MicroRNA-related transcription factor regulatory networks in human colorectal cancer

Shuhong Hao, MD<sup>a,b</sup>, Sibo Huo, MD<sup>c</sup>, Zhenwu Du, PhD<sup>a,d</sup>, Qiwei Yang, PhD<sup>a</sup>, Ming Ren, PhD<sup>d</sup>, Shuai Liu, MD<sup>e</sup>, Tongjun Liu, PhD<sup>c</sup>,* Guizhen Zhang, PhD<sup>a,b,d</sup>,*

**Abstract**

**Objective** Colorectal cancer (CRC) is an extremely common gastrointestinal malignancy. The present study aimed to identify microRNAs (miRNAs) and transcription factors (TFs) associated with tumor development.

**Methods** Three miRNA profile datasets were integrated and analyzed to elucidate the potential key candidate miRNAs in CRC. The starBase database was used to identify the potential targets of common differentially expressed miRNAs (DEMs). Transcriptional Regulatory Element Database and Transcriptional Regulatory Relationships Unraveled by Sentence-based Text databases were used to identify cancer-related TFs and the TF-regulated target genes. Functional and pathway enrichment analyses were performed using the Database for Annotation, Visualization and Integration Discovery (DAVID) database, and the miRNA–TF–gene networks were constructed by Cytoscape. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to detect the expression of genes and miRNAs.

**Results** In total, 14 DEMs were found in CRC. By bioinformatics analysis, 5 DEMs (miR-145, miR-497, miR-30a, miR-31, and miR-20a) and 8 TFs (ELK4 (ETS-family transcription factor), myeloblastosis proto-oncogene like (MYBL1), MYBL2, CEBPA, PPARA, PPARD, PPARG, and endothelial PAS domain protein (EPAS1)) appeared to be associated with CRC and were therefore used to construct miRNA–TF–gene networks. From the networks, we found that miR-20a might play the most important role as an miRNA in the networks. By qRT-PCR, we demonstrated that miR-20a was significantly upregulated in CRC tissues. We also performed qRT-PCR to identify the expression of miR-20a-related TFs (PPARA, PPARD, PPARG, and EPAS1). Three of them, PPARA, PPARG, and EPAS1, were downregulated in CRC tissues, with statistically significant differences, while the downregulation of PPARD in CRC tissues was not significantly different. Pathway enrichment analyses indicated that the phosphoinositide 3-kinase (PI3K)-Akt signaling pathway was the most significantly enriched pathway. Two main elements of the PI3K-Akt signaling pathway, phosphatase and tensin homolog deleted on chromosome 10 and B-cell lymphoma 2-associated agonist of cell death, were demonstrated to be downregulated in CRC.

**Conclusion** The present study identified hub miRNAs and miRNA-related TF regulatory networks in CRC, which might be potential targets for the diagnosis and treatment of CRC.

**Abbreviations:** BAD = B-cell lymphoma 2-associated agonist of cell death, BP = biological process, CC = cellular component, CEBPA = CCAAT/enhancer binding protein A, CRC = colorectal cancer, DAVID = Database for Annotation, Visualization and Integration Discovery, DEM = differentially expressed miRNA, GEO = Gene Expression Omnibus, GO = Gene Ontology, HIF = hypoxia-inducible factor, KEGG = Kyoto Encyclopedia of Genes and Genomes, logFC = log fold change, MF = molecular function, miRNA = microRNA, mRNA = messenger RNA, NCI = National Center for Biotechnology Information, NFI = nuclear factor I, PANT = paired adjacent normal tissues, PI3K = phosphoinositide 3-kinase, PPAR = peroxisome proliferator-activated receptor, PTEN = phosphatase and tensin homolog deleted on chromosome 10, qRT-PCR = quantitative reverse transcription polymerase chain reaction, STAT = signal transducer and activator of transcription, TCGA = The Cancer Genome Atlas, TF = transcription factor, TRED = Transcriptional Regulatory Element Database, TRRUST = Transcriptional Regulatory Relationships Unraveled by Sentence-based Text, VEGF = vascular endothelial growth factor.

**Keywords:** bioinformatics, colorectal cancer, microRNA, transcription factor
1. Introduction

Colorectal cancer (CRC) is the third most common cancer and the third leading cause of cancer-related deaths in men and women worldwide.¹¹ In America alone, in 2016, an estimated 134,490 cases of CRC were diagnosed, and 49,190 patients died of the disease.²² Despite considerable advancements in the diagnosis, treatment, and understanding of the molecular mechanisms of CRC, the recurrence and metastasis of CRC continue to be closely associated with poor prognosis.³³ Therefore, a deeper understanding of the molecular mechanisms in CRC progression, as well as the identification of new therapeutic strategies and diagnostic and prognostic biomarkers, is urgently warranted.

MicroRNAs (miRNAs) are small, regulatory, noncoding RNAs that are mostly involved in messenger RNA (mRNA) degradation and post-transcriptional repression.⁴⁴ MiRNAs were first discovered in 1993, after which nearly 2000 human miRNAs have been identified, which have been shown to play important roles in various biological processes, including cell proliferation, cell cycle, apoptosis, and differentiation.⁵⁵ In humans, mounting evidence has implicated miRNA dysregulation in multiple diseases, including cancers, with different functional, tumor-suppressing, or oncogenic consequences in different contexts.⁶⁶ A study by Sun et al⁷⁷ found that miR-195-5p is dramatically downregulated in human CRC tissues compared with normal colorectal tissues, and that miR-195-5p can serve as a prognostic marker to predict the outcome of CRC patients. Qu et al⁸⁸ suggested that miR-374b inhibits colon cancer cell proliferation and invasion through the downregulation of Liver receptor homolog-1 expression. These findings highlight the promising applications of miRNAs as a novel approach for the diagnosis and therapy of patients with CRC.

