Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls

Zhibin Hu1,2,3, Jingle Dong1,2, Li-E Wang1, Hongxia Ma2, Jibin Liu5, Yang Zhao2, Jinhai Tang2, Xi Chen3, Juncheng Dai3, Qingyi Wei4, Chenyu Zhang2 and Hongbing Shen1,2,3,*

1State key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 210029, China, 2Department of Epidemiology and Biostatistics, Ministry of Education Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing 210029, China, 3Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Nanjing Medical University, Nanjing 210029, China, 4Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA, 5Nantong Cancer Hospital, Nantong 226361, China, 6Department of General Surgery, Jiangsu Cancer Hospital, Nanjing 210009, China and 7State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210093, China

*To whom correspondence should be addressed. Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, 140 Hanzhong Road, Nanjing, Jiangsu 210029, China.
Tel/Fax: +0086 25 8686 2756; Email: hbshen@njmu.edu.cn

It has been demonstrated that there are abundant stable microRNAs (miRNAs) in plasma/serum, which can be detected and are potentially disease specific. However, the lack of suitable endogenous controls for serum miRNA detection is the restriction for the widely usage of this kind of biomarkers and for the between-laboratory comparison of the findings. We first systematically screened for endogenous control miRNAs (ECMs) by testing 10 pooling samples (using both Solexa sequencing and TaqMan low density array) and 50 individual samples (using quantitative reverse transcription–PCR) of different cancer traits and healthy controls. Then we assessed serum miRNAs used as potential biomarkers for breast cancer risk prediction based on a two-stage case–control analysis, including 48 breast cancer patients and 48 controls for the discovery stage and 76 breast cancer patients and 76 controls for validation. We identified two candidate ECMs (miRNA-191 and miRNA-484). Normalized by the two ECMs, we found four miRNAs (miR-16, miR-25, miR-222 and miR-324-3p) that were consistently differentially expressed between breast cancer cases and controls. The area under the receiver operating characteristic curve is 0.954 for the four-miRNA signature in the discovery stage (sensitivity = 0.917 and specificity = 0.896) and 0.928 in the validation stage (sensitivity = 0.921 and specificity = 0.934). In conclusion, the four-miRNA signature from serum may serve as a non-invasive prediction biomarker for breast cancer. Furthermore, we proposed the combination of miRNA-484 and miRNA-191 as endogenous control for serum miRNA detection, at least for most common cancers.

Introduction

MicroRNAs (miRNAs) are an abundant class of small non-protein-coding RNAs that function as negative gene regulator involving in crucial biological processes such as development, differentiation, apoptosis and proliferation (1,2). In mammals, mature miRNAs exert their regulatory effects by imperfect pairing within the 3′-untranslated region and coding region of genes (3–5). Bioinformatics data suggest that a single miRNA could bind to mRNA targets of as many as 200 genes with diverse functions (6,7). Therefore, the miRNA profiling may provide a more accurate and informative classification of cancer subtypes and other phenotypes than the use of the expression profiles of protein-coding genes (8).

In 2008, both we and Mitchell et al. discovered with excitement that human serum/plasma contains a large amount of stable miRNAs and the expression patterns of serum/plasma miRNAs have a great potential as the fingerprint for many diseases, including cancers (9,10). Later on, we used a two-stage study to investigate the role of serum miRNAs in predicting the prognosis of non-small-cell lung cancer by Solexa sequencing for pooled samples followed by individual-based quantitative reverse transcription–PCR (qRT–PCR) assays and found that the expression levels of four miRNAs (miR-486, miR-30d, miR-1 and miR-499) were significantly associated with non-small-cell lung cancer survival (11). Many additional miRNAs have been shown to be differentially expressed in various cancer phenotypes and related to cancer classification, diagnosis and prognosis (12–17). Current technologies for measurement of plasma/serum miRNAs make it possible to characterize the large-scale miRNA expression patterns. However, the chosen of endogenous controls for detection of serum miRNAs is a major issue to be solved. In contrast to tissue or cellular miRNAs, of which the expression levels were commonly normalized against U6, no specific serum miRNA was creditably endorsed as a suitable endogenous control to date. Two approaches could be used to reduce the experimental variations, one is to process the same ‘suitable’ volume of serum and the other is to use spiked-in synthetic non-human miRNA, like Caenorhabditis elegans miRNA-39 at the beginning of RNA extraction. However, the endogenous control miRNA (ECM) is still a chief restriction for clinical translation of serum miRNA biomarkers and for the between-laboratory comparison of the findings.

