Elevated circulating miR-182 acts as a diagnostic biomarker for early colorectal cancer

Xiangxiang Liu
Tao Xu
Xiuxiu Hu
XiaoXiang Chen
Kaixuan Zeng
Li Sun
Shukui Wang

Central Laboratory, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, People’s Republic of China

Purpose: Globally, colorectal cancer (CRC) is one of the most common cancers with high mortality. Although CRC patients in stages I–II are curable after surgical resection, due to the lack of sensitive and specific biomarkers, many patients are in the advanced stages when diagnosed. This study aimed to investigate whether circulating miRNAs in plasma could act as biomarkers for early CRC diagnosis.

Patients and methods: All healthy subjects and patients were from Nanjing First Hospital. We first selected 2 differential miRNAs by integrated analysis of 4 Gene Expression Omnibus (GEO) data sets and The Cancer Genome Atlas (TCGA) database. Next, the expression of these 2 miRNAs in tissue and plasma samples were examined through quantitative real-time polymerase chain reaction. Training phase and validation phase were designed to investigate the diagnostic utility of these differential miRNAs using receiver operating characteristic (ROC) curve analysis.

Results: After integrated analysis of 4 GEO and TCGA databases, upregulated miR-182 and miR-20a were selected to further investigate their diagnostic potential for CRC. We discovered that miR-182 and miR-20a were upregulated in CRC tissue and plasma and that circulating miR-182 and miR-20a in the plasma of CRC patients were tumor derived. The area under the ROC curve (AUC) of circulating miR-182 was 0.929 (95% CI 0.875–0.983) in the training phase and 0.891 (95% CI 0.821–0.961) in the validation phase. The AUC of circulating miR-20a expression was 0.801 (95% CI 0.695–0.906) in the training phase and 0.736 (95% CI 0.631–0.842) in the validation phase.

Conclusion: Circulating miR-182 is a novel potential biomarker for early CRC diagnosis.

Keywords: miRNAs, colorectal cancer, GEO database, TCGA database, biomarker

Introduction
Colorectal cancer (CRC) is the third leading cause of cancer death worldwide.1 Patient prognosis is inversely associated with the cancer stage when diagnosed, with almost 90% of deaths preventable if diagnosed at an early stage.2 Although traditional fecal occult blood test (FOBT) and colonoscopy are still widely used for the early detection of CRC, the lack of high sensitivity and particular specificity of FOBT, as well as the invasiveness and expensive cost of colonoscopy, makes both of them not suitable for screening of the general population. In addition, currently available serum tumor biomarkers, such as carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9), and carbohydrate antigen 72-4 (CA72-4), are neither very sensitive nor specific.3 Therefore, novel biomarkers for early direction of CRC are urgently in demand.

Correspondence: Shukui Wang
Central Laboratory, Nanjing First Hospital, Nanjing Medical University, No 32, Gongqingtuan Road, Nanjing, Jiangsu, People’s Republic of China
Tel +86 25 5227 1163
Fax +86 25 5226 9924
Email sk_wang@njmu.edu.cn
miRNAs are short single-stranded noncoding RNAs (about 22 nucleotides long) that cause either degradation of the target mRNA molecules or translational inhibition by binding to the 3′-untranslated region (UTR) of mRNAs. The expression profiles of miRNAs in cancerous tissues are significantly different from those in adjacent normal tissues (ANTS), indicating their potential role as diagnostic biomarkers. Emerging evidence has demonstrated that circulating miRNAs are promising potential biomarkers for detecting cancers with minimal invasiveness. Tumor-derived circulating miRNAs could exist in plasma and serum steadily because of their ability to resist endogenous ribonuclease activity, as well as extreme pH and temperature. There are several explanations for this phenomenon. On the one hand, circulating miRNAs could be packaged into some kinds of particles secreted by tumors, such as apoptotic bodies and exosomes. On the other hand, miRNAs could combine with protein complexes, such as argonaute (AGO) and high-density lipoproteins.

