A plasma microRNA panel for early detection of colorectal cancer

Shuyang Wang1,*, Jianbin Xiang2,*, Zhaoyong Li1,3,*, Shaohua Lu4, Jie Hu5, Xue Gao1, Lei Yu2, Lei Wang1, Jiping Wang6, Ying Wu1, Zongyou Chen2, Hongguang Zhu1,7

1Department of Pathology, Shanghai Medical College, Fudan University, Shanghai, P.R. China; 2Department of General Surgery, Huashan Hospital, Fudan University, Shanghai, P.R. China; 3School of Life Sciences, University of Science and Technology of China, Hefei, P.R. China; 4Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai, P.R. China; 5Liver Cancer Institute, Zhongshan Hospital, Fudan University, P.R. China; 6Division of Surgical Oncology, Brigham and Women’s Hospital, Harvard Medical School, USA; 7Division of Surgical Pathology, Huashan Hospital, Fudan University, Shanghai, P.R. China

Short title: Plasma microRNA and colorectal cancer

Key words: Colorectal cancer, colorectal cancer genes, microRNA

Abbreviations: CRC, colorectal cancer; miRNA: microRNA

*S.W., J.X. and Z.L. contributed equally to this work.

Grant sponsors: Science and Technology Commission and Health Bureau of Shanghai Municipality (06DZ22904, 11R21410700 and C106629)

Correspondence to: Hongguang Zhu, M.D., Ph.D., Department of Pathology, Shanghai Medical College, Fudan University, 138 Yi Xue Yuan Road Shanghai 200032, China; tel.: +86 21 5423 7616; fax: +86 21 5423 7596; e-mail: hongguang_702@163.com or to Zongyou Chen, M.D., Ph.D., e-mail: zongyouc@hotmail.com or to Ying Wu, M.D., Ph.D., e-mail: yingwu.holland@yahoo.co.uk.

A brief description of the novelty and impact of this study

It is a challenge to identify patients with early stage colorectal cancer (CRC). We investigated plasma microRNA expression in 241 subjects (124 CRC patients and 117 healthy individuals) and identified a panel of miR-409-3p, miR-7 and miR-93 that yielded high diagnostic accuracy in discriminating CRC from healthy group (AUC 0.866 and 0.897 for training and validation dataset, respectively). This plasma microRNA panel has potential clinical value on early CRC detection in general population.
Abstract

Colonoscopy remains the standard screening method for detecting colorectal cancer (CRC) at an early stage. However, many people avoid having a colonoscopy because of the fear for its potential complications. Our study aimed to identify plasma microRNAs for preliminarily screening CRC in general population, so that some unnecessary colonoscopies can be avoided. We investigated plasma microRNA expression in 3 independent cohorts including the discovery (n=80), training (n=112) and validation (n=49) phases recruited at 2 medical centers. Microarrays were used for screening 723 microRNAs in 80 plasma samples to identify candidate microRNAs. Quantitative reverse-transcriptase PCR was performed on the 161 training and validation plasma samples to evaluate the candidate microRNAs discovered from microarrays. A logistic regression model was constructed based on the training cohort and then verified by using the validation dataset. Area under the receiver operating characteristic curve (AUC) was used to evaluate the diagnostic accuracy. We identified a panel of miR-409-3p, miR-7 and miR-93 that yielded high diagnostic accuracy in discriminating CRC from healthy group (AUC 0.866 and 0.897 for training and validation dataset, respectively). Moreover, the diagnostic performance of the microRNA panel persisted in non-metastasis CRC stages (Dukes’ A-B, AUC 0.809 and 0.892 for training and validation dataset, respectively) and in metastasis CRC stages (Dukes’ C-D, AUC 0.917 and 0.865 for training and validation dataset, respectively). In conclusion, our study reveals a plasma microRNA panel that has
potential clinical value in early CRC detection and would play a critical role on preliminarily screening CRC in general population.
Introduction

Colorectal cancer (CRC) is the fourth most common cancer and the fourth leading cause of cancer deaths in the world. An estimated 1.23 million new CRC cases and 608,700 CRC deaths occurred in 2008. Colorectal cancer is curable if diagnosed at an early stage of development. At this early stage, most patients have no phenotypic symptoms of the disease. Therefore, screening is of paramount importance in early CRC detection. Evidence indicates that screening for CRC can significantly reduce its incidence and mortality.

According to American Cancer Society Screening Guidelines for early CRC detection, both men and women beginning at age 50 should follow one of the following testing schedules: 1) Colonoscopy every 10 years; 2) Flexible sigmoidoscopy every 5 years; 3) Double-contrast barium enema (BE) every 5 years or 4) Computed tomographic (CT) colonography every 5 years. In addition, fecal occult blood tests (FOBT) and carcinoembryonic antigen (CEA) blood test are other tests often used in CRC screening. However, all the current screening methods have their own disadvantages.