MiRNAs function mainly by binding to their complementary sequence on target mRNA to negatively regulate their target genes at the post-transcriptional level.⁹⁹ Some of the targets of miRNA are genes encoding transcription factors (TFs). As the terminal regulators of gene expression, a TF binds to the DNA helix at specific regulatory sequences to activate or inhibit transcription.¹⁰¹ MiRNA can affect the development of tumors by regulating TF expression. For example, Wang et al¹¹¹ found that miR-122 repressed c-Myc transcription by targeting the transcriptional activator E2F1 and coactivator Tfdp2 in hepatocellular carcinoma. Bao et al¹²² found that in lung cancer, miR-1269 promotes cell survival and proliferation by targeting TP53, which is a TF suppressing tumor growth through the regulation of dozens of target genes with diverse biological functions.¹³¹ However, studies on the systematic analysis of miRNA-related TF regulatory networks in CRC have been limited.

Therefore, in the present study, we analyzed 3 miRNA microarray datasets from Gene Expression Omnibus (GEO) to obtain differentially expressed miRNAs (DEMs) between CRC tissues and adjacent normal colorectal tissues by GEO2R. Using the prediction software, we obtained target genes corresponding to the DEMs. In addition, we selected cancer-related TFs from the target genes to construct miRNA-TF–gene networks. The potential mechanism by which miRNA regulates the proliferation and invasion of CRC through TFs was revealed.

2. Materials and methods

2.1. CRC miRNA profiling dataset analysis

The CRC datasets of miRNA array profiling with accession numbers GSE41655, GSE35834, and GSE48267 were selected by consulting the Gene Expression Omnibus DataSets portal (GEO DataSets), publicly available on the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov/geo/), including data from 125 CRC tissues and 99 paired adjacent normal tissues (PANT) collected from 3 different regions (China, Italy, and the USA). The DEMs between CRC and PANT were screened using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), which is an interactive web tool for identifying the differential expression of genes across experimental conditions in a GEO series. Log fold change (logFC) > 1 and P < .05 for the comparison between tumor and PANT were considered statistically significant. The DEMs from the 3 cohort profile datasets were overlapped by Venny 2.1 software (http://bioinfogp.cnb.csic.es/tools/venny/index.html).¹⁴¹ Before the 3 datasets were overlapped, the miRNAs contained in all the datasets were annotated according to the last nomenclature published by miRBase v22 (http://www.mirbase.org/) because of the different microarray platforms considered.¹⁵¹ The pipeline of the whole process of this study is shown in Figure 1.

Figure 1. A schematic flow chart of the analysis steps. CRC = colorectal cancer, DEM = differentially expressed miRNA, miRNA = microRNA, TCGA = The Cancer Genome Atlas, TF = transcription factor.
2.2. Gene target analysis of selected miRNAs

The starBase v2.0 (http://starbase.sysu.edu.cn/) database is a bioinformatics prediction tool used to systematically identify the RNA–RNA and RNA–protein interaction networks from 108 CLIP-Seq (cross-linking, ligation, and sequencing of hybrids, Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation, individual-nucleotide resolution cross-linking and immunoprecipitation, high throughput sequencing-cross-linking and immunoprecipitation)-generated datasets.[16] In our study, the presumed targets of common DEMs were identified by starBase v2.0 with the selection criteria of high stringency; number of cancer types ≥ 2; and prediction in at least 2 databases among TargetScan, picTar, RNA22, PITA, and miRanda. By consulting the Transcriptional Regulatory Element Database (TRED) (http://rulai.cshl.edu/TRED), it was possible to identify 36 cancer-related TF families that are known to be involved in carcinoma development.[17] Subsequently, by overlapping the predicted target genes and 36 cancer-related TF families, DEM-targeted TFs that affect CRC proliferation were obtained.

2.3. Construction of miRNA–TF–gene networks

Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining database (TRED) version 2.0 (www.grnpedia.org/trrust), a manually curated database of human transcriptional regulatory networks,[18] was used to obtain the target genes of TFs. For the validation of gene expression, UALCAN (http://ualcan.path.uab.edu), an interactive web portal for gene expression, and survival data analysis of The Cancer Genome Atlas (TCGA) database were used in this study, with P < .05 considered statistically significant.[19] Finally, miRNA–TF–gene networks were constructed based on miRNA expression profile, starBase, TRED, UALCAN, and TRED by Cytoscape 3.6.0 software (http://www.cytoscape.org/).

2.4. Functional and pathway enrichment analysis

Gene Ontology (GO) analysis is widely used to provide gene annotation terms for large-scale genomic or transcriptomic data,[20] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) is used for pathway enrichment analysis.[21] In the present study, the functional and pathway enrichment of genes in the miRNA–TF–gene networks was analyzed using DAVID (version 6.8) (https://david.nicifcr.gov/) by performing GO and KEGG analysis.[22] The results of GO analysis comprised 3 categories: biological process (BP), cellular component (CC), and molecular function (MF); P < .05 was selected as the threshold for significant GO terms and pathways.

2.5. Tissue samples

Twenty tissue samples, including 10 CRC and 10 PANT (distance from cancer > 5 cm), were acquired by pathological assessment of tissues retrieved from surgeries, together with complete clinical information from patients at The Second Affiliated Hospital of Jilin University (Changchun, China) between February 2017 and December 2017. Samples were snap-frozen in liquid nitrogen immediately after tumor resection and stored at −80°C until further processing. No local or systemic neoadjuvant, radiotherapy, chemotherapy, or targeted therapy was provided. The study was approved by the Ethics Committee of The Second Affiliated Hospital of Jilin University. Informed consent forms were signed by all patients for the acquisition and use of tissue samples.