Data from Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database demonstrated that miR-182 and miR-20a are significantly upregulated in CRC tissues compared to the normal intestinal mucosa. Hence, this study aimed to evaluate whether miR-182 or miR-20a could serve as noninvasive biomarkers for the early diagnosis of CRC. Initially, we validated the expression of miR-182 and miR-20a using 20 matched CRC tissues and ANTs. The expression levels of miR-182 and miR-20a were further demonstrated to be significantly lower in postoperative plasma specimens compared to those in preoperative plasma specimens. Next, the expression and the diagnostic role of miR-182 and miR-20a were validated and explored in plasma samples from 40 CRC patients and 40 healthy controls. Then, a larger cohort of 50 CRC patients, 50 adenoma patients, and 50 healthy controls was studied to further validate the diagnostic utility of these miRNAs. In summary, we identified plasma miR-182 as a novel potential noninvasive biomarker to discriminate CRC patients from healthy individuals.

Patients and methods

Study population
CRC patients, adenoma patients, and healthy individuals were recruited from Nanjing First Hospital affiliated to Nanjing Medical University. All CRC patients were confirmed through histopathological analysis of surgical resected tumors, and the tumor stage was determined according to the tumor–node–metastasis (TNM) system. None of the patients recruited in this study received chemotherapy or radiation therapy before specimen collection. Healthy controls were collected from those participating in the physical examination. Written informed consent was obtained from all patients and healthy individuals. This study was approved by the research and ethical committee of Nanjing First Hospital.

Study design
This study included 265 participants and 3 phases, as shown in Figure 1. In the discovery phase, we conducted an integrated analysis of miRNA expression using 4

![Flow diagram of the study design](https://www.dovepress.com/)

**Figure 1** Flow diagram of the study design. **Abbreviations:** TCGA, The Cancer Genome Atlas; CRC, colorectal cancer; qRT-PCR, quantitative real-time polymerase chain reaction; ROC, receiver operating characteristic.
GEO data sets. The expression of selected miRNAs was validated in TCGA database as well as 21 paired CRC tissues and ANTs. Fifteen paired pre- and postoperative plasma samples obtained from the same patient were tested to explore the source of miR-182 and miR-20a. In the training phase, the expression of the selected miRNAs was confirmed by quantitative real-time polymerase chain reaction (qRT–PCR) in 80 samples (40 healthy subjects and 40 CRC patients at stage I). These 80 samples were used to estimate the diagnostic parameters of miR-182 and miR-20a. In the validation phase, the diagnostic parameters of the miR-182 and miR-20a calculated in the training phase were independently assessed in additional 150 subjects (50 CRC patients at stage I, 50 adenoma patients, and 50 healthy subjects). There was no significant difference in the factors of age, gender, and drinking status among the 3 groups. The clinical characteristics of these subjects are presented in Table 1.

**Analysis of GEO and TCGA databases**

Two repositories (the GEO and TCGA databases) were used to identify the miRNA profiles significantly dysregulated in CRC. First, the differentially expressed miRNAs in the 3 GEO data sets were analyzed using the online tool GEO2R. Then, the expression levels of the selected miRNAs were explored in corresponding Series Matrix files. The expression levels of the miRNAs selected in TCGA database were downloaded and analyzed using R Studio.

**Sample processing**

Plasma was collected from venous blood in EDTA blood bottles after being centrifuged at 4000 rpm for 10 min, and subsequently stored at –80°C until further analysis. Total RNA was isolated from plasma using Trizol LS reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer’s protocol. *Caenorhabditis elegans* miR-39 (Cel-miR-39) (GenePharma, Shanghai, China) was added to each sample at a final concentration of 10–4 pmol/µL, acting as the external reference.

**qRT-PCR for miRNA quantitation**

Reverse transcription and qRT-PCR for miR-182, miR-20a, external reference miR-39, and endogenous control U6 snRNA were performed using Hairpin-it™ miRNA RT-PCR Quantitation Kit (GenePharma) according to the manufacturer’s instructions. After mixing 250 μL plasma and 750 µL Trizol LS reagent together, 1 µL cel-miR-39 at the concentration of 1 µmol/L was added into the mixture, serving as external control. The reactions were initiated with denaturation at 95°C for 3 min, followed by 40 cycles of 95°C for 15 s, and 62°C for 34 s. The relative expression levels of miR-182 and miR-20a were calculated using the 2–ΔΔCt method.