Breast cancer is the leading malignancy with 400 000 deaths annually among women worldwide, accounting for 23% of all cancers in women (18). In China, the incidence rate of breast cancer has increased dramatically in recent years, with a sharp rise of 38.5% between 2000 and 2005 (18,19). Early detection of breast cancer is vital to reduce the mortality of this disease (20,21). However, even with the most acceptable methods, such as mammography, ultrasonography and magnetic resonance imaging, for breast cancer detection, concerns remain for rates of misdiagnosis, missed diagnosis and the overdiagnosis (21). Several reported studies have investigated serum/plasma miRNA levels in breast cancer. Lodes et al. (22) used a pan-human microarray platform to evaluate serum miRNA expression patterns of five types of cancer, including breast cancer, and found that miRNA expression patterns could discriminate normal and cancer patients. Zhao et al. (23) performed a pilot study to compare the levels of plasma miRNAs from early-stage breast cancer patients and healthy controls of Caucasian American and African American and found that 31 miRNAs were differentially expressed in Caucasian American study and 18 in African American study. A recent study also reported that serum miR-21 had diagnostic and prognostic potential in breast cancer (24). Although these studies provided promising evidences that serum/plasma miRNA patterns would be a powerful tool for breast cancer classification, most of the studies used miR-16 as the reference miRNA without systematically screening for ECM. Furthermore, none of the studies validated their results in an independent sample set.

In this study, we firstly made an effort to seek the suitable ECM in serum by using both Solexa sequencing and TaqMan low density array (TLDA) chips followed by individual qRT–PCR assays and then tested the hypothesis that serum miRNAs could improve breast cancer risk estimation with a two-stage case–control analysis.
Materials and methods

Study design and study population

This study was approved by the Institutional Review Board of Nanjing Medical University. Only incident cancer patients before treatments of surgical operation or chemoradiotherapy were included. A written informed consent was obtained from each participant or their representatives. Each subject was face-to-face interviewed by trained interviewers using a pretested standard questionnaire to obtain information on demographic data and environmental exposure history.

We divided our study into two phases (Figure 1). The first phase of this study was designed to screen for ECMs in serum by comparing different cancer types. We used 10 pooled serum samples, including 2 lung cancer pooling samples (30 patients with longer survival) and 30 patients with shorter survival, respectively (11), 2 breast cancer pooling samples (24 patients each), 1 cervical cancer pooling sample (30 patients), 2 gastric cancer pooling samples (20 patients each), 1 hepatocellular carcinoma (HCC) pooling sample (30 patients) (15) and 2 healthy control samples (48 male samples and 48 female samples, respectively). These cancer patients were recruited from the Cancer Hospital of Jiangsu Province (Nanjing, China), the First Affiliated Hospital of Nanjing Medical University (Nanjing, China), and the Nantong Cancer Hospital (Nantong, Jiangsu, China) from July 2003 to 2009. The controls were recruited from the First Affiliated Hospital of Nanjing Medical University. All of the 10 pooling samples were detected by both Solexa sequencing and TLDA chip assays. Only the miRNAs that showed stability across the 10 samples from both assays were selected as candidate ECMs. Then, individual qRT–PCR was conducted for candidate ECMs in additional 50 serum samples, including 5 esophageal cancer, 5 colon cancer, 5 rectal cancer, 5 pancreatic cancer, 5 oral cancer, 5 gastric cancer, 5 lung cancer, 5 breast cancer, 5 HCC and 5 healthy controls. These 50 samples were recruited from the First Affiliated Hospital of Nanjing Medical University and the Cancer Hospital of Jiangsu Province from February 2010 to June 2010.

The second phase of this study was designed to discover serum miRNAs that were associated with breast cancer development, based on the above identified ECMs. The discovery stage samples were 48 incident histopathologically diagnosed breast cancer patients and 48 female controls that were used for ECMs identification in Solexa sequencing as well as TLDA chips. Then, individual qRT–PCR was conducted to refine the number of screened serum miRNAs. The lower the stability value is, the higher the miRNAs satisfied the following criteria were chosen for further confirmation by individual qRT–PCR: (i) having at least 50 copies by Solexa sequencing while miR-39. The NormFinder is an excel-add-in software and takes both intra- and intergroup variation into account (26). NormFinder can simultaneously provide the single most stable ECM and best combination of ECM pair that is more stable than single ECM. The lower the stability value is, the higher the miRNAs selected for validation

miRNAs selection for validation

In the first phase study, we employed Solexa sequencing as well as TLDA chips to identify the candidate ECMs. For Solexa sequencing, the final reads of each miRNA in each pooled sample were determined by normalization with the total reads of all called miRNAs in the sample. We selected those miRNAs that satisfied the following criteria as candidate ECMs for further individual qRT–PCR confirmation: (i) having at least 50 average copies among 10 pooling samples by Solexa sequencing, (ii) having at most 25 of the average Ct value and at most 5% of the coefficient of variation (CV) for Ct value by TLDA chips and (iii) showing the potential as ECMs in both Solexa sequencing and TLDA chips.

