Identification of colorectal cancer-restricted microRNAs and their target genes based on high-throughput sequencing data

Abstract: To identify potential key microRNAs (miRNAs) and their target genes for colorectal cancer (CRC). High-throughput sequencing data of miRNA expression and gene expression (ID: GSE46622) were downloaded from Gene Expression Omnibus, including matched colon tumor, normal colon epithelium, and liver metastasis tissues from eight CRC patients. Paired \( t \)-test and NOISeq separately were utilized to identify differentially expressed miRNAs (DE-miRNAs) and genes. Then, target genes with differential expression and opposite expression trends were identified for DE-miRNAs. Combined with tumor suppressor gene, tumor-associated gene, and TRANSFAC databases, CRC-restricted miRNAs were screened out based on miRNA-target pairs. Compared with normal tissues, there were 56 up- and 37 downregulated miRNAs in metastasis tissues, as well as eight up- and 30 downregulated miRNAs in tumor tissues. miRNA-1 was downregulated in tumor and metastasis tissues, while its target oncogenes \( \text{TWIST1} \) and \( \text{GATA4} \) were upregulated. Besides, miRNA-let-7f-1-3p was downregulated in tumor tissues, which also targeted \( \text{TWIST1} \). In addition, miRNA-133b and miRNA-4458 were downregulated in tumor tissues, while their common target gene \( \text{DUSP9} \) was upregulated. Conversely, miRNA-450-b-3p was upregulated in metastasis tissues, while its target tumor suppressor gene \( \text{CEACAM7} \) showed downregulation. The identified CRC-restricted miRNAs might be implicated in cancer progression via their target genes, suggesting their potential usage in CRC treatment.

Keywords: colorectal cancer, differentially expressed microRNAs, differentially expressed genes, oncogenes, tumor suppressor genes

Introduction
Colorectal cancer (CRC) is the third most common malignancy cancer, causing 1,400,000 new cases and 694,000 deaths in 2012 worldwide.\(^1\) With great progresses in diagnosis and treatment, the overall survival of patients with early-stage CRC has been improved. However, in terms of the patients with advanced CRC, chemotherapy usually fails to achieve satisfactory therapeutic effects.\(^2\) Therefore, it is greatly important to further understand the molecular mechanism of CRC so as to discover novel therapeutic targets.

MicroRNAs (miRNAs) are a class of small noncoding RNAs participating in diverse biological processes such as cell differentiation, proliferation, and apoptosis through a myriad of targets.\(^3\) Importantly, miRNAs act as downstream regulators of driver genes or protein kinases in cancers,\(^4\) and targeting miRNAs is considered as a potential strategy to increase the specificity of therapy and to overcome drug resistance. Recently, a link between the aberrant expression of miRNAs and CRC progression has been reported.\(^5,6\) For instance, miRNA-21 has been demonstrated to be overexpressed in CRC, mediating tumor progression and cancer cell growth.\(^7,8\) On the contrary, miRNA-338-3p,
miRNA-let-7, and miRNA-145 act as growth suppressors in CRC cells. In addition, downregulated miRNA-34a exerts suppressive effects on cell proliferation through regulating E2F signaling pathway in CRC. Serum miR-17-3p and miR-106a are biomarkers that can reflect tumor dynamics and predict disease recurrence for stage II/III CRC patients. Alteration in miRNA binding sites of cell envelope-associated proteinase genes, oculocutaneous albinism, and ocular albinism genes can inhibit translation process at genomic levels and disrupt cell cycle, which may in turn induce cancer or other diseases. Similarly, in silico evidence shows that single nucleotide polymorphisms in miRNA binding sites have effect on CRC. The miR-587/PPP2R1B/pAKT/XIAP signaling axis plays an essential role in regulating response to chemotherapy in CRC. Therefore, the identification of aberrantly expressed miRNAs and their targets is important to elucidate the initiation and progression of CRC.