Examination of the entire colon by colonoscopy remains the golden standard for screening CRC, but many people are hesitant to have a colonoscopy due to the embarrassment, preparation, fear of pain and discomfort that are associated with the procedure. Concerns on the potential complications and costs in an invasive procedure further decrease the success of screening colonoscopy. In contrast, the flexible sigmoidoscopy has a relatively low cost and is less invasive for early...
detection of neoplasia in the distal colon and rectum, however, the sigmoidoscopy is not able to view the entire colon.\textsuperscript{11} Double-contrast BE is radiological technique that can be used for CRC screening in general population. However, reports in the clinical literature commonly diminish its value due to the facts that it is not as accurate as colonoscopy and cannot perform biopsy.\textsuperscript{12} Also like double-contrast BE, CT colonography has been included in the screening guidelines, but it has not been widely used for colon cancer screening.\textsuperscript{13} On the other hand, the fecal occult blood test is a simple and cheap test, but it has less than ideal clinical specificity.\textsuperscript{14} CEA has been used for many years as a serum marker for CRC screening and prognosis.\textsuperscript{15,16}

However, CEA test has relatively low sensitivity and specificity which make it unsuitable for screening large asymptomatic patients. Currently, blood testing for CRC screening is more acceptable. However, no specific molecular markers have been identified so far that allow for a reliable diagnosis of CRC. Therefore, the discovery of new biomarkers would be of utmost clinical importance on screening CRC detection in routine healthcare for general population.

Many studies have demonstrated unique CRC microRNA profiles versus non-cancerous tissues.\textsuperscript{17-20} Additional studies have shown the stable existence of microRNAs in human serum or plasma\textsuperscript{21}, which has laid the foundation for studying their role in the diagnosis and prognosis of CRC. In fact, several studies have recently reported differential expression of plasma or serum microRNAs in CRC and healthy individuals.\textsuperscript{22-25} Nonetheless, those studies were limited by small number of screened microRNAs, small sample size or lack of independent validation.
Our study investigated plasma microRNA expression profiles to identify a panel of microRNAs for early CRC detection. In this paper, we present the successful discovery and validation of a plasma microRNA panel in discriminating CRC from healthy individuals. This microRNA panel has potential clinical value in early CRC detection and would play a critical role in screening CRC in general population.
Material and Methods

Clinical specimens

This study was approved by the local institutional review boards and written informed consent was obtained from all study participants. The study population consisted of 124 CRC patients and 117 healthy individuals recruited at Shanghai Huashan and Zhongshan Hospitals between October 2008 and March 2012. For selection of the samples in the CRC patients, we used routine histological classification according to the World Health Organization Classification of tumor system. All the lesions were diagnosed by 2 experienced pathologists and staged using Dukes classification. The blood samples from the cancer patients were obtained before the colorectal surgical resections. Patients who had received pre-operative radiotherapy or chemotherapy were excluded. For the selection of the samples in the healthy individuals, we used asymptomatic and apparently healthy volunteers without a previous history of cancer. All the volunteers were encountered during the medical check-up at the 2 medical centers and were confirmed that they were in the healthy condition without malignancy by the physical examinations. Individuals with system infection (lung, gastrointestinal tract, urinary tract) were excluded. Table 1 shows the eligibility criteria for the selection and exclusion of study subjects.

The clinical characteristics of the study subjects are presented in Table 2. There was no significant difference in the distribution of age and sex between the training and validation datasets for both healthy and CRC groups (all $p > 0.05$). The
healthy controls were significantly younger than the cancer patients \((p < 0.0001)\).

Overall, there were 48% (59/124) of the cancer patients with non-metastasis (Dukes’ A-B) stages, while 52% (65/124) of the cancer patients with metastasis (Dukes’ C-D) stages.

Study design

The study was designed in 3 different phases (Figure 1). The 241 blood samples were allocated in the chronological order. There was no sample overlapping among all the phases.

1. Discovery phase: 80 samples, each with 723 microRNAs, were screened with microarray platform. Mann-Whitney unpaired test was performed to discover differentially expressed microRNAs in the comparison of CRC and healthy group. From the differentially expressed microRNAs, 8 detectable microRNAs with \(p\) value \(< 0.05\) and fold expression change \(> 2.0\) were identified as the candidates for the further testing on quantitative reverse-transcriptase-PCR (quantitative RT-PCR). A detectable microRNA was defined as the microRNA had positive signals on microarrays in all plasma samples from any one of the 2 category subjects.

2. Training phase: The 8 candidate microRNAs discovered via microarray were tested using an independent cohort of 112 plasma samples with quantitative RT-PCR. Those 112 subjects were used as the training set to construct the diagnostic microRNA panel based on the logistic regression model for the differentiation of CRC patients and healthy individuals.
In addition, the diagnostic performance of the microRNA panel was further determined in distinguishing age-matched CRC patients (n = 41) and healthy controls (n = 27). The cancer patients were 15 men and 26 women with a mean age of 58 years (range 35 - 75 years), while the healthy volunteers were 17 men and 10 women with a mean age of 52 years (range 35 - 71 years). There was no significant difference in the distribution of age between the age-matched cancer patients and healthy volunteers (p = 0.0626).