\[
\Delta Ct = Ct_{miR-182/20a} - Ct_{cel-miR-39/U6} \quad (1)
\]

\[
\Delta\Delta Ct = \Delta Ct_{neoplasm} - \Delta Ct_{control} \quad (2)
\]

### Table 1 Clinical characteristics of the population enrolled in this study

<table>
<thead>
<tr>
<th>Variables</th>
<th>Discovery phase</th>
<th></th>
<th></th>
<th>Training phase</th>
<th></th>
<th></th>
<th>Validation phase</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANT</td>
<td>CRC</td>
<td>Preoperative</td>
<td>Postoperative</td>
<td>HC</td>
<td>CRC</td>
<td>HC</td>
<td>Adenoma</td>
<td>CRC</td>
</tr>
<tr>
<td>Number</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>40</td>
<td>40</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Age, years</td>
<td>56.7±9.3</td>
<td>56.7±9.3</td>
<td>53.2±10.6</td>
<td>53.2±10.6</td>
<td>55.8±9.8</td>
<td>57.2±10.1</td>
<td>56.3±8.7</td>
<td>55.8±9.2</td>
<td>57.1±7.7</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>25</td>
<td>24</td>
<td>32</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>Drinking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>16</td>
<td>14</td>
<td>19</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Ever</td>
<td>12</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>34</td>
<td>36</td>
<td>31</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>16</td>
<td>16</td>
<td>11</td>
<td>11</td>
<td>25</td>
<td>34</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue</td>
<td>21</td>
<td>21</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Plasma</td>
<td>–</td>
<td>–</td>
<td>15</td>
<td>15</td>
<td>40</td>
<td>40</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>–</td>
<td>40</td>
<td>–</td>
<td>–</td>
<td>50</td>
</tr>
<tr>
<td>II</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>III</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Abbreviations: ANT, adjacent normal tissue; CRC, colorectal cancer; HC, healthy control.
Statistical analyses
Clinicopathological characteristics among groups were compared by the chi-square test. The differential expression of miRNAs among groups was determined using analysis of variance (ANOVA) or the Student’s paired or unpaired t-test. The receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were established for discriminating patients with early-stage CRC from healthy individuals. The concomitant use of miR-182 and miR-20a in training and validation phases was evaluated with binary logistic regression analysis using SPSS 19.0 (IBM Corporation, Armonk, NY, USA). Cutoff values of the relative expression of miR-182 and miR-20a were determined by Youden’s index from the ROC curves. A P-value < 0.05 was considered statistically significant. The statistical analysis was performed using GraphPad 5.0 (GraphPad Software, La Jolla, CA, USA).

Results
Integrated analysis of 4 GEO data sets and TCGA database identified 2 significantly upregulated miRNAs in CRC tissues
Firstly, we manually found 4 GEO data sets (GSE35834, GSE38389, GSE54088, and GSE39845), using which, we conducted a comprehensive comparative analysis of miRNA expression in CRC tissue and normal intestinal mucosa. After merging these data sets, 8 consistently dysregulated miRNAs were identified (Figure 2A). The online analysis tool GEO2R revealed 5 upregulated miRNAs (miR-182, miR-133a, miR-182m, miR-20a, miR-31, miR-378m, miR-424, miR-592m, and miR-675).

Figure 2 Upregulated miR-20a and miR-182 were selected to further explore their potential diagnostic utility after an integrated analysis of GEO and TCGA databases.
Notes: (A) Venn graph of the integrated analysis of GSE38389, GSE35834, GSE39845, and GSE54088. (B–E) Although the miRNA expression values were calculated through different methods, the 4 GEO data sets showed consistent results. RMA is a kind of function in R Studio. (F) Eight selected miRNAs were further validated using the TCGA database. Data are presented as mean ± SD. *P<0.05, **P<0.01, ***P<0.001.
Abbreviations: GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; ns, not significant.
miR-20a, miR-31, miR-424, and miR-675) and 3 downregulated ones (miR-133a, miR-378, and miR-592) in the CRC tissue compared to the ANT (Figure 2B–E). Next, we further validated the 8 dysregulated miRNAs in TCGA colon adenocarcinoma (COAD) data set (268 CRC tissues, 8 normal intestinal mucosas). Only miR-133a, miR-20a, and miR-182 were significantly dysregulated in CRC tissues (Figure 2F). Considering the widely different expression of miR-20a and miR-182 between cancerous tissue and normal mucosa, as well as their abundant sequencing counts in CRC tissue, we selected them to further investigate whether the dysregulated miR-182 and miR-20a in the tumor could be secreted to the plasma, serving as potential diagnostic markers for CRC.