Next-generation sequencing can help globally identify candidate miRNAs implicated in cancer. In 2013, Röhr et al. applied high-throughput sequencing for miRNAs and mRNAs from paired normal, tumor, and metastasis tissues to explore the therapeutic application of miRNA-1 in CRC. However, although Röhr et al. have validated the great potential of miRNA-1 as therapeutic targets in CRC, there are still some important values to comprehensively and jointly analyze the miRNA and mRNA sequencing data using various bioinformatics methods.

In our study, we assumed that miRNA profile can be applied to predict clinical progression of CRC. To verify the hypothesis, bioinformatics approaches were first utilized to screen the differentially expressed miRNAs (DE-miRNAs) and genes (DEGs) in CRC based on the high-throughput data by Röhr et al. Thereafter, CRC-restricted miRNAs with differential expression were screened out according to the miRNAs-target pairs. The present findings revealed some new miRNAs implicated in CRC via their target genes, which can be used as novel therapeutic targets for CRC.

Materials and methods
Data acquisition
The high-throughput data GSE46622 deposited by Röhr et al. were downloaded from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), which contained both miRNA and mRNA sequencing data based on the platform of Illumina Genome Analyzer Ix (Illumina, San Diego, CA, USA). For miRNA sequencing, a total of 24 samples, including the matched primary colon tumor tissues (n=8), normal colon epithelium (n=8), and liver metastasis tissues (n=8) were collected from eight CRC patients undergoing surgical resection (mean age=66.63 years, six stable and two instable). In addition, GSE46622 also contained RNA-sequencing (RNA-seq) data of the paired samples collected from four of those patients. All samples were evaluated histopathologically and the required areas were macrodissected before RNA extraction. Then, RNA was extracted using the Trizol reagent (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer’s instructions. After smallRNA isolation and cDNA library preparation, RNA sequencing was performed by Illumina’s RNA-Seq prep kit and Illumina’s DGE smallRNA sample (Illumina). This study used microarray data downloaded from a public database, therefore, no ethics committee approval or patient consent were required.

Data preprocessing
The raw sequencing data were performed with quality control using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). The reads containing either more than 50% low quality bases (<18) or over 3% N content in each end of the full lengths were filtered out. Besides, such noncoding RNAs, including rRNA, tRNA, and small nuclear RNA (snRNA) were eliminated based on reference sequences from rfam11.0 (http://rfam.xfam.org/). Afterward, the clean reads were aligned against the reference human genome hg19 (http://www.genome.ucsc.edu/index.html) using Bowtie2 (http://computing.bio.cam.ac.uk/local/doc/bowtie2.html) and TopHat v1.3.3 (http://ccb.jhu.edu/software/tophat/index.shtml) with default parameters. The sequences of mature miRNAs and pre-miRNAs were obtained from mirbase20 (http://www.mirbase.org/) to identify miRNAs.

Differential analysis of miRNAs
MiRDeep2 (http://biowulf.nih.gov/apps/mirdeep2.html) emerges as a completely overhauled tool to predict miRNAs, which contains known and novel miRNAs in seven species. Herein, MiRDeep2 was used to predict the novel miRNAs and calculate the expression levels of those known and predicted miRNAs. After that, paired t-test was used to identify DE-miRNAs in the following pairwise comparisons: tumor versus normal, metastasis versus tumor, and metastasis versus normal. The P-value <0.05 was selected as the cutoff criterion.

Clustering analysis was performed for the DE-miRNAs between normal and metastasis tissues. In addition, Venn diagrams were constructed for the up- and downregulated miRNAs in the aforementioned three comparisons, respectively.
Differential analysis of genes

Cufflinks v2.1.1 (http://cufflinks.cbcb.umd.edu/index.html) 28 was applied to calculate the gene expression levels, namely fragments per kilobase of exon per million fragments mapped values. Subsequently, NOISeq (http://www.bioconductor.org/packages/release/bioc/html/NOISeq.html) 29 was used to screen DEGs in the following pairwise comparisons: tumor versus normal, metastasis versus tumor, and metastasis versus normal. The q-value $\geq 0.99$ was taken as the threshold.