3. Validation phase: The parameters of the logistic model from the training phase were applied to an independent cohort of 49 samples for validating the diagnostic performance of the selected microRNA panel.

**Plasma preparation and RNA isolation**

For plasma preparation, peripheral blood (4 ml) was drawn into EDTA tubes. Within 2 hours, the tubes were subjected to centrifuge at 820 g for 10 min. Then, 1 ml aliquots of the plasma was transferred to 1.5 ml tubes and centrifuged at 16,000g for 10 min to pellet any remaining cellular debris. Subsequently, the supernatant was transferred to fresh tubes and stored them at –80 °C.

For the plasma samples, total RNA was extracted by using mirVana PARIS miRNA Isolation kit according to the instructions from the manufacturer (Ambion, Austin, TX). The concentration was quantified by NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Waltham, MA). A RNA sample was discarded for further analysis if the total amount of RNA from 400 µl of plasma was less than 100 ng.
Microarray hybridization

Human microRNA microarrays from Agilent Technologies were used in plasma samples. The microarray contains probes for 723 human microRNAs from miRBase\textsuperscript{28}. Each slide is formatted with 8 identical arrays. Total RNA (100 ng) derived from plasma samples were labelled with Cy3. Microarray slides were scanned by XDR Scan (PMT100, PMT5). The labelling and hybridization were performed according to the protocols in the Agilent miRNA microarray system. The microarray image information was converted into spot intensity values using Feature Extraction Software Rev. 9.5.3 (Agilent Technologies, Santa Clara, CA). The signal after background subtraction was exported directly into the GeneSpring GX10 software (Agilent Technologies, Santa Clara, CA). The raw signals obtained for single-color CY3 hybridization were normalized by a stable endogenous control miR-1228\textsuperscript{29}. Then a log transform with base 2 was performed. A sample that showed intra-array coefficients of variation (CV) across replicated spots on an array above 15% or positive signals less than 5% was considered to be unreliable and excluded from further analysis.

Quantitative RT-PCR

In the training and validation phases, quantitative RT-PCR using Taqman MicroRNA Assays (Applied Biosystems, Foster City, CA) was performed on plasma samples according to the manufacturer’s instructions. Reverse transcription (RT) was performed with Taqman microRNA RT kits according to the instruction from Applied
Biosystem. The cDNA was synthesized from total RNA (100 ng) using microRNA-specific primers in a 40 µl reaction volume. The RT reaction was performed using the following thermal cycling program: 30 min at 16°C, 30 min at 42°C, 5 min at 85°C and then held at 4°C. The RT product was diluted 10-fold, and 4 µl of the product were used in a total reaction volume of 10 µl for relative quantification by real-time PCR with an ABI 7900HT fast system (Applied Biosystems, Foster City, CA). The thermal cycling program used for the quantification was as follows: 96°C for 5 min and followed by 50 cycles of 95°C for 15 s and 60°C for 1 min. The cycle threshold (Ct) was calculated using the second derivative method in the ABI software.

Following the guidelines on the minimum information for publication of quantitative real-time PCR experiments (MIQE), the Taqman MicroRNA assays were first performed on 112 plasma samples for 8 candidate microRNAs (miR-25, miR-331-3p, miR-345, miR-409-3p, miR-671-3p, miR-7, miR-92a and miR-93). Then the assays were applied on 49 plasma samples for 3 significant microRNAs (miR-409-3p, miR-7 and miR-93). For normalization, the expression level of miR-1228 was used as an endogenous control. All assays were carried out in triplicate. A RNA sample was discarded for further analysis if the endogenous control miR-1228 or target microRNA showed Ct values above 35 cycles.

Analysis of microarray data

Mann-Whitney unpaired test with Benjamini-Hochberg correction (p value < 0.001, fold expression change > 2.0) was performed to determine differentially
expressed microRNAs between the CRC patients and healthy individuals. Hierarchical clustering was performed with Pearson correlation using 52 differentially expressed microRNAs. The mean signal from the biological replicate samples was used for the clustering. GeneSpring GX10 software (Agilent Technologies, Santa Clara, CA) was used to perform the analysis of microarray data.

Statistical analysis

For quantitative RT-PCR, delta Ct (ΔCt) value was used to represent individual microRNA expression level. The delta Ct of the microRNA was calculated by subtracting the Ct value of the endogenous control miR-1228. Mann-Whitney unpaired test was used to determine differentially expressed microRNAs between the CRC patients and healthy individuals. Logistic regression model\(^{32}\) was used to select diagnostic microRNA markers based on the training dataset. The predicted probability of being diagnosed with CRC was used as a surrogate marker to construct receiver operating characteristic (ROC) curve. Area under the ROC curve (AUC)\(^{33}\) was used as an accuracy index for evaluating the diagnostic performance of the selected microRNA panel. MedCalc (10.4.7.0) software (Mariakerke, Belgium) was used to perform the Mann-Whitney unpaired test, ROC curve and regression analysis. All the \(p\) values were two sided.
Results

MicroRNA expression profiles

In the discovery phase, we measured the expression of 723 human microRNAs in 80 plasma samples from 47 CRC patients and 33 healthy individuals. Figure 2 shows hierarchical clustering of plasma microRNA expression profiles on the microarrays. The clustering of 52 microRNAs placed 34/47 CRC patients in one group and 21/33 healthy individuals in another group.