Circulating miR-182 and miR-20a were tumor-derived miRNAs

Twenty paired CRC tissues and ANTs were used to validate the dysregulated expression of miR-182 and miR-20a, revealing a consistently differential expression with the GEO data sets and TCGA database (Figure 3A). In order to investigate whether circulating miR-182 and miR-20a in the plasma are tumor derived, 15 pairs of pre- and postoperative plasma samples were collected, and qRT-PCR showed that the expression levels of circulating miR-182 and miR-20a were significantly decreased after surgical resection of CRC tissue (Figure 3B). These results indicated that CRC tissue is the source of the upregulated circulating miR-182 and miR-20a in plasma.

MiR-182 and miR-20a act as novel potential diagnostic biomarkers for CRC

Circulating miR-182 and miR-20a were validated in an independent cohort comprising 40 CRC patients at stage I and 40 healthy individuals. The relative expression levels of miR-182 and miR-20a were significantly higher in the CRC plasma samples than in the healthy controls (Figure 4A).

![Figure 3](https://www.dovepress.com/)

**Figure 3** CRC tissue was the main source of miR-20a and miR-182 in plasma.

**Notes:** (A) miR-182 and miR-20a were validated to be upregulated in CRC tissues compared with ANTs. (B) miR-182 and miR-20a in the plasma of CRC patients were significantly downregulated after surgical resection. The boxes represent the intervals between the 25th and 75th percentiles, and the whiskers mark the interval between the minimum and the maximum. *P*<0.05, **P*<0.01, ***P*<0.001. **Abbreviations:** ANT, adjacent normal tissue; CRC, colorectal cancer; ns, not significant.
To assess the potential diagnostic role of miRNAs, ROC curve analysis was performed. Individually, the AUC was 0.929 (95% CI 0.875–0.983) for miR-182, 0.801 (95% CI 0.695–0.906) for miR-20a, and 0.905 (95% CI 0.841–0.968) for the 2-miRNA combination (Figure 4B). The cutoff values for miR-182, miR-20a, and the 2-miRNA combination were 3.165, 1.355, and 2.255, respectively.

Validation of miRNAs as CRC diagnostic biomarkers

To further validate the potential diagnostic role of miR-182 and miR-20a, a larger cohort composed of 50 CRC patients, 50 adenoma patients, and 50 healthy controls was investigated. As shown in Figure 5A, the relative expression levels of miR-182 and miR-20a were significantly higher in the CRC patients than in the adenoma patients and healthy controls. ROC curve analysis revealed that the AUC was 0.891 (95% CI 0.821–0.961) for miR-182, 0.736 (95% CI 0.631–0.842) for miR-20a, and 0.831 (95% CI 0.746–0.914) for the 2-miRNA combination (Figure 5B). The cutoff values for miR-182, miR-20a, and the 2-miRNA combination were 2.620, 1.355, and 2.147, respectively.

**Discussion**

CRC is the third most common cancer and the second main cause of cancer death worldwide.\(^\text{15}\) Many cases of mortality could be prevented if CRC patients could be diagnosed at an early stage and treated before malignancy develops into an advanced stage.\(^\text{16}\)

Frattini et al\(^\text{17}\) reported that some kinds of plasma nucleic acids are tumor derived and that their alternations in plasma are detectable, identifying promising biomarkers for diagnosing and monitoring a variety of cancers with noninvasive techniques. miRNAs are involved in many biological and pathological processes.\(^\text{18,19}\) Increasing studies have demonstrated that tumor-derived circulating miRNAs are tissue specific and more stable than DNA, RNA, and proteins in plasma.\(^\text{10,20}\) Circulating miRNAs have been investigated for early diagnosis in several cancers, such as breast cancer,\(^\text{21}\) gastric cancer,\(^\text{8}\) bladder cancer,\(^\text{22}\) prostate cancer,\(^\text{23}\) and CRC.\(^\text{9}\)
miR-182 has been reported in various cancers, such as lung cancer, CRC, gastric cancer, and breast cancer. Li et al demonstrated miR-182 could inhibit the epithelial-to-mesenchymal transition (EMT) and the metastasis of lung cancer cells by targeting the \textit{Met} gene. In addition, miR-182 was reported to be upregulated in CRC and to promote chemoresistance through the PI3K/AKT pathway. Although circulating miR-182 has been reported to be upregulated in the plasma of CRC patients, its potential use as a promising diagnostic biomarker for early CRC has not been further investigated.