Screening of target genes for DE-miRNAs

miRNAs function through transcriptionally regulating expression levels of target genes in vivo. Thus, two kinds of databases were introduced to identify targets for DE-miRNAs, including predicted databases (miRanda, PITA, MirTarget2, PicTar, and TargetScan) 30, and validated databases (miRWalk 31 and miRecords 32). The genes, which appeared in no less than two of the predicted databases or in at least one of the validated databases, were screened out to be potential targets of DE-miRNAs.

Mining of CRC-restricted miRNAs

In the aforementioned analysis, target genes of DE-miRNAs were extracted according to the predicted and the validated databases. Then, the target genes were compared with the DEGs to identify differentially expressed targets. Following, differentially expressed targets were further mined to obtain targets with opposite expression trends with the miRNAs, termed as candidate targets of DE-miRNAs. Finally, the candidate targets of DE-miRNAs were input into tumor suppressor gene database, 33 tumor-associated gene database, 34 or TRANSFAC database 35 to identify CRC-related candidate targets. Based on the pairs of miRNAs and targets, the miRNAs corresponding to CRC-related candidate targets were selected as CRC-restricted miRNAs.

Results

Screening of DE-miRNAs

Compared with normal tissues, there separately were 56 up- and 37 downregulated miRNAs in metastasis tissues, as well as eight up- and 30 downregulated miRNAs in tumor tissues (Table 1). Besides, there were 32 up- and one downregulated miRNAs in metastasis tissues relative to tumor tissues (Table 1). The number of upregulated miRNAs in metastasis tissues was more than that in tumor tissues, implying a more serious dysregulation in advanced colon cancer. Then, the cluster analysis was performed to explore the changes of the DE-miRNAs, indicating a clear separation between normal tissues and metastasis tissues (Figure 1). In addition, using MiRDeep2, a total of one, 18, and 24 upregulated novel miRNAs separately were identified in the following comparisons: tumor versus normal, metastasis versus tumor, and metastasis versus normal. On the other hand, two and six downregulated novel miRNAs were obtained from tumor and metastasis tissues in comparison to the normal tissues, respectively.

Based on the Venn diagram of upregulated miRNAs in the three comparisons (Figure 2A), a total of three common miRNAs, including miRNA-548ap-5p, miRNA-548j-5p, and miRNA-6850-5p were identified in tumor and metastasis tissues in comparison to normal tissues. A total of four common miRNAs, including miRNA-21-3p, miRNA-3648, miRNA-4306, and miRNA-936 were found in the two comparisons (metastasis versus tumor and metastasis versus normal), suggesting persistent upregulation of those miRNAs in the three kinds of tissues. According to the Venn diagram of downregulated miRNAs in different comparisons (Figure 2B),

Table 1 Differentially expressed miRNAs in different comparisons (tumor vs normal, metastasis vs tumor, and metastasis vs normal)