Selection of candidate microRNAs

Eight differentially expressed microRNAs were identified by microarray platform. Of the 8 candidates, 4 microRNAs including miR-7 (p < 0.0001 with fold change 0.2), miR-25 (p = 0.0005 with fold change 0.4), miR-92a (p = 0.0009 with fold change 0.5) and miR-93 (p = 0.0002 with fold change 0.4) had significantly lower expression levels in the CRC group than in the healthy group. In contrast, other 4 microRNAs including miR-409-3p (p = 0.0024 with fold change 4.5), miR-345 (p = 0.0055 with fold change 4.4), miR-671-3p (p = 0.0057 with fold change 4.5) and miR-331-3p (p = 0.0176 with fold change 3.3) had significantly higher expression levels in the CRC group than in the healthy group. The diagnostic accuracy, measured by AUC, ranged from 0.658 to 0.759 (Table 2: microarrays).

Evaluation of candidate microRNAs
The expression profiles of the 8 candidate microRNAs were evaluated with quantitative RT-PCR on the training dataset, an independent cohort of 112 plasma samples including 55 CRC patients and 57 healthy individuals. Six of the 8 candidate microRNAs had significant differential expression levels between the CRC and healthy group (Table 2: quantitative RT-PCR). Low expression levels of miR-93 ($p < 0.0001$ with fold change 0.3), miR-345 ($p < 0.0001$ with fold change 0.5), miR-331-3p ($p < 0.0001$ with fold change 0.3), miR-25 ($p < 0.0001$ with fold change 0.4) and miR-7 ($p = 0.02$ with fold change 0.7) were observed in CRC patients as compared with those in the healthy group. In contrast, high expression level of miR-409-3p ($p = 0.04$ with fold change 1.8) was observed in CRC patients as compared with those in the healthy group. The diagnostic accuracy (AUC) ranged from 0.604 to 0.828. There were similar expression levels in miR-671-3p ($p = 0.52$ with fold change 1.2) and miR-92a ($p = 0.40$ with fold change 0.8) between the CRC and healthy group.

Of 8 candidate microRNAs, the expression profiles acquired on microarrays could be confirmed for 4 (50%) cases by quantitative PCR (Table 2 and 3). Three microRNAs (miR-25, miR-7 and miR-93) were confirmed as down-regulated, while one microRNA (miR-409-3p) could be verified as up-regulated in CRC patients as compared with the healthy group. In contrast, 2 microRNAs (miR-331-3p and miR-345) demonstrated the opposite regulations, while other 2 microRNAs (miR-671-3p and miR-92a) did not show significant differential expression.

Establishment of the predictive microRNA panel
A logistic regression model was applied to estimate the risk of being diagnosed with CRC on the training dataset (112 plasma samples) with 4 significant microRNAs (miR-25, miR-409-3p, miR-7 and miR-93). Three of the 4 significant microRNAs turned out to be significant predictors (see the regression model in Table 2). The predicted probability of being diagnosed with CRC from the logit model based on the panel of 3 significant predictors, logit(p) = 7.7084+0.5297*(miR-409-3p)+1.4463*(miR-7)-1.1592*(miR-93), was used to construct the ROC curve. The diagnostic performance for the established microRNA panel was evaluated by using ROC analysis. In all CRC stages (Dukes’ A-D), the AUC for the microRNA panel was 0.866 with 91% sensitivity and 88% specificity (Fig. 3A). In the non-metastasis (Dukes’ A-B) stages, the AUC for the microRNA panel was 0.809 with 85% sensitivity and 88% specificity (Fig. 3B). While in the metastasis (Dukes’ C-D) stages, the AUC was 0.917 with 96% sensitivity and 88% specificity (Fig. 3C).

The diagnostic performance of the microRNA panel in distinguishing age-matched CRC patients and healthy controls was further evaluated (Fig. 3D-3F). The corresponding AUC for the patients with Dukes’ A-D, A-B and C-D was 0.894, 0.850 and 0.937, respectively.