2.6. Total RNA extraction and quantitative polymerase chain reaction

Total RNA from clinical tissues was extracted with TRizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer’s protocol. RNA concentration was measured using the NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA). RNA (1 μg) was converted into cDNA with a First Strand cDNA Synthesis Kit (TransGen Biotech, Beijing, China), and quantitative polymerase chain reaction (qPCR) was performed with TransStart Top Green qPCR SuperMix (TransGen Biotech) on the Applied Biosystems 7500 Sequence Detection System (Thermo Scientific). For miRNA, quantitative reverse transcription PCR (qRT-PCR) was performed with the All-in-OneTM miRNA qRT-PCR Detection Kit (GeneCoopsia, Rockville, MD) on the Applied Biosystems 7500 Sequence Detection System. The relative expression levels of mRNAs and miRNAs were calculated using the 2^−ΔΔCt method[23] and were normalized to 18S and U6 RNA, respectively.

2.7. Statistical analysis

Statistical analysis of the results was performed using SPSS 22 software (SPSS, Inc., Chicago, IL). The comparisons of the mean values of the analyzed parameters were performed using Student t test. The obtained data (miRNA and mRNA levels) are presented as means ± standard deviation; P < .05 was considered to indicate a statistically significant difference.

3. Results

3.1. Identification of DEMs in CRC

CRC and PANT miRNA expression profiles of GSE41655, GSE35834, and GSE48267 were obtained from the NCBI-GEO database (Table 1). We extracted 63, 96, and 110 DEMs from the expression profile datasets GSE35834, GSE48267, and GSE41655, respectively, with thresholds of |log2FC| > 1.0 and P < .05. On overlapping the DEMs from the 3 cohort profile data sets, we obtained 14 common DEMs (Table 2, Fig. 2A), including 8 upregulated miRNAs and 6 downregulated miRNAs in the CRC tissues, compared to PANT (Table 3). Employing Multi-

| Table 1 | The information of microRNA expression profiles. |
|---------|---------------------------------------------|
| Series   | First author | Publication year | Country  | Platform   | Tumor site | Number of samples (normal/tumor) |
| GSE41655 | Shi X        | 2015            | China    | GPL11487   | CRC        | 48 (15/33) |
| GSE35834 | Bortoluzzi S | 2014            | Italy    | GPL8786    | CRC        | 54 (23/31) |
| GSE48267 | Li E         | 2015            | USA      | GPL10850   | CC         | 122 (61/61) |

CC = colon cancer, CRC = colorectal cancer.
Experiment Viewer (MeV version 4.7) (http://mev.tm4.org/), we developed a heatmap of the 8 upregulated and 6 downregulated DEMs of the 3 miRNA expression profiles, showing the significantly differential distribution of the 14 DEMs (Fig. 3).

### 3.2. Target gene prediction and cancer-related TF identification

To identify the potential target genes of the 14 DEMs, target prediction was performed using the bioinformatics tool starBase. A total of 1928 putative targets were identified (data not shown). By consulting TRED, we obtained 36 cancer-related TF families that are known to be involved in carcinoma development (Table 4). Subsequently, by overlapping the predicted target genes and TF families, we obtained 12 results (Fig. 2B), which were putative target genes of 8 miRNAs (Table 5). Because miRNAs mainly negatively regulate their target genes, the expression of target genes should run counter to the expression of their corresponding miRNAs. We further verified the expression of the 12 target genes in the TCGA colon adenocarcinoma database by UALCAN (including 41 normal colon tissues and 286 primary tumors). The results showed that the expression levels of ELK4, MYBL1, and MYBL2 were significantly higher in primary tumors than those in the normal colon for CRC patients from TCGA, in contrast to the expression levels of their corresponding miRNAs, miR-145, miR-497, and miR-30a, respectively ($P < .05$). The expression level of CEBPA was significantly lower in primary tumors than that in the normal colon for CRC patients from TCGA, in contrast to the expression level of its corresponding miRNA, miR-31 ($P < .05$). The expression levels of PPARA, PPARD, PPARG, and EPAS1 were significantly lower in primary tumors than that in the normal colon for CRC patients from TCGA, unlike the expression levels of their corresponding miRNA, miR-20a ($P < .05$) (Fig. 4). Except for the 8 genes, the expression levels of the other genes were consistent with their corresponding miRNAs or were not significantly different between primary tumor and normal colon tissue for CRC patients from TCGA ($P > .05$) (Table 5). Based on these results, we chose 5 miRNAs (miR-145, miR-497, miR-30a,...

---

**Table 2**

The number of differentially expressed microRNAs of expression profiles.

|          | Upregulation | Downregulation | Total |
|----------|--------------|----------------|-------|
| GSE41655 | 56           | 54             | 110   |
| GSE35834 | 28           | 35             | 63    |
| GSE48267 | 52           | 44             | 96    |
| Commonly DEMs | 8           | 6              | 14    |

DEMs = differentially expressed miRNAs.

---

**Table 3**

The commonly differentially expressed microRNAs of 3 microRNA expression profiles.

| Commonly DEMs | Gene symbol            |
|---------------|------------------------|
| Upregulated DEMs | hsa-miR-31, hsa-miR-224, hsa-miR-183, hsa-miR-18a, hsa-miR-20a, hsa-miR-1246, hsa-miR-552, hsa-miR-21 |
| Downregulated DEMs | hsa-miR-378, hsa-miR-378*, hsa-miR-145, hsa-miR-497, hsa-miR-30a, hsa-miR-30a* |

DEMs = differentially expressed miRNAs.

---

Figure 2. Venn diagrams. (A) Identification of common DEMs of the 3 cohort profile datasets. Different color areas represent different datasets. The overlapping areas are the common DEMs. (B) Identification of cancer-related TFs in the target genes of DEMs. The blue area represents the putative target genes of the common DEMs, and the yellow area represents 36 cancer-related TF families. DEM = differentially expressed miRNA, TF = transcription factor.
Figure 3. Clustering of the 14 DEMs in CRC vs. PANT across each independent dataset. (A) Clustering analysis of GSE41655. (B) Clustering analysis of GSE35834. (C) Clustering analysis of GSE48267. Each column represents a sample, and each row represents the expression level of a given miRNA. The color scale represents the raw score ranging from blue (low expression) to red (high expression). Dendrograms by each heatmap correspond to the hierarchical clustering by the expression of the 14 miRNAs. CRC = colorectal cancer, DEM = differentially expressed miRNA, miRNA = microRNA, PANT = paired adjacent normal tissues.