<table>
<thead>
<tr>
<th>Differentially expressed miRNAs</th>
<th>Comparisons</th>
<th>Number</th>
<th>List of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated miRNAs</td>
<td>Tumor vs normal</td>
<td>8</td>
<td>hsa-miR-548ap-5p, hsa-miR-548j-5p, hsa-miR-6850-5p, hsa-miR-503-3p, hsa-let-7a-5p, hsa-miR-629-5p, etc</td>
</tr>
<tr>
<td>Metastasis vs normal</td>
<td>32</td>
<td></td>
<td>hsa-miR-21-3p, hsa-miR-3648, hsa-miR-4306, hsa-miR-936, etc</td>
</tr>
<tr>
<td>Metastasis vs normal</td>
<td>56</td>
<td></td>
<td>hsa-miR-548ap-5p, hsa-miR-548j-5p, hsa-miR-6850-5p, hsa-miR-21-3p, hsa-miR-3648, hsa-miR-4306, hsa-miR-936, etc</td>
</tr>
<tr>
<td>Downregulated miRNAs</td>
<td>Tumor vs normal</td>
<td>30</td>
<td>hsa-miR-1, hsa-miR-1265, hsa-miR-133a-3p, hsa-miR-133a-5p, hsa-miR-133b, hsa-miR-338-3p, hsa-miR-497-5p, hsa-miR-605-3p, hsa-miR-6780a-5p, hsa-miR-433-3p, etc</td>
</tr>
<tr>
<td>Metastasis vs tumor</td>
<td>1</td>
<td></td>
<td>hsa-miR-548ab</td>
</tr>
<tr>
<td>Metastasis vs normal</td>
<td>37</td>
<td></td>
<td>hsa-miR-1, hsa-miR-1265, hsa-miR-133a-3p, hsa-miR-133a-5p, hsa-miR-133b, hsa-miR-338-3p, hsa-miR-497-5p, hsa-miR-605-3p, hsa-miR-548ab, hsa-miR-192-5p, etc</td>
</tr>
</tbody>
</table>

Note: Number, the number of up- or downregulated miRNAs in each comparison.

Abbreviations: miRNAs, microRNAs; vs, versus.
a total of eight common miRNAs (miRNA-1, miRNA-1265, miRNA-133a-3p, miRNA-133a-5p, miRNA-133b, miRNA-338-3p, miRNA-497-5p, and miRNA-605-3p) were found in tumor and metastasis tissues compared with normal tissues.

### Screening of DEGs

Relative to normal tissues, there separately were 1,552 up- and 207 downregulated genes in tumor tissues, as well as 992 up- and 861 downregulated genes in metastasis tissues. In addition, there were 485 up- and 1,323 downregulated genes in metastasis tissues compared with tumor tissues (Table 2).

### Identification of CRC-restricted miRNAs

To screen out the CRC-restricted miRNAs, the differentially expressed target genes were screened out and further...
minter, miRNA-1 was downregulated in both tumor and metastasis tissues relative to normal tissues, while its target oncoprotein twist family bHLH transcription factor 1 (TWIST1) and GATA binding protein 4 (GATA4) were upregulated. In addition, miRNA-133b, miRNA-let-7f-1-3p, miRNA-4458, and miRNA-338-3p were downregulated in tumor tissues and associated with CRC. Dual specificity phosphatase 9 (DUSP9) was the common target of miRNA-133b and miRNA-4458 and was upregulated in tumor tissues. Besides, miRNA-450b-3p was upregulated in metastasis tissues, and its target gene carcinoembryonic antigen (CEA)-related cell adhesion molecule 7 (CEACAM7) exhibited downregulation.

**Discussion**

In recent years, miRNAs have been identified not only as diagnostic and prognostic molecules but also as underlying therapeutic targets. In the current study, by analyzing matched normal, tumor, and metastasis tissues of eight CRC patients, we found that miRNA-1 and miRNA-let-7f-1-3p were downregulated, while their common target TWIST1 was upregulated. Besides, miRNA-133b and miRNA-4458 were found to be downregulated in tumor tissues, while their common target DUSP9 exhibited upregulation.