In the second phase study, based on both scientific and applicable considerations, miRNAs satisfied the following criteria were chosen for further confirmation by individual qRT–PCR: (i) having at least 50 copies by Solexa sequencing in either of the two pooled samples, (ii) having at most 35 of Ct value by TLDA in both of the two pools, (iii) having the same direction of results from Solexa sequencing and TLDA, (iv) showing 4-fold altered expression by Solexa sequencing while ΔΔCt > 3 by TLDA or (v) showing 8-fold altered expression by Solexa sequencing while ΔΔCt > 2 by TLDA.

Statistical analysis

The stability of each candidate ECMs was evaluated by the NormFinder by inputting the expression value of each candidate ECM that normalized by cel-mir-39. The NormFinder is an excel-add-in software and takes both intra- and intergroup variation into account (26). NormFinder can simultaneously provide the single most stable ECM and best combination of ECM pair that is more stable than single ECM. The lower the stability value is, the higher the

Serum miRNA and breast cancer

Serum preparation and RNA isolation

Five ml venous blood was collected from each participant. The whole blood was separated into serum and cellular fractions within 1 h for the 50 samples that used for ECM identification and within 24 h for the samples that used for the breast cancer association study, by centrifugation at 4000 r.p.m. for 10 min, followed by a 15 min high-speed centrifugation at 12 000 r.p.m. to completely remove cell debris. The supernatant serum was stored at −80°C until analysis.

We used a modified method to isolate the total RNAs (25). In brief, the Trizol Reagent (Invitrogen, Carlsbad, CA) was used for serum denaturing and Qiagen miRNeasy Mini kit (Qiagen, Valencia, CA) for RNA collection and purification according to the manufacturer’s protocol. To control sample to serum variability in RNA extraction and/or purification procedures, after the initial denaturing step, we spiked-in synthetic C. elegans miR-39 (cel-mir-39) (Takara, Japan) to a final concentration of 10−4 mmol/l. Equal volume of serum was processed for each sample.

Solexa sequencing and TLDA chip assays

The Solexa sequencing assay has been described previously (9,11). Sequencing analysis was performed with the Illumina’s Solexa Sequencer (Illumina, San Diego, CA), and the clean readouts were compared with the miRBase database (http://microrna.sanger.ac.uk, release 12.0).

For the TLDA Chip (Applied Biosystems, CA, V2.0) screening, 960 μl serum from each pooled sample was used. Megaplex RT reactions and preamplification reactions were run according to the manufacturer’s protocol. Then, 75 μl of each pooled sample was determined by PreAmp product, and 9 μl diluted PreAmp product was used to run the RT–PCR reactions by dispensing 100 μl of the PCR reaction mix into each port of the TaqMan miRNA array. The default PCR procedure was used and the analysis was performed by using RQ manager software (Applied Biosystems). ΔCt and ΔΔCt were calculated using the following mathematical formula: ΔCt = Ct sample − Ct cel−miR-39, ΔΔCt = ΔCt case − ΔCt control. Finally, the ΔΔCt was normalized against the cel-mir-39, miR-191 and miR-484, respectively.

miRNAs selection for validation

In the first phase study, we employed Solexa sequencing as well as TLDA chips to identify the candidate ECMs. For Solexa sequencing, the final reads of each miRNA in each pooled sample were determined by normalization with the total reads of all called miRNAs in the sample. We selected those miRNAs that satisfied the following criteria as candidate ECMs for further individual qRT–PCR confirmation: (i) having at least 50 average copies among 10 pooling samples by Solexa sequencing, (ii) having at most 25 of the average Ct value and at most 5% of the coefficient of variation (CV) for Ct value by TLDA chips and (iii) showing the potential as ECMs in both Solexa sequencing and TLDA chips.

