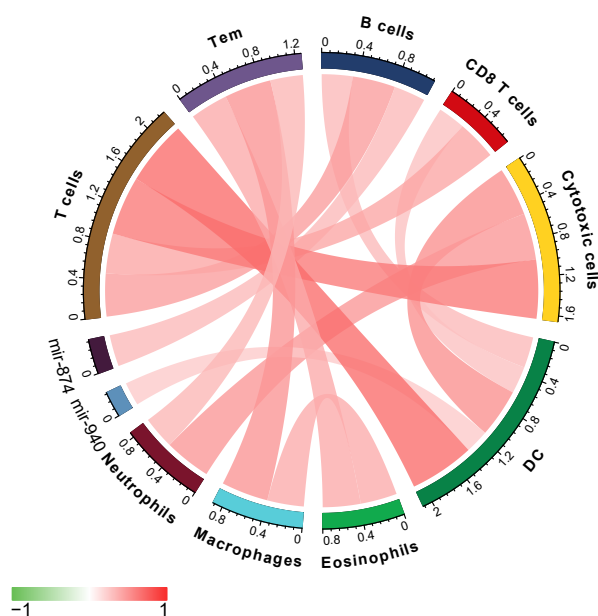


Figure 5 The correlation of expression of mir-874 and mir-940 with the immune cells in the tumors; DC, dendritic cell.

these properties, miRNAs represent promising biomarkers for the identification and therapy of CRC. By examining the ceRNA regulation network in CRC, we identified 3 core miRNAs (i.e., mir-874, mir-92a-1, and mir-940) in this study. MiR-874 has been shown to be downregulated in 22 tumors and abnormally expressed in 18 diseases other than cancer (20). MiR-874 malfunction is not only associated with the diagnosis and prognosis of tumor patients, but also has a significant effect on the efficacy of chemotherapy and radiotherapy (21). MiR-874 is engaged in multiple cancer-related signaling pathways, including the Wnt/-catenin signaling pathway and the Hippo signaling pathway. The ceRNAs in the MCF2LAS1/miR-874-3p/CCNE1 axis have been linked to apoptosis, the inhibition of cancer cell proliferation, invasion, and migration in CRC (22).

Additionally, the ceRNAs of the circ 0005576/miR-874/CDK8 axis have been shown to enhance the advancement of cancer in CRC (23). In CRC, circ 0007142 knockdown has been shown to induce ferroptosis via the regulation of the circ 0007142/miR-874-3p/GDPD5 axis, and thus to improve the efficacy of chemotherapy or radiotherapy and prevent the malignant growth of CRC cells (21). By targeting FOXD1/FOXJ1, regulating NF- κ B pathway activity, and promoting CDX2 and the downstream intestinal indicators, miR-92a-1-5p has been shown to suppress the malignant progression of CRC (24). MiR-

92a-1 is significantly more upregulated in the serum of CRC patients than that of healthy individuals (25). In addition, the expression of miR-92a-1 is associated with the TNM stage, the histological stage, lymph node metastasis, and distant metastases (25). Thus, serum miR-92a-1 may serve as a diagnostic biomarker for CRC. MiR-940 is engaged in the BPs that promote cancer development, including the Wnt/-catenin pathway, the MAPK pathway, the PD-1 pathway, and the PI3K-Akt pathway (26). The expression of miR-940 is closely associated with the diagnosis, prognosis, and chemotherapy efficacy of cancer patients (26). According to a research on exosomes in the serum of CRC patients, miR-940 may serve as a novel possible biomarker for the early diagnosis of CRC (27).

Using RNA-sequencing expression data from TCGA database, we identified the mRNAs, lncRNAs, and miRNAs that significantly between the normal and CRC tissue samples. Using a WGCNA, gene co-expression networks and binding information were merged to further identify the turquoise module. In addition, the lncRNA-miRNA-mRNA ceRNA network was constructed, and the core miRNAs of the module (i.e., mir-874, mir-92a-1, and mir-940) were identified. Among them, mir-874 was negatively correlated with the OS of patients, and it regulated genes in the network, including IZUMO4, WT1, NPEPL1, TEX22, PPFIA4, and SFXN3, and the associated lncRNAs, including LINC00858 and PRR7-AS1. These genes were also validated in multiple databases and shown to be significantly overexpressed. However, the effect of a single gene on the mechanism of CRC remains unclear. Thus, additional research needs to be conducted to verify our findings and validate the molecular mechanisms involved in the progression of CRC.

Our study identified some core genes and miRNAs that were significantly associated with the OS of CRC patients. Among them, mir-874 was negatively correlated with OS, suggesting that it may act as a tumor suppressor in CRC. In contrast, IZUMO4, WT1, NPEPL1, TEX22, PPFIA4, SFXN3, LINC00858 and PRR7-AS1 were positively correlated with OS, indicating that they may function as oncogenes in CRC. These genes and miRNAs may serve as potential prognostic biomarkers for CRC patients. Moreover, some of these genes and miRNAs have been reported to be involved in various BPs and pathways related to CRC development and progression, such as cell cycle, apoptosis, epithelial-mesenchymal transition (EMT), Wnt signaling, and immune response. Therefore, these genes and miRNAs may also provide candidate targets for the

treatment of CRC. Although our study identified a novel ceRNA network associated with CRC and revealed the potential roles of some core genes and miRNAs in CRC progression and prognosis, it still has some limitations. One of the main limitations is the lack of experimental validation and functional research on the key ceRNAs and key genes of the key modules. Future studies should perform *in vitro* and *in vivo* experiments to confirm the ceRNA interactions and to elucidate the molecular mechanisms and biological functions of the key ceRNAs and key genes in CRC development and progression.