Quantitative real-time polymerase chain reaction (PCR) shows that miRNA-1 is downregulated, and it can mediate proliferation in CRC by targeting oncogene c-met (MET). Our results were in accordance with previous study that the expression level of miRNA-1 was lower in CRC. Especially, miRNA-let-7f-1-3p was also downregulated in the current study, whose correlation with CRC has not been recorded in previous literatures. However, miRNA-let-7 has already been considered as a growth suppressor in CRC cells, suggesting the potential involvement of miRNA-let-7f-1-3p in CRC. Moreover, as miRNAs act through their target genes, we identified a list of DEGs as the potential target genes of miRNA-1 and miRNA-let-7f-1-3p. Among those target genes, downregulated TWIST1 was commonly targeted by miRNA-let-7f-1-3p and miRNA-1. As previously reported, TWIST1, a highly conserved transcription factor, involves in the metastasis via the regulation of epithelial-to-mesenchymal transition. The aberrant expression of TWIST1 can lead to the loss of cell-cell adhesion that may further cause the arrest of G2/M phase as well as apoptosis in CRC cells. Real-time PCR declares that expression level of TWIST1 mRNA is upregulated, thus, TWIST1 has been suggested to be a potential marker of poor outcome in CRC patients. Consistently, upregulation of TWIST1 in our analysis further confirmed that TWIST1 might be a therapeutic target for inhibiting the progression of CRC. As TWIST1 was targeted by miRNA-let-7f-1-3p and miRNA-1, we inferred that miRNA-1 and miRNA-let-7f-1-3p might be considered as therapeutic targets to inhibit the initial and development of CRC.

Meanwhile, miRNA-133b and miRNA-4458 were also found to be downregulated in tumor tissues in our analysis. The downregulation of miRNA-133b in bladder cancer has been revealed by real-time PCR that might inhibit cell growth and invasion in the progression of cancer via targeting Fascin homolog 1 gene (FSCN1). At present, no published reports demonstrate the involvement of miRNA-4458 in cancer. In this study, we screened out a common target for miRNA-133b and miRNA-4458, namely overexpressed DUSP9. DUSP9 is a member of protein tyrosine phosphatases family important for controlling cell growth and cell survival in tumorigenesis. DUSP9 belongs to DUSPs identified as the negative regulators of MAPKs, which play crucial roles in CD147-promoted invasion and epithelial-to-mesenchymal transition of CRC cells. A strong expression of DUSP9 has been demonstrated in the early stages of CRC in a mouse model, we thus inferred that miRNA-133b and miRNA-4458 targeting DUSP9 might be used as novel targets for CRC treatment.

Conversely, the level of miRNA-450b-3p was identified to be overexpressed in metastasis tissues in our study, which has never been reported in CRC. CEACAM7 was the target...
Figure 3 Regulatory network of miRNAs and corresponding target genes.
Notes: The green triangles indicate the downregulated miRNAs in tumor tissues. The red circles denote the upregulated target genes of miRNAs. Blue hexagons represent upregulated target genes potentially associated with colorectal cancer.
Abbreviation: miRNAs, microRNAs.
gene of miRNA-450b-3p, whose expression level showed a downward trend in CRC. The CEA family genes have been revealed to encode a group of intercellular adhesion molecules\(^{45,46}\) and to promote metastasis of human CRC\(^{10}\). CEACAM7, a member of CEA family, has been demonstrated to be downregulated in the transgensics’ colons\(^{50}\) and in human CRC,\(^{48}\) which may function by suppressing CRC growth.\(^{51}\) Based on the previous researches, we deduced that the overexpressed miRNA-450b-3p might act as oncogene by downregulating CEACAM7 and could be suggested as a novel target to treat advanced CRC.

**Conclusion**

The screened miRNAs, especially miRNA-1 and miRNA-let-7f-1-3p targeting TWIST1, miRNA-133b and miRNA-4458 targeting DUSP9, as well as miRNA-450b-3p targeting CEACAM7, might function as tumor suppressors or oncogenes in CRC development. Those miRNAs could be suggested as potential therapeutic targets. However, their involvement in CRC was determined from bioinformatics prospective. In future works, selected miRNAs and DEGs will be validated using real-time PCR and Western blotting.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**

27. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 demonstrates to be downregulated in the transgenics’ colons and could be suggested as a potential cancer biomarker. *World J Gastroenterol*. 2011;17(9):1159–1164.