In the second phase study, based on both scientific and applicable considerations, miRNAs satisfied the following criteria were chosen for further confirmation by individual qRT–PCR: (i) having at least 50 copies by Solexa sequencing in either of the two pooled samples, (ii) having at most 35 of Ct value by TLDA in both of the two pools, (iii) having the same direction of results from Solexa sequencing and TLDA, (iv) showing 4-fold altered expression by Solexa sequencing while ΔΔCt > 3 by TLDA or (v) showing 8-fold altered expression by Solexa sequencing while ΔΔCt > 2 by TLDA.

qRT–PCR assay

We used TaqMan miRNA probes (Applied Biosystems) to perform qRT–PCR assay according to the manufacturer’s instructions. In each step from serum purification to qRT–PCR, an equal volume (100 μl) of serum sample was processed. The total RNA was reverse transcribed to complementary DNA by using TaqMan miRNA RT Kit and stem-loop RT primers (Applied Biosystems). RT–PCR was performed using the TaqMan PCR kit on the ABI 7900 Real-Time PCR System (Applied Biosystems). The reactions were initiated in a 384-well optical plate at 95°C for 5 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. We assigned equal number of patients and controls in one plate and run the RT–PCR for target miRNAs and control miRNAs simultaneously. All reactions, including no template controls, were run in triplicate. The CT values were determined using the fixed threshold settings. The relative expression level of target miRNAs were normalized by the spike-in cel-mir-39, miR-191, miR-484 and the combination of miR-191 and miR-484, respectively.

Statistical analysis

The stability of each candidate ECMs was evaluated by the NormFinder by inputting the expression value of each candidate ECM that normalized by cel-mir-39. The NormFinder is an excel-add-in software and takes both intra- and intergroup variation into account (26). NormFinder can simultaneously provide the single most stable ECM and best combination of ECM pair that is more stable than single ECM. The lower the stability value is, the higher the

Fig. 1. Workflow chart of the experimental design.
gene stability of the ECM is. In addition, we also calculated the CV of each miRNA to assess their variations among samples.

Differences in demographic and clinical characteristics and mean expression levels of miRNAs were evaluated by t-test or the student’s t-test between cases and controls. The intensity value was classified into low and high expression according to the 85th percentile of miRNA expression levels of control samples in the discovery stage and was applied directly to the risk score function to see the consistency of the predictive effect of the identified miRNAs. The associations between miRNA expression levels and breast cancer risk were estimated by the odds ratios and their 95% confidence intervals from both univariate and multivariate logistic regression analyses.

To investigate the effectiveness of the four-miRNA signature (miR-16, miR-25, miR-222 and miR-324-3p) for breast cancer risk prediction, risk scoring was constructed. The upper 95% reference interval of each miRNA value in controls was set as the threshold to code the expression level of the corresponding miRNA for each sample in discovery and validation stages. The risk score of each miRNA, denoted as \( s_i \), was calculated as follows:

\[
s_i = \begin{cases} 
1 & \text{(high expression level)} \quad \text{if } r_i > t_i, \\
0 & \text{(low expression level)} \quad \text{otherwise}
\end{cases}
\]

Here, we used \( i \) to denote the \( i \)th sample, \( j \) to denote the \( j \)th miRNA.

When taking into account the strength of each miRNA associated with breast cancer risk, we assigned each patient a risk score function according to a linear combination of the expression level of the corresponding miRNA for each sample in discovery and validation stages. The risk score of each miRNA, denoted as \( s_i \), was calculated as follows:

\[
rsf_i = \sum_{j=1}^{k} \text{sign}(s_j) \cdot w_j \cdot x_{ij},
\]

where \( s_j \) is the expression level of the \( j \)th miRNA in the \( i \)th sample, \( w_j \) is the weight of the risk score of miRNA \( j \), and \( x_{ij} \) is the expression level of the \( j \)th miRNA in the \( i \)th sample.

The area under the receiver operating characteristic curve (AUC) was calculated for each of the combination of miRNAs, respectively, in order to assess the individual and accumulative effects of the miRNAs on breast cancer risk prediction. All the statistical analyses were performed with Statistical Analysis System software (version 9.1.3; SAS Institute, NC). A P-value of <0.05 was considered statistically significant, and all tests were two tailed.

Results

Stability of candidate ECMs

Thirty-two miRNAs satisfied the criteria of Solexa sequencing while 12 miRNAs of TDLA. Three miRNAs (miR-191, miR-320 and miR-484) overlapped from the two platforms and were subjected to further comparison (Figure 2). Because miR-16 was frequently used as an endogenous control in previous studies, we also included it in further analysis. The stability of these four miRNAs was then tested in 50 individual samples by qRT–PCR and evaluated by NormFinder. As a result, the CV for cel-miR-39 across 50 samples is 2.7%, suggesting that the experimental variation is acceptable. As shown in Supplementary Figure 1 and Table 1, available at Carcinogenesis Online, with a stability value of 0.337, NormFinder selected miR-484 as the most stably expressed single miRNA across different cancer samples and health controls. The best combination of two miRNAs, miR-484 and miR-191 (mean of the two miRNAs), further reduced the NormFinder stability value to 0.305. Obviously, miR-16 was not a suitable ECM according to its stability value and CV (Supplementary Figure 1 and Table 1, available at Carcinogenesis Online). To better evaluate the stability and the application of the ECMs, we then performed parallel analysis by normalizing with miR-484, miR-191 and cel-miR-39, respectively, in further analysis.