Table 4
Curated 36 cancer-related transcription factor families.

| Family | Full name | Members (official gene symbols) |
|--------|-----------|---------------------------------|
| AP1    | Activator protein 1 | FOS, FOSB, JUN, JUNB, JUND |
| AP2    | Activator protein 2 | TFAP2A, TFAP2B, TFAP2C, TFAP2D, TFAP2E |
| AR     | Androgen receptor | AR |
| ATF    | Activating transcription factor | ATF1–7 |
| BCL    | B-cell CLL/lymphoma | BCL2, BCL6 |
| BRCA   | Breast cancer susceptibility protein | BRCA1–3 |
| CEBP   | CCAAT/enhancer binding protein | CEBPA, CEBPB, CEBPD, CEBPE, CEBPG |
| CREB   | cAMP responsive element binding protein | CREB1–5, CREM |
| E2F    | E2F transcription factor | E2F1–7 |
| EGR    | Early growth response protein | EGR1–4 |
| ELK    | Member of ETS oncogene family | ELK1, ELK3, ELK4 |
| ER     | Estrogen receptor | ESR1, ESR2 |
| ERG    | ETS-related gene | ERG |
| ETS    | ETS-domain transcription factor | ETS1, ETS2, ETV4, SP1 |
| FLI1   | Friend leukemia integration site 1 | FLI1 |
| GLI    | Glioma-associated oncogene homolog | GLI1–4 |
| HIF    | Hypoxia-inducible factor | HIF1A, ARNT, EPAS1, HIF3A |
| HLF    | Hepatic leukaemia factor | HLF |
| HOX    | Homeobox gene | HOXA, HOXB, HOXD series, CHX10, MSX1, MSX2, TLX1, PBX2 |
| LEF    | Lymphoid enhancing factor | LEF1 |
| MYB    | Myeloblastosis oncogene | MYB, MYBL1, MYBL2 |
| MYC    | Myelocytomatosis viral oncogene homolog | MYC |
| NFI    | Nuclear factor I, COAT-binding transcription factor | NFI1, NFI2B, NFI3, NFI8 |
| NFKB   | Nuclear factor kappa B, reticuloendothelial oncogene | NFKB1, NFKB2, RELA, RELB, REL |
| OCT    | Octamer binding proteins | POL2F1–3, POL2F1–2, POL2F1, POL2F1 |
| p53    | P53 family | TP53, TP73, TP73 |
| PAX    | Paired box gene | PAX1–9 |
| PPAR   | Peroxisome proliferator-activated receptor | PPARA, PPARD, PPARG |
| PR     | Progesterone receptor | PGR |
| RAR    | Retinoid X receptor | RARA, RARB, RARG |
| SMAD   | Mothers against decapentaplegic homolog | SMAD1–9 |
| SP     | Sequence-specific transcription factor | SP1–8 |
| STAT   | Signal transducer and activator of transcription | STAT1–6 |
| TAL1   | T-cell acute lymphocytic leukemia-1 protein | TAL1 |
| USF    | Upstream stimulatory factor | USF1, USF2 |
| WT1    | Wilms tumor 1 (zinc finger protein) | WT1 |

cAMP = adenosine-3’-5’-cyclic monophosphate, CLL = chronic lymphocytic leukemia, ELK = ETS-family transcription factor, EPAS = endothelial PAS domain protein, ETS = E Twenty Six-specific, LRH = liver receptor homolog, MYBL = myeloblastosis proto-oncogene like.
miR-20a might be playing the most important miRNA role in the networks. MiR-20a interacted with 4 TFs (PPARA, PPARD, PPARG, and EPAS1) and indirectly interacted with 95 target genes in these networks. Second, miR-31, which indirectly interacted with 39 target genes through 1 transcription factor (CEBPA), also seemed to play an important role in this network.

3.4. Functional annotation and pathway enrichment analysis of the genes in the miRNA–TF–gene networks

We performed GO and KEGG pathway analysis for the 116 target genes in the miRNA–TF–gene networks using DAVID software. The annotation of the 116 target genes was mainly classified into 3 functional groups: BP, CC, and MF (Table 7, Fig. 6). As shown in Figure 6, in the BP group, the top 3 enriched terms were mainly cellular response to hypoxia, release of cytochrome c from mitochondria, and response to hydrogen peroxide. In terms of CC, the top 3 enriched terms were extracellular space, extracellular region, and extracellular matrix. In addition, MF analysis also revealed that the target genes were significantly enriched in growth factor activity, identical protein binding, and hormone activity. Moreover, KEGG pathway analysis indicated that the phosphoinositide 3-kinase (PI3K)-Akt, FoxO, and hypoxia-inducible factor-1 signaling pathways were the top 3 significantly enriched pathways (Table 8, Fig. 7).