Validation of the predictive microRNA panel

We next validated the panel of 3 significant microRNAs: miR-409-3p, miR-7 and miR-93 (described above) on the independent validation dataset (49 plasma samples). The parameters estimated from the training dataset were used to predict the
probability of being diagnosed with CRC for the validation dataset. Similarly, the predicted probability was used to construct the ROC curve. In all CRC stages (Dukes’ A-D), the AUC of the microRNA panel was 0.897 with sensitivity 82% and specificity 89% (Fig. 4A). In the non-metastasis (Dukes’ A-B) stages, the AUC for the microRNA panel was 0.892 with 82% sensitivity and 89% specificity (Fig. 4B). While in the metastasis (Dukes’ C-D) stages, the AUC was 0.862 with 82% sensitivity and 89% specificity (Fig. 4C).
DISCUSSION

Colorectal cancer is common and many of these cancers are preventable, because most CRCs arise from precursor adenomatous polyps. The major aim of the screening is to detect sporadic CRC cases in people over the age of 50 years. Screening colonoscopy, which is so far the most reliable method, has the advantage of visualizing the entire colon, but the procedure is invasive, involves substantial discomfort and has a risk of complications such as bowel perforation. In addition, the screening interval of 10 years for colonoscopy is too long. In such a 10-year screening interval, the detection rate of early CRC by colonoscopy is only 18% - 35%. It had been reported that the range of growth rates of CRC is 25% - 75% compared with the baseline. Thus, colonoscopy is not suitable for screening CRC in general population.

Flexible sigmoidoscopy in asymptomatic persons appears to be a highly sensitive method for early detection of both cancer and adenomatous polyps in the distal colon and rectum. However, recent studies demonstrated that the sigmoidoscopic screening failed to detect a substantial proportion of asymptomatic neoplasia due to its inability in delivering ~50% protection against CRC in the proximal colon. Similarly, the sigmoidoscopy is not suitable for screening CRC in general population.

CEA, which performs screening on a serum level, has the remarkable advantage of convenience and simplicity in which it can be performed on a patient annually. Moreover, its lack of sensitivity in detecting early CRC makes CEA
determination especially poor for screening. When using 2.5 mg/ml of the serum CEA as the upper limits of normal, the sensitivity for Dukes’ A and B lesions is 36%, compared with 74% for Dukes’ C and 83% for Dukes’ D disease. In contrast, our study reveals that plasma miR-409-3p, miR-7 and miR-93 were potential circulating markers for early CRC detection. The microRNA panel with the 3 significant microRNAs from the logistic regression model demonstrate high diagnostic accuracy (AUC 0.897, sensitivity 82% and specificity 89%) in discriminating CRC from healthy group. Moreover, the high diagnostic accuracy of the microRNA panel persisted in the non-metastasis CRC stages (AUC 0.892, sensitivity 82% and specificity 89%) and in the metastasis CRC stages (AUC 0.865, sensitivity 82% and specificity 89%). This makes the microRNA panel an optimal tool for screening CRC in general population.

At the serum or plasma level, more than 100 circulating microRNAs had been identified in the blood of healthy individuals and their profiles significantly differed from that of CRC patients. Compared to those studies of circulating microRNAs in diagnosing CRC, our study is unique for the following reasons: First, we screened a large number of plasma microRNAs and samples via microarrays, which enabled us to have better chance to identify potential diagnostic markers. Furthermore, we identified a panel of 3 significant microRNAs, which would have better detection reliability and/or accuracy than a single molecular marker. In addition, we validated the plasma microRNA panel in 2 independent cohorts of plasma samples recruited at the 2 medical centers. Our microRNA panel demonstrates the satisfactory diagnostic
performance in detection of non-metastasis CRC stages. Using this microRNA panel for annual screening CRC in general population would improve the detection of early CRC and avoid unnecessary colonoscopy examinations.

The accurate quantification of plasma microRNA content is reliant upon the selection of stable endogenous control for normalizing microarray and quantitative RT-PCR data. Currently, there is no standard endogenous control for the circulating microRNA studies. Chen et al.\textsuperscript{22} discovered 69 microRNAs in discriminating CRC from healthy group using Solexa sequencing analysis. The expression levels of target microRNAs were directly normalized to total RNA. Ng et al.\textsuperscript{23} found the significantly elevated expression level of miR-92 in the plasma of CRC patients. The expression levels of target microRNAs were normalized to RNU6B on quantitative RT-PCR. In addition, Huang et al.\textsuperscript{24} found that miR-29a and miR-92a were significantly elevated in the CRC plasma samples when compared with normal controls. Their study selected miR-16 as a normalization control in the data analysis of quantitative RT-PCR. Our study showed the lower expression of miR-92a in the comparison of CRC versus healthy group in the microarray analysis using miR-1228 as a stable endogenous circulating control. However, such differential expression could not be confirmed in an independent cohort of 112 plasma samples using quantitative RT-PCR. Nevertheless, another study found that miR-92a dramatically decreased in the plasma of acute leukemia patients.\textsuperscript{25} Moreover, none of the 3 significant microRNAs in our predictive panel has been reported in the serum or plasma of CRC patients. The discrepancy on the candidate microRNAs across the different studies was thought to
be the difference in the endogenous controls for normalization. Furthermore, the heterogeneity of CRC and difference in technology platforms and patient numbers might contribute to the discordance. Therefore, identification and characterization of appropriate endogenous control microRNA(s) for the data normalization is a critical process to obtain a biologically meaningful microRNA expression comparison.