Detection of differentially expressed miRNAs of breast cancer in the discovery stage

Characteristics of study populations for breast cancer cases and controls are shown in Supplementary Table 2, available at Carcinogenesis Online. Overall, 51 miRNAs showed more than 4-fold altered expression between the two pooled samples by Solexa sequencing, whereas for TDLA chips, 203 miRNAs showed \( \Delta \text{AC}_{t} > 2 \) (i.e. 4-fold) when normalized by cel-miR-39, 383 miRNAs when normalized by miR-191 and 385 miRNAs when normalized by miR-484. As a result,

In the above equation, \( s_j \) is the risk score for miRNA \( j \) on sample \( i \), \( W_i \) is the weight of the risk score of miRNA \( j \). To determine the signs and \( W_i \), four univariate logistic regression models were fitted using the disease status with each of the risk scores. The regression coefficient of each risk score was used as the weight to indicate the contribution of each miRNA to the risk scoring function.

Table I. Results of Solexa sequencing and TDLA in the discovery stage of 10 miRNAs for breast cancer risk

<table>
<thead>
<tr>
<th>Patient group (n = 48)</th>
<th>Control group (n = 48)</th>
<th>Ratio</th>
<th>( \Delta \text{AC}_{t}^{a} )</th>
<th>( \Delta \text{AC}_{t}^{b} )</th>
<th>( \Delta \text{AC}_{t}^{c} )</th>
<th>( \Delta \text{AC}_{t}^{d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies ( ^{a} )</td>
<td>Mean ± SD</td>
<td>Median</td>
<td>Copies ( ^{a} )</td>
<td>Mean ± SD</td>
<td>Median</td>
<td>(cel-39)</td>
</tr>
<tr>
<td>let-7b</td>
<td>6884</td>
<td>-1.784</td>
<td>4.08 ± 4.83</td>
<td>2.71</td>
<td>1674</td>
<td>2.220</td>
</tr>
<tr>
<td>mir-151-3p</td>
<td>693</td>
<td>1.065</td>
<td>8.07 ± 3.82</td>
<td>6.05</td>
<td>18</td>
<td>4.118</td>
</tr>
<tr>
<td>mir-16</td>
<td>146</td>
<td>-5.766</td>
<td>67.73 ± 55.51</td>
<td>52.27</td>
<td>48</td>
<td>0.214</td>
</tr>
<tr>
<td>mir-222</td>
<td>508</td>
<td>-3.791</td>
<td>1.51 ± 1.53</td>
<td>0.98</td>
<td>106</td>
<td>1.173</td>
</tr>
<tr>
<td>mir-25</td>
<td>2575</td>
<td>1.238</td>
<td>2.57 ± 2.97</td>
<td>1.84</td>
<td>46</td>
<td>5.214</td>
</tr>
<tr>
<td>mir-30a</td>
<td>75</td>
<td>0.009</td>
<td>3.28 ± 3.18</td>
<td>2.11</td>
<td>15</td>
<td>4.119</td>
</tr>
<tr>
<td>mir-451</td>
<td>418</td>
<td>-0.755</td>
<td>38.32 ± 48.47</td>
<td>19.96</td>
<td>64</td>
<td>3.222</td>
</tr>
<tr>
<td>mir-486</td>
<td>97088</td>
<td>0.241</td>
<td>54.67 ± 62.12</td>
<td>38.49</td>
<td>9509</td>
<td>3.212</td>
</tr>
</tbody>
</table>

\( ^{a} \)Solexa sequencing results.  
\( ^{b} \)TDLA results. \( \Delta \text{CT} = \text{CT}_{\text{sample}} - \text{CT}_{\text{U6}} \).  
\( ^{c} \)Copies of patients group/copies of control group.  
\( \Delta \text{AC}_{t} = \Delta \text{AC}_{t}^{a} \cdot \Delta \text{AC}_{t}^{b} \cdot \Delta \text{AC}_{t}^{c} \cdot \Delta \text{AC}_{t}^{d} \).  
\( ^{d} \)Student’s t-test for mean ± SD. Significantly differentially expressed miRNAs are marked in bold.
9 miRNAs were identified when normalized by cel-miR-39, 13 miRNAs when normalized by miR-191 and 13 miRNAs when normalized by miR-484. Among them, 9 overlapped miRNAs were subjected to further analyses (Supplementary Figure 2 and Table 1, available at Carcinogenesis Online). miR-16 satisfied the criteria of TLDA and was approaching the threshold of Solexa sequencing (Solexa sequencing: 3.042-fold; TLDA: $\Delta \Delta C_T$ cel-miR-39 $= -5.670$, $\Delta \Delta C_T$ miR-191 $= -5.967$, $\Delta \Delta C_T$ miR-484 $= -5.975$), therefore, we included it for further analysis.