3.5. Quantitative reverse transcription polymerase chain reaction

Twenty samples from 10 patients, including 6 men and 4 women, were used to perform qRT-PCR. The age at diagnosis ranged from 43 to 79 years. Primary tumors were observed in the colon of 3 patients and in the rectum of 7 patients. The histological subtypes pure adenocarcinoma and mucinous adenocarcinoma were found in 9 and 1 patients, respectively. The staging of

Table 5

| miRNA | Target genes/TFs | Expression (miRNA/target genes) |
|-------|------------------|----------------------------------|
| hsa-miR-145 | ELK4 | ↑ |
| hsa-miR-497 | MYBL1 | ↑ |
| hsa-miR-30a | MYBL2 | ↑ |
| hsa-miR-31 | CEBPA | ↑ |
| hsa-miR-20a | PPARA, PPARD, PPARG, EPAS1 | ↑ |
| hsa-miR-18a | HIF1A, NFIB | ↓ |
| hsa-miR-21 | NFIA, NFIB | ↓ |
| hsa-miR-224 | NFIA, NFIB | ↓ |

1 = higher expressed in colorectal cancer, ↓ = lower expressed in colorectal cancer, = no differential expression between colorectal cancer and normal colon, CEBPA = CCAAT enhancer binding protein A, HIF1A = hypoxia-inducible factor, miRNA = microRNA, MYBL = myeloblastosis proto-oncogene like, NF1 = nuclear factor I, PPAR = peroxisome proliferator-activated receptor, TF = transcription factor.
Table 6
Differently expressed target genes of the transcription factors.

| Transcription factor | Target genes |
|----------------------|--------------|
| ELK4                 | NR2C2        |
| MYBL1                | DHRS2, NCL   |
| MYBL2                | CCNA1, CCN1, COL1A1, MYBL2, MYC, NCL |
| CEBPA                | APF, ALB, BCL2, CD163, CSF1, CSF2, CSF3, DEFA1, DEFA3, ELANE, FN1, GAPDH, HAMP, HK3, HOMER3, HS01711, HS014, ICAM1, IGF1, IGF2, IL10, IL5, IL6, INS, ITGAX, MAD11, MS1, MYC, PKC1, PKC2, PL1, S100A9, SAA1, SLC10A1, SP1, SP2, STAR, STAT3, TLR9, VLDLR |
| PPARA                | ACADL, ACOX1, ACSL1, APOB, APOB1, CD36, CETP, CG5, CPT1A, FABP4, G0S2, HMOX1, HGFPB1, IL6, IL6R, LPL, MLYCD, MMP9, NF18A, PDK4, PLN2, PTGS2, SERPINE1, SLC25A20, SLC0A1, SOD2, SREBF2, TUBA1B, TXNIP, UGT1A9 |
| PPARD                | ANG, CAT, HS01182, PKD4, SAT1, SOD1, TIMP1, TXN |
| PPARG                | ABOG, ABO, BCL1, BCL2, BCL2, CAT, CAV1, CAV1, CCN1, CCN1, CD36, CD41A, CEBP, CDD1, CPT1B, CPT1B, DEF1, DEF3, DEF1, EGFR, ERO1A, FASL, G0S2, HMOX1, ICAM1, KLF4, MMP1, MPP, MYC, NOTCH4, NR1H3, PDK2, PON1, PTEN, PTGS2, REI, RETN, SAT1, SERPINE1, SLC2A4, SLC0A1, SOD1, TGF, TNFRSF11B, TP53, THSR, TXNIP, TSHR, UGT1A9, VLDLR |
| EPAS1                | CA9, CCR7, COL10A1, FLT1, MMP14, MEC, PTPRZ1, SERPINE1, VEGFA |

CEBPA = CCAAT/enhancer binding protein A, PPAR = peroxisome proliferator-activated receptor, MYBL = myeloblastosis proto-oncogene like.

Figure 5. MiRNA–TF–gene regulatory networks for CRC. The red square nodes represent miRNAs; green diamond nodes, TFs; and blue ellipsoidal nodes, the target genes of TFs. The lines represent the interaction between 2 nodes, with red for activation, green for repression, and gray for unknown. CRC = colorectal cancer, miRNA = microRNA, TF = transcription factor.
Table 7
Gene annotation of the target genes of transcription factor.

| GO name                                                                 | GO_ID      | P value     | Count |
|-------------------------------------------------------------------------|------------|-------------|-------|
| **1. Top 10 enriched biological process of target genes.**              |            |             |       |
| Cellular response to hypoxia                                            | GO:0071456 | 2.35E-06    | 6     |
| Release of cytochrome c from mitochondria                               | GO:0019538 | 8.08E-05    | 4     |
| Response to hydrogen peroxide                                           | GO:0042442 | 2.18E-04    | 4     |
| Response to gamma radiation                                             | GO:0013332 | 3.83E-04    | 4     |
| Regulation of multicellular organism growth                             | GO:0040014 | 3.83E-04    | 4     |
| Regulation of mitochondrial membrane Permeability                       | GO:0046602 | 3.85E-04    | 3     |
| Transcription, DNA-templated                                           | GO:0006351 | 5.00E-04    | 10    |
| Regulation of mitochondrial membrane potential                           | GO:0051881 | 6.14E-04    | 4     |
| Negative regulation of apoptotic process                                | GO:0045606 | .01055201   | 8     |
| Intrinsic apoptotic signaling pathway in response to DNA damage         | GO:0008630 | .0103699    | 4     |
| **2. Top 10 enriched cellular component of target genes.**              |            |             |       |
| Extracellular space                                                     | GO:0005615 | 9.42E-09    | 24    |
| Extracellular region                                                    | GO:0005576 | .01073224   | 11    |
| Extracellular matrix                                                    | GO:0031012 | .00490995   | 5     |
| Mitochondrion                                                           | GO:0005739 | .00654697   | 13    |
| Cytoplasm                                                               | GO:0005829 | .00838687   | 13    |
| Mitochondrial outer membrane                                            | GO:0005741 | .00370206   | 4     |
| RNA polymerase II transcription factor complex                           | GO:0090575 | .01002836   | 3     |
| Myelin sheath                                                           | GO:0043209 | .01535444   | 5     |
| Extracellular exosome                                                   | GO:0070652 | .01605403   | 21    |
| Mitochondrial matrix                                                   | GO:0005759 | .02473456   | 4     |
| **3. Top 10 enriched molecular function of target genes.**              |            |             |       |
| Growth factor activity                                                  | GO:0005833 | 6.60E-05    | 6     |
| Identical protein binding                                               | GO:0042802 | 3.03E-04    | 5     |
| Hormone activity                                                        | GO:0005179 | 9.41E-04    | 5     |
| Kinase activity                                                         | GO:0016301 | .00219826   | 5     |
| Cytokine activity                                                       | GO:0005125 | .00344331   | 6     |
| Protein heterodimerization activity                                     | GO:0046082 | .004499169  | 4     |
| Ubiquitin protein ligase binding                                        | GO:0031625 | .01610962   | 4     |
| Phosphoenolpyruvate carboxykinase activity                              | GO:0004611 | .017124514  | 2     |
| Phosphoenolpyruvate carboxykinase (GTP) activity                        | GO:0004613 | .017124514  | 2     |
| Palmitoyl-CoA oxidase activity                                          | GO:0016401 | .017124514  | 2     |