We compared the expression profiles obtained on microarrays with those on quantitative RT-PCR. Only 4 (50%) of the 8 candidate microRNAs had similar expression profiles on the 2 different platforms. In the discovery and validation phases, the total RNA source and labelling procedures used in the experiments were not identical. The discrepancy might be due to the differences in the technology platforms and plasma samples. Further large validations should be performed to ensure the clinical applicability of our biomarker panel. Furthermore, the stable control miR-1228 we utilized needs to be validated in more studies. Moreover, the patients’ follow-up time in this study was only 1-4 years, limiting our current ability for prognostic analysis.

In our study, many individuals in the healthy group are below the age range that is usually considered optimal for CRC screening (50-60 years). Although it has been reported that the expression levels of circulating microRNAs in the healthy human subjects are reproducibly consistent, this might still introduce a bias in the study outcomes. Therefore, we determined the diagnostic performance of the microRNA panel in distinguishing age-matched CRC patients and healthy controls. Our study indicated that the diagnostic performance of the microRNA panel was
independent of subject age, which made it an optimal diagnostic tool in CRC screening.

The association at the tissue level between CRC and 2 (miR-7 and miR-93) of the 3 significant microRNAs in our predictive panel had been previously reported. In our previous studies, we showed that miR-7 had increased in the expression level with the adenoma to carcinoma transition.\(^\text{46}\) It strongly expressed in the epithelial cells of colorectal tissues.\(^\text{47}\) Further functional study is needed to confirm the role of miR-7 in CRC. Recently, it was reported that miR-93 had substantially different expression levels in early and non-early relapse CRC patients.\(^\text{48}\) It could inhibit tumour growth and early relapse of human colorectal cancer by affecting genes involved in the cell cycle.\(^\text{48}\) MicroRNA-409-3p was reported to be significantly down-regulated in gastric cancer cell lines and tissues.\(^\text{49,50}\) It suppressed tumour cell invasion and metastasis by directly targeting radixin\(^\text{49}\) and suppressed cell proliferation and induced cell apoptosis by targeting PHF10.\(^\text{50}\) Nevertheless, our study is the first to report the importance of miR-409-3p expression profile in association with CRC.

In conclusion, this study reveals a plasma microRNA panel that differentiates CRC with a high degree of accuracy from healthy group, especially for patients with non-metastasis CRC stages. Our study demonstrates that this plasma microRNA panel has potential clinical value in early CRC detection and would play a critical role on screening CRC in general population.

Acknowledgments
We thank Xianxin Meng, MS and Zhidong Zhu, MS at Shanghai Biotechnology Corporation for microarray hybridization and statistical assistance.
References


8 Seeff LC, Nadel MR, Klabunde CN, Thompson T, Shapiro JA, Vernon SW, Coates RJ. Patterns and predictors of colorectal cancer test use in the adult U.S.


16 Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: is it...


Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 1947;18:50-60.


Table 1. Eligibility criteria for the selection and exclusion of study subjects

<table>
<thead>
<tr>
<th>General inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age ≥ 18 years and ≤ 80 years</td>
</tr>
<tr>
<td>2. Not currently residing in an institution, such as a prison, nursing home, or shelter</td>
</tr>
<tr>
<td>3. Not severely ill in the intensive care unit</td>
</tr>
<tr>
<td>4. With the capability to give informed consent</td>
</tr>
<tr>
<td>5. Encountered between August 2008 and March 2012</td>
</tr>
</tbody>
</table>

**Healthy individuals (control group)**

1. Had the medical check-up in Shanghai Huashan or Zhongshan Hospital
2. Asymptomatic and apparently healthy without a previous history of cancer
3. Confirmed healthy condition without malignancy in the physical examinations
   - No lung tumor examined by chest X-ray and/or computed tomography
   - No liver, kidney, pancreatic, pelvis, bladder, thyroid tumor scanned by B-mode ultrasound
   - No breast tumor examined by B-mode ultrasound
   - No indication for cancer in the tumor marker blood tests.
4. No system infection (lung, gastrointestinal tract, urinary tract)

**Colorectal cancer patients (CRC group)**

1. Had colonoscopy biopsy and colorectal surgical resections
2. Diagnosed by two experienced pathologists
3. No pre-operative chemotherapy and radiotherapy
Table 2. Characteristics of patients and tumors in the study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Healthy Count (%)</th>
<th>Colorectal cancer Count (%)</th>
<th>Healthy vs. CRC (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Microarrays</td>
<td>Quantitative RT-PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discovery phase</td>
<td>Training phase</td>
<td>Validation phase</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13 (39)</td>
<td>34 (60)</td>
<td>18 (67)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (61)</td>
<td>23 (40)</td>
<td>9 (33)</td>
</tr>
<tr>
<td>Age— year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean± SD</td>
<td>43 ± 14</td>
<td>44 ± 12</td>
<td>42 ± 8</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>18 (38)</td>
<td>23 (42)</td>
<td>13 (59)</td>
</tr>
<tr>
<td>Female</td>
<td>29 (62)</td>
<td>32 (58)</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Age— yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean± SD</td>
<td>64 ± 13</td>
<td>64 ± 13</td>
<td>64 ± 14</td>
</tr>
<tr>
<td>Dukes' stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>10 (21)</td>
<td>7 (13)</td>
<td>5 (23)</td>
</tr>
<tr>
<td>B</td>
<td>12 (26)</td>
<td>19 (35)</td>
<td>6 (27)</td>
</tr>
<tr>
<td>C</td>
<td>18 (38)</td>
<td>19 (35)</td>
<td>10 (45)</td>
</tr>
<tr>
<td>D</td>
<td>7 (15)</td>
<td>10 (18)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Healthy vs. CRC (P value)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.7341</td>
<td>0.8155</td>
<td>0.9647</td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Chi-square test was used to calculate p values for sex, while unpaired unequal variance t-test was used to calculate p values for age.