**Individual qRT–PCR confirmation and validation**

We then performed individual qRT–PCR detection on the 96 discovery samples to quantify each of the 10 identified miRNAs. Cel-miR-39, miR-191, miR-484 and the combination of miR-191 and miR-484 were used to normalize the expression of each miRNA, respectively. As a result, the expression levels of four miRNAs (miR-16, miR-25, miR-222 and miR-324-3p) were confirmed to be significantly overexpressed in the breast cancer patients than that in the controls ($P < 0.001$) (Table I; Figure 2 for normalization on the combination of miR-191 and miR-484; Supplementary Figure 3 and Table 3, available at Carcinogenesis Online).

The intensity values of the 95th percentile of each miRNA expression levels among controls in the discovery stage were used to classify low and high expression of the miRNA, and these cutoff values were applied directly to the validation samples and also combined samples (Table II and Supplementary Table 4, available at Carcinogenesis Online). The four miRNAs were consistently significantly overexpressed among breast cancer cases (Table II and Figure 3; Supplementary Figure 3 and Table 4, available at Carcinogenesis Online). As listed in Supplementary Table 5, available at Carcinogenesis Online, high expression levels of four serum miRNAs were all individually and jointly associated with significantly increased risk of breast cancer after adjustment for age, age at menarche and menopausal status.

A linear combination of the expression levels of the four miRNAs weighted by the regression coefficient was used to form a risk score of the four-miRNA signature on all of the 248 subjects. We then used the four-miRNA signature to divide samples into high- and low-risk groups according to the median risk score and found that the stratified effects of high-risk group by age, age at menarche, menopause status and estrogen receptor status all showed increased breast cancer risk without heterogeneity, suggesting independent prediction role of the four-miRNA signature (Supplementary Table 6, available at Carcinogenesis Online).

**Receiver operating characteristic curve and AUC analyses**

We conducted the ROC curve and AUC analyses to assess the sensitivity and specificity of the miRNA signature individually and in combination for breast cancer risk estimation.

When normalized with the combination of miR-191 and miR-484, for single miRNA, the AUC was 0.927, 0.823, 0.865 and 0.781 in the

<p>| Table II. Serum expression levels of the four miRNAs and risk of breast cancer |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Stage</th>
<th>miRNA</th>
<th>$\beta$</th>
<th>Cutoff point</th>
<th>SE</th>
<th>$\chi^2$</th>
<th>$P$</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>miR16</td>
<td>5.287</td>
<td>18.75</td>
<td>0.863</td>
<td>37.524</td>
<td>&lt;0.0001</td>
<td>0.927</td>
</tr>
<tr>
<td>(n = 96)</td>
<td>miR25</td>
<td>3.924</td>
<td>1.04</td>
<td>0.787</td>
<td>24.886</td>
<td>&lt;0.0001</td>
<td>0.823</td>
</tr>
<tr>
<td>miR222</td>
<td>4.349</td>
<td>0.64</td>
<td>0.800</td>
<td>29.561</td>
<td>&lt;0.0001</td>
<td>0.865</td>
<td></td>
</tr>
<tr>
<td>miR324-3p</td>
<td>3.558</td>
<td>0.12</td>
<td>0.780</td>
<td>20.798</td>
<td>&lt;0.0001</td>
<td>0.781</td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>miR16</td>
<td>4.968</td>
<td>18.75</td>
<td>0.673</td>
<td>54.452</td>
<td>&lt;0.0001</td>
<td>0.908</td>
</tr>
<tr>
<td>(n = 152)</td>
<td>miR25</td>
<td>4.090</td>
<td>1.04</td>
<td>0.641</td>
<td>40.693</td>
<td>&lt;0.0001</td>
<td>0.836</td>
</tr>
<tr>
<td>miR222</td>
<td>4.056</td>
<td>0.64</td>
<td>0.545</td>
<td>55.367</td>
<td>&lt;0.0001</td>
<td>0.868</td>
<td></td>
</tr>
<tr>
<td>miR324-3p</td>
<td>2.785</td>
<td>0.12</td>
<td>0.563</td>
<td>24.494</td>
<td>&lt;0.0001</td>
<td>0.711</td>
<td></td>
</tr>
</tbody>
</table>

*Normalized by the combination of miR-191 and miR-484. The code of the miRNA expression value was 0 or 1 according to the 95th percentile of control samples’ miRNAs expression levels in the discovery stage.