GO = gene ontology.

patients following the American Classification Joint Committee on Cancer showed 2 patients classified as stage I, 3 patients as stage II, and 5 patients as stage III.

From the miRNA–TF–gene networks, we found that miR-20a might be playing the most important miRNA role for CRC. Therefore, we performed qRT-PCR to examine the expression of miR-20a and TFs that interacted with miR-20a, including PPARA, PPARD, PPARG, and EPAS1 in CRC. The relative expression of miR-20a was $1.84 \pm 0.73$ folds upregulated in tumor tissues versus the adjacent nontumor tissues ($P < .01$). The relative expression levels of PPARA, PPARD, and EPAS1 were $0.36 \pm 0.15$ ($P < .01$), $0.26 \pm 0.18$ ($P < .01$), and $0.48 \pm 0.28$ ($P < .05$) folds downregulated in tumor tissues versus adjacent nontumor tissues, respectively. The relative expression of PPARD was $0.87 \pm 0.57$ folds downregulated in tumor tissues versus the adjacent nontumor tissues, but this was not statistically significant. Because the most enriched KEGG pathway of the target genes was the PI3K-Akt signaling pathway, we chose the main PI3K-Akt signaling pathway-related elements from the genes in the miRNA–TF–gene networks, including phosphatase and tensin homolog deleted on chromosome 10 (PTEN) and B-cell lymphoma 2-associated agonist of cell death (BAD), to perform qRT-PCR verification. The relative expression levels of PTEN and BAD were $0.50 \pm 0.48$ and $0.25 \pm 0.36$ folds downregulated in tumor tissues versus adjacent nontumor tissues, respectively (Fig. 8).

4. Discussion

CRC is a digestive tract tumor with a relatively high incidence rate among various malignant tumors.[1] In the past several decades, numerous studies have been conducted to understand the causes and underlying mechanisms of CRC occurrence and development; however, the incidence and mortality of CRC remain very high globally. This might be because studies often focus on a single genetic effect.[24] However, cancer genomics is complex, with underlying networks involving a variety of factors such as noncoding RNAs, coding genes, and TFs. The overall aim of the current study was to explore the correlation among miRNAs, TFs, and target genes in CRC.

Initially, we integrated 3 cohorts of profile datasets of individuals from 3 different geographical regions (China, Italy,
and the USA) and applied bioinformatics analysis to identify 14 commonly altered DEMs. By predicting the targets of the DEMs and the associated cancer-related TFs, we finally identified 5 miRNAs (miR-145, miR-497, miR-30a, miR-31, and miR-20a) that were considered to regulate CRC proliferation through TFs. By consulting the literature, we found that all the 5 miRNAs are related to the generation and proliferation of malignant tumors.[25–29] Among the 5 miRNAs, miR-20a was considered to be the most important miRNA in CRC because it interacted with 4 TFs (PPARA, PPARD, PPARG, and EPAS1), while the others interacted with only 1 TF. MiR-20a, a member of the miR-17–92 cluster, has been shown to function as an oncomir in CRC. A study performed by Cheng et al.[30] demonstrated that miR-20a was upregulated in CRC and that it promoted CRC invasion and metastasis by downregulating Smad4. Xu et al.[31] reported that miR-20a enhances the epithelial-to-mesenchymal transition of CRC cells by modulating matrix metalloproteinases. In the present study, on the basis of miRNA microarray data and qRT-PCR verification, we demonstrated that miR-20a was upregulated in CRC. The results were consistent with several studies. However, the result of qRT-PCR was not exactly the same as that of the microarray. In the qRT-PCR test, the relative expression of miR-20a was $1.84 \pm 0.73$ folds upregulated in CRC tissues compared to normal tissues, which was less than 2 folds. The strength of microarray-based techniques lies in their ability to quantify large numbers of miRNAs simultaneously in a single experiment. However, the specificity of microarray-based miRNA tests is lower than that of qRT-PCR tests.[32] Therefore, we believe that the results of qRT-PCR are more reliable. Of course, we cannot exclude the errors in the results of qRT-PCR tests, which could partly be due to the small sample number. Fan et al.[33] reported that miR-20a overexpression suppresses the expression levels of PPAR signaling pathways in bone marrow stem cells, indicating that PPAR might be the direct target of miR-20a. The result of this study is consistent with the putative target of miR-20a in our study. Besides, in the present study, we found that the PI3K/Akt signaling pathway was the most significantly enriched pathway in the miRNA-TF-gene networks. This finding was similar to the report of Jiang et al.[34] which demonstrated that miR-20a

![Gene Ontology of target genes in the miRNA-TF-gene networks. GO analysis classified the target genes into 3 groups (i.e., biological process, cellular component, and molecular function).](image)

Figure 6. Gene Ontology of target genes in the miRNA–TF–gene networks. GO analysis classified the target genes into 3 groups (i.e., biological process, cellular component, and molecular function). GO = Gene Ontology, miRNA = microRNA, TF = transcription factor.