<sup>a</sup>Significant difference between the training and validation datasets.

<sup>b</sup>Significant difference between the CRC and healthy datasets.
<table>
<thead>
<tr>
<th>Microarrays</th>
<th>p value</th>
<th>Fold change</th>
<th>AUC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsasmiRs25</td>
<td>4.5E-04</td>
<td>0.4</td>
<td>0.727</td>
<td>0.615 - 0.820</td>
</tr>
<tr>
<td>hsasmiRs331-3p</td>
<td>1.8E-02</td>
<td>3.3</td>
<td>0.658</td>
<td>0.544 - 0.761</td>
</tr>
<tr>
<td>hsasmiR-345</td>
<td>5.5E-03</td>
<td>4.4</td>
<td>0.680</td>
<td>0.566 - 0.780</td>
</tr>
<tr>
<td>hsasmiR-409-3p</td>
<td>2.4E-03</td>
<td>4.5</td>
<td>0.697</td>
<td>0.584 - 0.795</td>
</tr>
<tr>
<td>hsasmiR-671-3p</td>
<td>5.7E-03</td>
<td>4.5</td>
<td>0.679</td>
<td>0.565 - 0.779</td>
</tr>
<tr>
<td>hsasmiR-7</td>
<td>&lt; 0.0001</td>
<td>0.2</td>
<td>0.759</td>
<td>0.650 - 0.848</td>
</tr>
<tr>
<td>hsasmiR-92a</td>
<td>9.3E-04</td>
<td>0.5</td>
<td>0.712</td>
<td>0.600 - 0.808</td>
</tr>
<tr>
<td>hsasmiR-93</td>
<td>2.2E-04</td>
<td>0.4</td>
<td>0.739</td>
<td>0.629 - 0.831</td>
</tr>
<tr>
<td>Quantitative RT-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsasmiR-25</td>
<td>&lt; 0.0001</td>
<td>0.4</td>
<td>0.773</td>
<td>0.684 - 0.847</td>
</tr>
<tr>
<td>hsasmiR-331-3p</td>
<td>&lt; 0.0001</td>
<td>0.3</td>
<td>0.723</td>
<td>0.630 - 0.803</td>
</tr>
<tr>
<td>hsasmiR-345</td>
<td>&lt; 0.0001</td>
<td>0.3</td>
<td>0.816</td>
<td>0.732 - 0.883</td>
</tr>
<tr>
<td>hsasmiR-409-3p</td>
<td>4.0E-02</td>
<td>1.8</td>
<td>0.604</td>
<td>0.507 - 0.695</td>
</tr>
<tr>
<td>hsasmiR-671-3p</td>
<td>5.2E-01</td>
<td>1.2</td>
<td>0.526</td>
<td>0.430 - 0.622</td>
</tr>
<tr>
<td>hsasmiR-7</td>
<td>2.0E-02</td>
<td>0.7</td>
<td>0.593</td>
<td>0.496 - 0.685</td>
</tr>
<tr>
<td>hsasmiR-92a</td>
<td>4.0E-01</td>
<td>0.8</td>
<td>0.620</td>
<td>0.521 - 0.713</td>
</tr>
<tr>
<td>hsasmiR-93</td>
<td>&lt; 0.0001</td>
<td>0.3</td>
<td>0.828</td>
<td>0.745 - 0.893</td>
</tr>
</tbody>
</table>

Regression model (miR-409-3p+7+93) 0.866 0.788 - 0.923

Logit(p=CRC)=7.7084+0.5297*(miR-409-3p)+1.4463*(miR-7)-1.1592*(miR-93)
Figure legends

Figure 1  Study design. The microRNA profiles of 241 plasma samples from 124 CRC patients and 117 healthy individuals recruited at 2 medical centers were used to generate outcomes in 3 different phases. The candidate microRNAs discovered on 80 plasma samples using microarrays were validated in 2 independent cohorts of 161 plasma samples using quantitative RT-PCR. A logistic regression model was constructed based on a training cohort (n = 112) and then validated using an independent cohort (n = 49). CRC: colorectal cancer.