Fig. 3. Expression levels of the four miRNAs normalized with the combination of miR-191 and miR-484. DS, discovery stage; VS, validation stage; com: DS + VS.
Identification of biomarkers for cancer diagnosis and prognosis is of particular interest because the ability to detect cancers at early stages is a key factor in reducing death rates and improving patients’ survival. In the past 10 years, early detection has resulted in 2.3% annual decrease in death rates of breast cancer (27). However, the currently available breast cancer screening techniques may miss the patients with early stages or misguide patients who have benign abnormalities to conduct invasive examine such as surgical biopsy (28). Plasma/serum miRNAs as a class of novel promising non-invasive biomarkers for cancer risk prediction have their unique merits. For example, they are stable, easily to be detected and potentially cancer specific (9). In this study, we identified the four-miRNA signature (i.e. miR-16, miR-25, miR-222 and miR-324-3p) derived from the genome-wide serum miRNA expression profiling may serve as a non-invasive predictor for the risk of breast cancer. Accurate normalization of miRNA expression is a prerequisite to minimize variation that can mask or exaggerate biologically meaningful changes and to obtain reliable results. In this study, we firstly systematically screened candidate ECMs and identified the combination of two miRNAs (miR-191 and miR-484) may serve as candidates.

Our recent studies and others’ have implied that the patterns of plasma/serum miRNAs may serve as non-invasive biomarkers for cancers (9,11,12,17,22,29–32). For example, Brase et al. (30) screened 667 miRNAs in plasma samples from patients with metastatic and localized prostate cancer and found that miR-375 and miR-141 were the most pronounced biomarkers for high-risk prostate cancer. Significant upregulation of miR-21, miR-92, miR-93, miR-126 and miR-29a and downregulation of miR-155, miR-127 and miR-99b were found in plasma of patients with ovarian cancer compared with controls (31). Ng et al. (17) reported that the expression levels of plasma miR-92 were significantly elevated in colorectal cancer. For the four miRNA identified in the current study, the aberrant expression of plasma/serum miR-16 was reported to be associated with the risk of pancreatic cancer and HCC (33,34), miR-25 was found to be a novel serum marker for HCC (15) and miR-222 was also a serum classifier for HCC (35). These studies all indicated that the use of plasma/serum miRNAs as biomarkers for cancer predictor is both available and feasible.

In addition to those from the plasma/serum, Heneghan et al. (36) suggested that miRNA signatures from blood cells can also be used for breast cancer diagnosis and found that the expression levels of miR-195 and let-7a were significantly decreased in breast cancer patients. Although the exact mechanisms how the small RNAs have entered the plasma/serum and whether or not they are biologically functional warrant further investigations, we believe that plasma/serum miRNAs may serve as non-invasive biomarkers for cancer risk prediction have their unique merits. For example, they are stable, easily to be detected and potentially cancer specific (9). In this study, we firstly systematically screened candidate ECMs and identified the combination of two miRNAs (miR-191 and miR-484) respectively, the AUC was 0.920, 0.936 and 0.966 in the discovery stage, 0.901, 0.882 and 0.901 in the validation stage and 0.903, 0.883 and 0.911 in the combined 248 samples (Supplementary Table 4, available at Carcinogenesis Online).

Discussion

Identification of biomarkers for cancer diagnosis and prognosis is of particular interest because the ability to detect cancers at early stages is a key factor in reducing death rates and improving patients’ survival. In the past 10 years, early detection has resulted in 2.3% annual decrease in death rates of breast cancer (27). However, the currently available breast cancer screening techniques may miss the patients with early stages or misguide patients who have benign abnormalities to conduct invasive examine such as surgical biopsy (28). Plasma/serum miRNAs as a class of novel promising non-invasive biomarkers for cancer risk prediction have their unique merits. For example, they are stable, easily to be detected and potentially cancer specific (9). In this study, we identified the four-miRNA signature (i.e. miR-16, miR-25, miR-222 and miR-324-3p) derived from the genome-wide serum miRNA expression profiling may serve as a non-invasive predictor for the risk of breast cancer. Accurate normalization of miRNA expression is a prerequisite to minimize variation that can mask or exaggerate biologically meaningful changes and to obtain reliable results. In this study, we firstly systematically screened candidate ECMs and identified the combination of two miRNAs (miR-191 and miR-484) may serve as candidates.