| Pathway name                  | Pathway ID   | P value     | Count |
|-------------------------------|--------------|-------------|-------|
| PI3K-Akt signaling pathway    | ssc04151     | 4.48E-13    | 24    |
| FoxO signaling pathway        | ssc04068     | 1.24E-11    | 16    |
| HIF-1 signaling pathway       | ssc04066     | 6.10E-11    | 14    |
| Pathways in cancer            | ssc05200     | 1.19E-10    | 23    |
| Hepatitis B                   | ssc05161     | 9.62E-09    | 14    |
| Insulin resistance            | ssc04031     | 3.73E-08    | 12    |
| PPAR signaling pathway        | ssc03320     | 4.54E-08    | 10    |
| Proteoglycans in cancer       | ssc05205     | 1.51E-07    | 14    |
| Bladder cancer                | ssc05219     | 1.65E-07    | 8     |
| Prostate cancer               | ssc05215     | 3.71E-07    | 10    |
| Small cell lung cancer        | ssc05222     | 4.11E-07    | 10    |
| Transcriptional misregulation in cancer | ssc05202 | 9.78E-07 | 12 |
| Adipocytokine signaling pathway | ssc04920 | 1.07E-06 | 9 |
| Jak-STAT signaling pathway    | ssc04630     | 3.18E-06    | 11    |
| Cytokine–cytokine receptor interaction | ssc04060 | 5.51E-06 | 13 |
| AMPK signaling pathway        | ssc04152     | 5.58E-06    | 10    |
| MicroRNAs in cancer           | ssc05206     | 4.24E-05    | 12    |
| Focal adhesion                | ssc04510     | 5.35E-05    | 11    |
| Pancreatic cancer             | ssc05212     | 6.79E-05    | 7     |
| Fatty acid degradation        | ssc00071     | 7.63E-05    | 6     |

AMPK = Mitogen-activated protein kinase; HIF = hypoxia-inducible factor, PPAR = peroxisome proliferator-activated receptor, STAT = signal transducer and activator of transcription.
suppresses multiple myeloma progression by modulating the PTEN/PI3K/Akt signaling pathway.

TFs are the final players of signal transduction cascades, which often begin with extracellular ligand-binding events, followed by signal integration and processing, ultimately resulting in the initiation or repression of target gene transcription.[35] MiRNAs and TFs combine together in a functional network to alter gene expression in cancer.[36]

In CRC, Wang et al.[37] generated a miRNA–TF regulatory network, in which 2 miRNAs (hsa-mir-25 and hsa-mir-31), 1 TF breast cancer susceptibility protein (BRCA1), and 2 genes (ADAMTSL3 and AXIN1) were identified as the hub molecules...
and as having crucial roles in CRC pathogenesis. MiRNA and TF regulatory networks are activated by the effect of the TFs on the target gene and post-transcription interactions between miRNAs and the target genes separately. In the present study, by predicting the targets of the DEMs and associated cancer-related TFs, we constructed miRNA–TF–gene networks related to the modes of transcription regulation in CRC. From the networks, we found that the most important TFs are peroxisome proliferator-activated receptors (PPARs), which are the predicted targets of miR-20a and interact with the largest number of target genes in the networks. PPARs are ligand-activated TFs that belong to the nuclear-hormone receptor superfamily. There are 3 distinct members of the PPAR family: PPARα, PPARβ/δ, and PPARγ. All 3 members of the PPAR subfamily of nuclear receptors have been shown to participate in energy metabolism: PPARα and PPARβ/δ function mostly as catabolic regulators of energy expenditure, while PPARγ regulates anabolic metabolism, playing a role in energy storage. Besides energy metabolism, PPAR subfamily members also regulate many biological processes, including cell proliferation, survival, apoptosis, and tumor growth. Gao et al. reported that PPARα functions as an E3 ubiquitin ligase to induce Bcl2 ubiquitination and degradation, inducing colon cancer cell apoptosis. PPARγ is considered to have anticarcinogenic effects in many different cancers, because of its antiproliferation, prodifferentiation, and proapoptotic properties. Shimada et al. reported that the PPARγ ligand induces apoptosis in colon cancer cells by compensating for deregulated c-Myc expression caused by mutated adenomatous polyposis coli. Whether PPARγ ligands potentiate or suppress colon carcinogenesis is debated. For example, in the above study by Shimada et al., no apoptotic effect of PPARγ was found in colon cancer cells. Meanwhile, Zhang et al. indicated that PPARγ directly regulated the transcription of neutral amino acid transporter solute carrier family 1 member 5 and glucose transporter-1 genes, leading to the uptake of neutral amino acid, activation of mammalian target of rapamycin signaling, and tumor progression in colon cancer cells. However, a study by Yang et al. demonstrated that PPARδ knockout promotes the growth of colon cancer by inducing less differentiation and accelerating vascular endothelial growth factor (VEGF) expression in tumor cells in vivo; in other words, the study indicated that PPARδ attenuates colon carcinogenesis. In the present study, the expression levels of PPARα and PPARγ were lower in CRC tumor tissues compared with adjacent nontumor tissues, as confirmed by qRT-PCR, but there was no difference in PPARδ expression, suggesting that PPARα and PPARγ might have an antitumor role in CRC. However, whether they are direct targets of miRNA-20a and how warrant further experimental verification.