Conclusions

In this study, we identified the lncRNAs, mRNAs, and miRNAs that were differentially expressed between CRC and normal tissues using RNA-sequencing data from TCGA database. We also established a CRC-related ceRNA network based on the gene expression and clinical features, and validated the core genes and miRNAs in other independent data sets. Our findings revealed that mir-874 was negatively correlated with the OS of CRC patients, and that IZUMO4, WT1, NPEPL1, TEX22, PPFIA4, SFXN3, LINC00858 and PRR7-AS1 were significantly highly expressed in CRC.

Our study provides novel insights into the molecular mechanisms of CRC development and progression, and suggests potential biomarkers for CRC diagnosis and prognosis. Our study contributes to the existing literature on CRC by applying a comprehensive bioinformatics approach to identify and analyze the ceRNAs associated with CRC. We hope that our study will inspire more research on the role of lncRNAs in CRC and other cancers, and facilitate the development of novel therapeutic strategies for CRC patients.

Acknowledgments

Funding: This work received funding from the Clinical Research Award of the First Affiliated Hospital of Xi'an Jiaotong University, China (No. XJTU1AF-CRF-2018-013), and the Natural Science Foundation of Shaanxi Province, Youth Project (No. 2022JQ-866).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at <https://jgo.amegroups.com/article/view/10.21037/jgo-23-244/rc>

[amegroups.com/article/view/10.21037/jgo-23-244/rc](https://jgo.amegroups.com/article/view/10.21037/jgo-23-244/rc)

Peer Review File: Available at <https://jgo.amegroups.com/article/view/10.21037/jgo-23-244/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-23-244/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
2. Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)* 2022;135:584-90.
3. Zamani M, Hosseini SV, Mokarram P. Epigenetic biomarkers in colorectal cancer: premises and prospects. *Biomarkers* 2018;23:105-14.
4. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2014;383:1490-502.
5. Bender U, Rho YS, Barrera I, et al. Adjuvant therapy for stages II and III colon cancer: risk stratification, treatment duration, and future directions. *Curr Oncol* 2019;26:S43-52.
6. Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat Rev Mol Cell Biol* 2019;20:5-20.

7. Bartel DP. Metazoan MicroRNAs. *Cell* 2018;173:20-51.
8. Kouhsar M, Azimzadeh Jamalkandi S, Moeini A, et al. Detection of novel biomarkers for early detection of Non-Muscle-Invasive Bladder Cancer using Competing Endogenous RNA network analysis. *Sci Rep* 2019;9:8434.
9. Chen JB, Zhu YW, Guo X, et al. Microarray expression profiles analysis revealed lncRNA OXCT1-AS1 promoted bladder cancer cell aggressiveness via miR-455-5p/JAK1 signaling. *J Cell Physiol* 2019;234:13592-601.
10. Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov* 2013;3:1113-21.
11. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.
12. Lin L, Zeng X, Liang S, et al. Construction of a co-expression network and prediction of metastasis markers in colorectal cancer patients with liver metastasis. *J Gastrointest Oncol* 2022;13:2426-38.
13. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-9.
14. Kanehisa M, Goto S, Furumichi M, et al. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010;38:D355-60.
15. Van der Jeught K, Xu HC, Li YJ, et al. Drug resistance and new therapies in colorectal cancer. *World J Gastroenterol* 2018;24:3834-48.
16. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
17. Condrat CE, Thompson DC, Barbu MG, et al. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* 2020;9:276.
18. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat Rev Drug Discov* 2017;16:203-22.
19. Fesler A, Jiang J, Zhai H, et al. Circulating microRNA testing for the early diagnosis and follow-up of colorectal cancer patients. *Mol Diagn Ther* 2014;18:303-8.
20. Zhang Q, Zhong C, Yan Q, et al. miR-874: An Important Regulator in Human Diseases. *Front Cell Dev Biol* 2022;10:784968.
21. Wang Y, Chen H, Wei X. Circ_0007142 downregulates miR-874-3p-mediated GSDMD on colorectal cancer cells. *Eur J Clin Invest* 2021;51:e13541.
22. Huang FK, Zheng CY, Huang LK, et al. Long non-coding RNA MCF2L-AS1 promotes the aggressiveness of colorectal cancer by sponging miR-874-3p and thereby up-regulating CCNE1. *J Gene Med* 2021;23:e3285.
23. Yu C, Li S, Hu X. Circ_0005576 Promotes Malignant Progression Through miR-874/CDK8 Axis in Colorectal Cancer. *Onco Targets Ther* 2020;13:7793-805.
24. Li T, Guo H, Li H, et al. MicroRNA-92a-1-5p increases CDX2 by targeting FOXD1 in bile acids-induced gastric intestinal metaplasia. *Gut* 2019;68:1751-63.
25. Shi Y, Liu Z. Serum miR-92a-1 is a novel diagnostic biomarker for colorectal cancer. *J Cell Mol Med* 2020;24:8363-7.
26. Li H, Li Y, Tian D, et al. miR-940 is a new biomarker with tumor diagnostic and prognostic value. *Mol Ther Nucleic Acids* 2021;25:53-66.
27. Shi Y, Zhuang Y, Zhang J, et al. Four circulating exosomal miRNAs as novel potential biomarkers for the early diagnosis of human colorectal cancer. *Tissue Cell* 2021;70:101499.

(English Language Editor: L. Huleatt)

Cite this article as: Li L, Ruan J, Ma Y, Xu X, Qin H, Tian X, Hu J. Identification of key modules and micro RNAs associated with colorectal cancer via a weighted gene co-expression network analysis and competing endogenous RNA network analysis. *J Gastrointest Oncol* 2023;14(3):1320-1330. doi: 10.21037/jgo-23-244