Our recent studies and others’ have implied that the patterns of plasma/serum miRNAs may serve as non-invasive biomarkers for cancers (9,11,12,17,22,29–32). For example, Brase et al. (30) screened 667 miRNAs in plasma samples from patients with metastatic and localized prostate cancer and found that miR-375 and miR-141 were the most pronounced biomarkers for high-risk prostate cancer. Significant upregulation of miR-21, miR-92, miR-93, miR-126 and miR-29a and downregulation of miR-155, miR-127 and miR-99b were found in plasma of patients with ovarian cancer compared with controls (31). Ng et al. (17) reported that the expression levels of plasma miR-92 were significantly elevated in colorectal cancer. For the four miRNA identified in the current study, the aberrant expression of plasma/serum miR-16 was reported to be associated with the risk of pancreatic cancer and HCC (33,34), miR-25 was found to be a novel serum marker for HCC (15) and miR-222 was also a serum classifier for HCC (35). These studies all indicated that the use of plasma/serum miRNAs as biomarkers for cancer predictor is both available and feasible.

In addition to those from the plasma/serum, Heneghan et al. (36) suggested that miRNA signatures from blood cells can also be used for breast cancer diagnosis and found that the expression levels of miR-195 and let-7a were significantly decreased in breast cancer patients. Although the exact mechanisms how the small RNAs have entered the plasma/serum and whether or not they are biologically functional warrant further investigations, we believe that plasma/serum miRNAs may serve as non-invasive biomarkers for breast cancer (normalized by the combination of miR-191 and miR-484). (A) Discovery stage. AUC = 0.954, sensitivity = 0.917, specificity = 0.896, cutoff value = 3.924. (B) Validation stage. AUC = 0.928, sensitivity = 0.921, specificity = 0.934, cutoff value = 3.924. (C) Combined discovery and validation stages. AUC = 0.919, sensitivity = 0.919, specificity = 0.919, cutoff value = 3.924.
Serum microRNAs are a unique diagnostic system compared with those from cancer tissues and blood cells (9,11). No doubt plasma/serum microRNAs are cancer related, but the releasing mechanisms may be complicated; we and other groups showed that tumor-derived microvesicles/exosomes might be involved rather than a simply leaking from cancer cells (37–39). In this sense, the signatures from blood cells, although easier to be detected, may not be cancer specific.

A better understanding of the mRNA targets and the molecular mechanisms by which the microRNAs regulate breast cancer risk may promote their clinical application. Among the four miRNAs, miR-222 has been extensively evaluated in the context of breast cancer. miR-222 was a target of ESR1, and was reported to be dramatically higher in ESR1 (−) cells, inhibit ERα expression and induce the cell cycle checkpoint transition from G1 to S phase (40–42). miR-222 can trigger a malignant transformation by altering the expression levels of many important genes involved in apoptosis, cell cycle, tumor invasion and metastasis, transformation by altering the expression levels of many important genes involved in apoptosis, cell cycle, tumor invasion and metastasis, etc.

miR-222 is known to cooperate with its host gene MCM7 in cellular transformation both in vitro and in vivo in transgenic mice (46). miR-16 was reported to be located at chromosome 1q24, a region deleted in more than half of B-cell chronic lymphocytic leukemia and showed an important role in the regulation of Wip1 phosphatase in the DNA damage response and mammalian tumorigenesis (47,48). It was also reported that miR-16 can regulate HuR translation and link this regulatory pathway to human breast cancer (49). Additional investigation of the regulatory mechanism of these miRNAs and their target mRNAs may improve our understanding of the molecular pathogenesis of breast cancer as well as the effectiveness in identifying potential therapy targets and surveillance markers for breast cancer. Moreover, studies also showed the functions of the two ECMs in disease. For example, Elyakim et al. (50) reported the suppression of miR-191 can decrease cell proliferation and induce apoptosis in vitro and significantly reduce tumor masses in mouse model of HCC. Additionally, miR-191 regulated cell growth and migration rate of follicular thyroid cell lines by mir-39 to control the experimental variation and conducted qRT-PCR and qPCR with well quality control. Limitation of our study was the relatively small sample size. Therefore, the generalization of the two ECMs and the four-miRNA signatures identified in this study warrants further investigations in additional studies involving different ethnic populations and prospective study designs.

Supplementary material

Supplementary Figures 1–3 and Tables 1–6 can be found at http://carcin.oxfordjournals.org/

Funding

This work was supported in part by National Natural Science Foundation of China (81071715), Jiangsu Natural Science Foundation (BK2011028), the Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China (122031), Foundation for the Author of National Excellent Doctoral Dissertation (201081), the Program for Changjiang Scholars and Innovative Research Team in University (IRT0631) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Conflict of Interest Statement: None declared.

References

identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.*, **64**, 5245–5250.


Received November 22, 2011; revised January 7, 2012; accepted January 26, 2012.