The PI3K/Akt pathway is an important intracellular signal transduction pathway to control the progression of tumor cells, including apoptosis, transcription, translation, metabolism, and angiogenesis. During malignant transformation, various genetic alterations may occur in any of the PI3K pathway components, such as the receptor tyrosine kinase genes EGFR, HER2, KIT, PTEN, PIK3CA, and AKT. By KEGG analysis in the present study, we found that the PI3K/Akt signaling pathway was the most significantly enriched pathway in the miRNA–TF–gene networks. Simultaneously, we found 2 important genes in the miRNA–TF–gene networks: PTEN and BAD. As a negative regulator of PI3K/Akt signaling, PTEN is closely correlated with the carcinogenesis, progression, and prognosis of CRC. A study by Hsu demonstrated that the expression of PTEN in CRC was lower than that in normal colon mucosa, and negative expression of PTEN was correlated with tumor size and poor prognosis in CRC. As an important downstream target of PI3K/Akt signaling, BAD can promote apoptosis by binding B-cell lymphoma 2 and inhibiting its function. In the present study, the expression levels of PTEN and BAD were lower in CRC tumor tissues compared with adjacent nontumor tissues, as confirmed by qRT-PCR, suggesting that PTEN and BAD might play an inhibitory role in CRC. Nevertheless, further research needs to explore the underlying mechanisms in more detail.

In conclusion, using profile datasets of multiple cohorts and integrated bioinformatics analysis, we identified 5 miRNAs that might regulate CRC proliferation through 8 TFs and 116 target genes. On the basis of the results, we constructed miRNA-related TF regulatory networks underlying human CRC. The findings of the potential antitumor roles of PPARα and PPARγ and inhibitory roles of PTEN and BAD in CRC have improved the understanding of the molecular basis of CRC. The identified pivotal miRNAs, genes, and pathways could be used as therapeutic targets and diagnostic biomarkers in CRC.

Author contributions

Conceptualization: Guizhen Zhang.
Data curation: Zhenwu Du.
Funding acquisition: Guizhen Zhang.
Investigation: Shuhong Hao, Zhenwu Du.
Methodology: Guizhen Zhang, Shuhong Hao.
Project administration: Guizhen Zhang.
Resources: Sibo Huo, Tongjun Liu.
Software: Shuhong Hao, Ming Ren, Shui Liu.
Visualization: Qiwei Yang.
Writing – original draft: Shuhong Hao.
Writing – review & editing: Shuhong Hao.

References

[1] Siegel, R., Desantis, C., Jemal, A. Colorectal cancer statistics, 2014. CA Cancer J Clin 2014;64:10–17.
[2] Siegel, R.L., Miller, K.D., Jemal, A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7–30.
[3] Wang, F., Wang, J., Cao, X., et al. Hsa_circ_0014717 is downregulated in colorectal cancer and inhibits tumor growth by promoting p16 expression. Biomed Pharmacother 2018;98:775–82.
[4] Jia, H., Zeng, X.Q., Huang, F., et al. Integrated microRNA and mRNA sequencing analysis of age-related changes to mouse thymic epithelial cells. JUBMB Life 2018;70:678–90.
[5] Xu, F., Zhu, Y., Sun, B., Xiao, Z. Colorectal cancer characterization and therapeutic target prediction based on microRNA expression profile. Sci Rep 2016;6:20616.
[6] Bracken, C.P., Scott, H.S., Goodall, G.J. A network-biology perspective of microRNA function and dysfunction in cancer. Nat Rev Genet 2016;17:719–32.
[7] Sun, M., Song, H., Wang, S., et al. Integrated analysis identifies microRNA-195 as a suppressor of Hippo-YAP pathway in colorectal cancer. J Hematol Oncol 2017;10:79.
[8] Qu, R., Hao, S., Jin, X., et al. MicroRNA-374b reduces the proliferation and invasion of colon cancer cells by regulation of LRH-1/Wnt signaling. Gene 2018;642:354–61.
[9] Dai, Q., Li, J., Zhou, K., et al. Competing endogenous RNA: a novel post transcriptional regulatory dimension associated with the progression of cancer. Oncol Let 2015;10:2683–90.
[10] Lambertz, M., Jambon, S., Depauw, E., et al. Targeting transcription factors for cancer treatment. Molecules 2018;23:
[11] Wang B, Hsu SH, Wang X, et al. Reciprocal regulation of microRNA-122 and c-Myc in Hepatocellular Cancer: role of E2F1 and transcription factor dimerization partner 2. Hepatology 2014;59:535–46.

[12] Bao M, Song Y, Xia J, et al. miR-1269 promotes cell survival and proliferation by targeting tp53 and caspase-9 in lung cancer. Oncotargets Ther 2018;11:1721–32.

[13] Sullivan KD, Galbraith MD, Andrysik Z, et al. Mechanisms of transcriptional regulation by p53. Cell Death Differ 2018;25:133–43.

[14] Olivieros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn’s diagrams. Available at: http://bioinfogp.cnb.csic.es/tools/venny/index.html.

[15] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014;42:D68–73.

[16] Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-nlncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 2014;42:D92–7.

[17] Zhao F, Xuan Z, Liu L, et al. TRED: a transcriptional regulatory element database and a platform for in silico gene regulation studies. Nucleic Acids Res 2005;33:12–7.

[18] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014;42:D68–73.

[19] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. Gene Ontology Consortium. Nat Genet 2000;25:25–9.

[20] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:D27–30.

[21] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2 -ΔΔCt method. Methods 2001;25:402–8.

[22] Michalik L, Desvergne B, Wahli W. Peroxisome-proliferator-activated receptors and cancers: complex stories. Nat Rev Cancer 2004;4:61–70.

[23] Jiang Y, Chang H, Chen G. Effects of microRNA-20a on the proliferation, migration and apoptosis of multiple myeloma via the Pten/PI3K/AKT signaling pathway. Oncol Lett 2015;12:4759–65.

[24] Jiang Y, Chang H, Chen G. Effects of microRNA-20a on the proliferation, migration and apoptosis of multiple myeloma via the Pten/PI3K/AKT signaling pathway. Oncol Lett 2015;12:4759–65.

[25] Yokota J, Nakamura M, Fujii T, et al. High expression of RAC2 and c-Myc in Hepatocellular Cancer: role of E2F1 and transcriptional regulation by p53. Cell Death Differ 2018;25:133–43.