In the present study, we first systematically screened the differentially expressed miRNAs in CRC tissues using 4 GEO data sets (GSE35834, GSE38389, GSE54088, and GSE39845). Then, 8 consistently dysregulated miRNAs were further validated in TCGA database. Finally, 2 upregulated miRNAs (miR-182 and miR-20a), confirmed in independent samples, were selected to explore their potential diagnostic role in CRC. In the training phase and validation phase, the expression of plasma miR-182 and miR-20a in patients with stage I CRC was significantly higher than that in healthy individuals and adenoma patients. The AUC of miR-182 expression in plasma was 0.929 (95% CI 0.875–0.983) in the training phase and 0.891 (95% CI 0.821–0.961) in the validation phase for discriminating early CRC from healthy subjects and adenoma patients. However, the AUC of circulating miR-20a expression was only 0.801 (95% CI 0.695–0.906) in the training phase and 0.736 (95% CI 0.631–0.842) in the validation phase. In addition, the combined use of miR-182 and miR-20a showed that miR-182 provided independent diagnostic information, as miR-20a dropped out from the logistic regression analysis model ($p=0.33$). Thus, our findings suggested that circulating miR-182 in plasma could be a potential biomarker with 78% sensitivity and 91% specificity for the early diagnosis of CRC patients.

Although Carpelan-Holmström et al reported that the sensitivity of serum CEA, CA19-9, and CA72-4 was 55%, 36%, and 25% and the specificity of serum CEA, CA19-9, and CA72-4 was 66%, 71%, and 66%, respectively, in diagnosing CRC, which were lower than those of miR-182. There were still several limitations in our study. First, the number

---

**Figure 5** MiR-182 and miR-20a were further validated to be potential biomarkers for early CRC diagnosis.

Notes: (A) Circulating miR-20a and miR-182 were upregulated in patients with stage I CRC compared to adenoma patients and healthy controls (B) ROC curve analysis revealed that the AUC was 0.891 (95% CI 0.821–0.961) for miR-182, 0.736 (95% CI 0.631–0.842) for miR-20a, and 0.831 (95% CI 0.746–0.914) for the 2-miRNA combination. The boxes represent the intervals between the 25th and 75th percentiles, and the whiskers mark the interval between the minimum and maximum. *$p<0.05$. ***$p<0.001$.

Abbreviations: HC, healthy control; CRC, colorectal cancer; ROC, receiver operating characteristic; AUC, area under the ROC curve; ns, not significant.
of subjects enrolled in our study was small. Moreover, the clinicopathological characteristics of patients were not fully detailed. Further, selection of the identified miRNAs came from the integrated analysis of 4 GEO data sets and TCGA database, which may disregard other miRNAs that function as novel promising diagnostic biomarkers. Finally, in the validation phase of our study, the number of adenoma patients at low risk (n=46) was much more than the number of those at high risk (n=4). Therefore, larger cohorts with more detailed clinicopathological features of patients need to be implemented in the future.

Conclusion
The expression of plasma miR-182 was significantly upregulated in patients with early-stage CRC compared with that in adenoma patients and healthy subjects, and miR-182 in plasma was demonstrated to be a promising diagnostic biomarker for early CRC.

Acknowledgments
This project was supported by grants from the National Nature Science Foundation of China (No. 81472027) to SKW; Key Project of Science and Technology Development of Nanjing Medicine (ZDX16001); Innovation team of Jiangsu Provincial Health-Strengthening Engineering by Science and Education (CXTDB2017008) to SKW; Jiangsu Youth Medical Talents Training Project to BH (number QNRC2016066) and YP (number QNRC2016074); and Nanjing Medical Science and Technique Development Foundation to BSH (number JQX13003).

Disclosure
The authors report no conflicts of interest in this work.

References