Figure 2. Hierarchical clustering of plasma microRNA expression profiles on the microarrays. Hierarchical clustering was performed with 52 differentially expressed microRNAs in the comparison of CRC patients (n = 47) and healthy individuals (n = 33). The clustering of 52 microRNAs placed 34/47 CRC patients in one group and 21/33 healthy individuals in another group. Colored bars indicate the range of normalized log2-based signals. CRC: colorectal cancer

Figure 3. ROC curve analysis of the logit model with miR-409-3p, miR-7 and miR-93 on the training dataset. A) AUC estimation of the logit(p) value for the plasma microRNA panel in discriminating the CRC patients (Dukes’ A-D) from the healthy group. B) AUC estimation of the logit(p) value for the microRNA panel in discriminating the non-metastasis CRC patients (Dukes’ A-B) from the healthy group. C) AUC estimation of the logit(p) value for the microRNA panel in discriminating the
metastasis CRC patients (Dukes’ C-D) from the healthy group. D) AUC estimation of
the logit(p) value for the plasma microRNA panel in discriminating the age-matched
CRC patients (Dukes’ A-D) from the age-matched healthy group. E) AUC estimation
of the logit(p) value for the microRNA panel in discriminating the age-matched non-
metastasis CRC patients (Dukes’ A-B) from the age-matched healthy group. F) AUC
estimation of the logit(p) value for the microRNA panel in discriminating the age-
matched metastasis CRC patients (Dukes’ C-D) from the age-matched healthy group.
CRC: colorectal cancer. Logit(p=CRC)=7.7084+0.5297*(miR-409-3p)+1.4463*(miR-7)-1.1592*(miR-93).

Figure 4. ROC curve analysis of the logit model with miR-409-3p, miR-7 and
miR-93 on the validation dataset. A) AUC estimation of the logit(p) value for the
plasma microRNA panel in discriminating the CRC patients (Dukes’ A-D) from the
healthy group. B) AUC estimation of the logit(p) value for the microRNA panel in
discriminating the non-metastasis CRC patients (Dukes’ A-B) from the healthy group.
C) AUC estimation of the logit(p) value for the microRNA panel in discriminating the
metastasis CRC patients (Dukes’ C-D) from the healthy group. CRC: colorectal
cancer. Logit(p=CRC)=7.7084+0.5297*(miR-409-3p)+1.4463*(miR-7)-1.1592*(miR-93).
The microRNA profiles of 241 plasma samples from 124 CRC patients and 117 healthy individuals recruited at 2 medical centers were used to generate outcomes in 3 different phases. The candidate microRNAs discovered on 80 plasma samples using microarrays were validated in 2 independent cohorts of 161 plasma samples using quantitative RT-PCR. A logistic regression model was constructed based on a training cohort (n = 112) and then validated using an independent cohort (n = 49). CRC: colorectal cancer.
Hierarchical clustering of plasma microRNA expression profiles on the microarrays. Hierarchical clustering was performed with 52 differentially expressed microRNAs in the comparison of CRC patients (n = 47) and healthy individuals (n = 33). The clustering of 52 microRNAs placed 34/47 CRC patients in one group and 21/33 healthy individuals in another group. Colored bars indicate the range of normalized log2-based signals. CRC: colorectal cancer
ROC curve analysis of the logit model with miR-409-3p, miR-7 and miR-93 on the training dataset. A) AUC estimation of the logit(p) value for the plasma microRNA panel in discriminating the CRC patients (Dukes’ A-D) from the healthy group. B) AUC estimation of the logit(p) value for the microRNA panel in discriminating the non-metastasis CRC patients (Dukes’ A-B) from the healthy group. C) AUC estimation of the logit(p) value for the microRNA panel in discriminating the non-metastasis CRC patients (Dukes’ A-B) from the healthy group. D) AUC estimation of the logit(p) value for the plasma microRNA panel in discriminating the age-matched CRC patients (Dukes’ A-D) from the age-matched healthy group. E) AUC estimation of the logit(p) value for the microRNA panel in discriminating the age-matched non-metastasis CRC patients (Dukes’ A-B) from the age-matched healthy group. F) AUC estimation of the logit(p) value for the microRNA panel in discriminating the age-matched metastasis CRC patients (Dukes’ C-D) from the age-matched healthy group. G) AUC estimation of the logit(p) value for the microRNA panel in discriminating the age-matched metastasis CRC patients (Dukes’ C-D) from the age-matched healthy group.

CRC: colorectal cancer. Logit(p=CRC)=7.7084+0.5297\times(miR-409-3p)+1.4463\times(miR-7)-1.1592\times(miR-93)

John Wiley & Sons, Inc.
ROC curve analysis of the logit model with miR-409-3p, miR-7 and miR-93 on the validation dataset. A) AUC estimation of the logit(p) value for the plasma microRNA panel in discriminating the CRC patients (Dukes’ A-D) from the healthy group. B) AUC estimation of the logit(p) value for the microRNA panel in discriminating the non-metastasis CRC patients (Dukes’ A-B) from the healthy group. C) AUC estimation of the logit(p) value for the microRNA panel in discriminating the metastasis CRC patients (Dukes’ C-D) from the healthy group. CRC: colorectal cancer. Logit(p=CRC)=7.7084+0.5297*(miR-409-3p)+1.4463*(miR-7)-1.1592*(miR-93).